



**UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA  
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA**



**DEYSE VALVERDE GOMES DE ANDRADE**

**DETERMINAÇÃO ESTRUTURAL DO MODELO  
MUTANTE DA CISTEÍNA PROTEASE DO *THEOBROMA*  
CACAO POR MODELAGEM COMPARATIVA E CÁLCULOS  
QM/MM**

Feira de Santana, BA  
Julho de 2009

**DEYSE VALVERDE GOMES DE ANDRADE**

**DETERMINAÇÃO ESTRUTURAL DO MODELO  
MUTANTE DA CISTEÍNA PROTEASE DO *THEOBROMA*  
CACAO POR MODELAGEM COMPARATIVA E CÁLCULOS  
QM/MM**

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Biotecnologia da Universidade Estadual de Feira de Santana, como requisito a defesa e obtenção do título de Mestre em Biotecnologia.

Orientador: Prof. Dr. Alex Gutterres Taranto  
Co-orientador: Prof. Dr. Aristóteles Góes Neto

Feira de Santana, BA  
Julho de 2009

**Dedicado a Deus, minha família,  
ao meu namorado e a todas as  
pessoas que fazem parte da  
minha vida.**

## AGRADECIMENTOS

À Deus, Meu Pai, Senhor e Salvador, pela paz, sabedoria, cuidado e amor.

Ao Prof. Dr. Alex Gutteres Taranto, Orientador desta dissertação, pela confiança, assistência, empenho, compreensão e apoio. Gostaria de ratificar a sua competência, paciência e exímia participação que muito contribuíram para o meu aprendizado e desenvolvimento deste trabalho. Muito obrigada por ter sido um referencial á minha formação científica, profissional e didática.

Ao Prof. Dr. Aristóteles Góes Neto, Co-orientador desta dissertação e Coordenador do Programa de Pós-Graduação em Biotecnologia e Recursos Naturais Renováveis – UEFS pela oportunidade de crescimento, aprendizado, realização profissional e pessoal.

À Profa. Dra. Fabienne Florence Lucienne Micheli, pela concessão da seqüência primária da cisteína protease do *Theobroma cacao*.

Ao Prof. Dr. Ramon dos Santos El-Bachá e colegas do Laboratório de Neuroquímica e Biologia Molecular – LABNQ/UFBA. Em especial ao Prof. Dr. Lisandro Diego Giraldez Alvarez, meu pai científico.

À Helton Ricardo, Secretário do Programa de Pós-Graduação em Biotecnologia e Recursos Naturais Renováveis – PPGBIOTEC/UEFS, pela Eficiência, Eficácia, Competência e Carisma (EECC) no exercício de suas atividades administrativas. Muito obrigada, amigo!

Aos meus colegas do PPGBIOTEC, Bruno Andrade, pelas palavras e pelo apoio em momentos difíceis, à Viviane Galvão, pela amizade e interdisciplinaridade, à Rafaela e Catinha pelos momentos de integração e socialização do conhecimento.

Ao Grupo de Pesquisa do Laboratório de Modelagem Molecular – LMM/UEFS, em Especial a André Teles Lacerda pelas co-ooooorientações; a Franco Henrique, Vivian Ubelino, JotaJota e Manoelito Coelho, pela contribuição e amizade, as minhas colegas

Monalisa e Alesandra Marques, pelos momentos de descontração e aprendizado.

Aos Profs. Drs. Ângelo Amâncio Duarte, Sandra Aparecida de Assis, Fernando Flores Santos Ribeiro e Carla Rodrigues Cardoso Branco por aceitarem participar da Banca de Defesa desta Dissertação, proporcionando discussões e sugestões que servirão para crescimento, aprendizado e incentivo à pesquisa.

Aos meus familiares, minhas raízes que sempre me deram amor e força, valorizando meus potenciais. Em especial aos meus pais, meus fãs de certidão, clubinho, carteirinha e camisa, pois tenho ciência que ao longo da minha vida sempre contarei com o amor, a dedicação, as orientações e apoio incondicional.

Ao meu namorado Samuel Akira Kimura, pelo companheirismo, cumplicidade, amor e compreensão nos momentos de stress científico e TPM.

À todos os meus amigos e amigas que sempre estiveram presentes me aconselhando e incentivando com carinho e dedicação. Em especial a Mônica Gomes, Silvia Falcão, Angélica Andrade, Keila Loren, Help, Daise e Mark Zanone, as MSP's : Silvia Carla e Lucinéia Gomes.

A Fundação de Amparo a Pesquisa do Estado da Bahia (FAPESB), pelo apoio financeiro para o desenvolvimento deste trabalho.

À todas as pessoas que, direta ou indiretamente, contribuíram para a execução dessa Dissertação de Mestrado.

O conhecimento é a maior riqueza para valorização da vida.

## RESUMO

A cultura do cacau no Brasil foi altamente prejudicada pelo *Moniliophthora perniciosa*, o fungo que provoca a praga da vassoura-de-bruxa no cacau, provocando uma diminuição significativa na produção de cacau no Brasil. Uma importante estratégia para controlar a vassoura-de-bruxa consiste no estudo da interação *Theobroma cacao* – *M. perniciosa*. No entanto, há pouca informação sobre o mecanismo de interação molecular envolvido no processo de resistência/vulnerabilidade do cacaueiro. Este fato incentivou a implantação do Projeto Genoma do cacau e, a partir deste foi verificado que genes que codificam as proteases estão envolvidos na fitopatogênese da praga, principalmente a cisteína protease. Esta é expressa durante o processo de maturação das sementes e está presente no período necrofítico da doença. Além de seus aspectos biológicos, as proteases são muito importantes na área biotecnológica devido à sua grande aplicação industrial. As proteases são amplamente empregadas na indústria alimentícia, na produção de detergentes e indústrias farmacêuticas. Neste trabalho determinamos a estrutura 3D da cisteína protease do *T. cacao* através de Modelagem Comparativa. Desta forma, a seqüência primária da cisteína protease do *T. cacao* foi submetida ao BLASTp obtendo-se a proteína 1PCI|A com 36% de similaridade estrutural como molde. A seguir, três modelos foram construídos (dv1, dv2 e dv3), os quais foram refinados e validados pelos softwares AMBER 9.0 e PROCHECK, respectivamente. Dentre os modelos construídos, o modelo dv1 proporcionou um melhor Gráfico de Ramachandran apresentando 95% dos aminoácidos em regiões energeticamente favoráveis. Este modelo final é constituído por 171 aminoácidos, formados por 2693 átomos, unidos por 2719 ligações químicas. A estrutura 3D desta enzima apresenta 7  $\alpha$ -hélices, 23 *turns* e 2 folhas- $\beta$ . A região do sítio ativo esta conservada, sendo representada pelos resíduos Cys25 e His159. A partir deste modelo, foi então gerado um modelo mutante pela substituição do resíduo His/Gly159. Este ao ser validado também apresentou similares características de dv1. Estudos da interação entre a estrutura mutante com os íons metálicos  $Zn^{+2}$ ,  $Cu^{+1}$ ,  $Cu^{+2}$  e  $Cd^{+2}$  foram realizados através do método QM/MM implementado no programa Gaussian 03W. Como resultado, a proteína mutante é capaz de quelar todos os íons estudados, e mostrou-se ser seletiva para o  $Cd^{+2}$  com energia de ligação de  $-59801 \times 10^3$  kcal/mol. Com isso o presente trabalho sugere que modificações na seqüência da cisteína protease do *T. Cacao* pode levar ao desenvolvimento de novos produtos de interesse comercial.

**Palavras-chave:** *Moniliophthora perniciosa*, *Theobroma cacao*, cisteína protease, modelagem comparativa, QM/MM.

## ABSTRACT

The culture of cacao in Brazil was highly harmed by *Moniliophthora perniciosa*, the fungus that causes witches' broom disease of cocoa. This disease decreases significantly the cocoa production. An important strategy for control of the witches' broom is the molecular study of the interaction between cacao-*M. perniciosa*. However, there is little information about of the mechanism of molecular interaction involved in resistance/vulnerability of cacao tree. To address this problem, the Genome Project showed genes of cysteine protease involved in the mechanism of resistance/vulnerability. Cysteine protease is expressed during the process of the maturation of the seed and it is present in necrotic period of the disease. Furthermore, proteases have a wide application in feed products, detergents and pharmaceutical industries. This work constructed the 3D structure of the cysteine protease of *T. cacao* by comparative modeling. Thus, the primary sequence of the cysteine protease of *T. cacao* was submitted to BLASTp obtaining the protein 1PCI|A with 36% of structural similarity as such. Below, three models were constructed (dv1, dv2 and dv3), which were refined and validated by AMBER 9.0 and PROCHECK software, respectively. Among these models, dv1 showed a better Ramachandran Plot with 95% of amino acids in favorable energy region. The final model consists of 171 amino acids, formed by 2693 atoms linked by 2719 chemical bonds. The 3D structure of this enzyme has 7  $\alpha$ -helix, 23 turns and 2  $\beta$ -sheets. The region of the conserved active site is represented by residues Cys25 and His159. From this model, a mutant model was then generated by replacing His159/Gly. This also was evaluated showing similar dv1 characteristics. Studies of the interaction between the mutant structure with the metal ions  $^{+2}$ ,  $\text{Cu}^{+1}$ ,  $\text{Cu}^{+2}$  and  $\text{Cd}^{+2}$  were carried out by QM/MM approach implemented in Gaussian 03W. As a result, the mutant protein is able to complex with all of them, and showed selective for  $\text{Cd}^{+2}$  with  $-59801 \times 10^3$  kcal/mol. Therefore this study suggests that changes in the sequence of the cysteine protease of *T. Cacao* can lead to the development new products of commercial interest.

**Keywords:** *Moniliophthora perniciosa*, *Theobroma cacao*, cysteine protease, comparative modeling, QM/MM.



## LISTA DE TABELAS

|   |    |
|---|----|
| <b>Tabela 1.</b> Lista dos principais bancos de dados disponíveis para análise comparativa de genomas.  | 30 |
| <b>Tabela 2.</b> Resultado do alinhamento do <i>BLASTp</i> da sequência primária da cisteína protease do <i>T. cacao</i> com seus respectivos moldes. | 52 |
| <b>Tabela 3.</b> Comparação entre os valores de energia do modelo mutante e sua interação com metais.   | 87 |

## LISTA DE ILUSTRAÇÕES

|  |    |
|--|----|
| <b>Figura 1:</b> Comparação entre os galhos e frutos do cacauero com a presença dos sintomas da vassoura-de-bruxa. Em A e C está representado o aspecto fenotípico do processo de dicarionização e fibulação do micélio fúngico secundário com a invasão das hifas nos ramos dos tecidos do hospedeiro culminando com a morte dos ramos e necrose do fruto. Em B, C, D e E está mostrado a presença das vassouras necróticas de coloração amarronzada no fruto. Fonte: CEPEC/CEPLAC, 2008. | 22 |
| <b>Figura 2:</b> Equação clássica dos métodos empíricos.   | 33 |
| <b>Figura 3:</b> Cálculo da energia total resultante pela aplicação do método ONIOM.   | 40 |
| <b>Figura 4.</b> Etapas da Modelagem Comparativa para a construção do modelo 3D da cisteína protease do <i>T. cacao</i> . Em azul encontram-se os programas utilizados em cada etapa.  | 42 |
| <b>Figura 5.</b> Seqüência de aminoácidos da cisteína protease do <i>T. cacao</i> .  | 43 |
| <b>Figura 6.</b> Seqüência de aminoácidos da cisteína protease com a permanência de <i>gaps</i> no modelo dv1.   | 45 |
| <b>Figura 7.</b> Seqüência de aminoácidos da cisteína protease com a deleção de <i>gaps</i> no modelo dv2.   | 45 |
| <b>Figura 8.</b> Seqüência primária da cisteína protease com a inserção de aminoácidos aos <i>gaps</i> no modelo dv3.  | 46 |
| <b>Figura 9.</b> Resultado do alinhamento do BLASTp da seqüência de aminoácidos da cisteína protease do <i>T. cacao</i> com seus respectivos moldes.   | 51 |
| <b>Figura 10.</b> Modelo estrutural da 1PCI A.   | 52 |
| <b>Figura 11.</b> Alinhamento entre a seqüência primária da proteína 1PCI A dissociada da molécula complexada.   | 54 |
| <b>Figura 12.</b> O alinhamento qualitativo da molécula complexante demonstrou melhor disposição entre os resíduos de aminoácidos da seqüência primária do molde e a seqüência da cisteína protease do <i>T. cacao</i> .   | 54 |
| <b>Figura 13.</b> Gráfico de Ramachandran para os 322 (A) e 216 (B) resíduos de aminoácidos da 1PCI A. As regiões representadas em vermelho e amarelo representam regiões mais favoráveis e as regiões em branco como desfavorável.  | 55 |
| <b>Figura 14.</b> Representação estrutural da seqüência primária da 1PCI A constituída por 216 aminoácidos, 08 $\alpha$ -hélices, 18 folhas $\beta$ , 22 <i>gaps</i> , 1634 átomos, 1675 ligações de hidrogênio  | 57 |
| <b>Figura 15.</b> (A) Estrutura 3D do modelo dv1 com 6 $\alpha$ -hélices, 17 <i>loops</i> e 14 folhas- $\beta$ ; (B) Estrutura do  | 58 |

modelo dv2 com 6  $\alpha$ -hélices, 16 *loops* e 10 folhas- $\beta$ . (C) Estrutura do modelo dv3 com 8  $\alpha$ -hélices, 22 *loops* e 22 folhas- $\beta$ .

**Figura 16.** (A) Modelagem das estruturas moleculares para obtenção do modelo dv1 visualizado na cor azul e na cor verde 1PCI|A. (B) Em destaque modelagem da região do sítio ativo. 59

**Figura 17.** (A) Modelagem das estruturas moleculares para obtenção do modelo dv2 (cor vermelha). 60  
(B) Em destaque os aminoácidos Cys15 e His49 do modelo sobreposto ao molde.

**Figura 18.** (A) Modelagem das estruturas moleculares para obtenção do modelo dv3 (cor vermelha). 61  
(B) Em destaque o alinhamento da Cys 25 e com ausencia de alinhamento entre as seqüências com a His159.

**Figura 19.** Representação da energia potencial de cada átomo na cadeia da proteína através do gráfico do ANOLEA para a validação dos modelos dv1(A), dv2(B) e dv3(C). O valor de energia é dado em E/kT, onde k corresponde a constante de Boltzman (0,582 Kcal/mol) e T corresponde a temperatura absoluta. As projeções em vermelho referem-se a um valor positivo de energia potencial para cada resíduo, e as projeções em verde identificam os valores de energia negativa, ou seja, valores satisfatórios. Em destaque as regiões do sítio ativo. 62

**Figura 20.** (A) Gráfico de Ramachandran do modelo dv1 com 72,9% dos aminoácido da cadeia principal nas regiões favoráveis (A e B) e 20,0% nas regiões permitida (a,b,l,p). Em destaque região do sítio; B e C. Ramachandran da Glicina e da Prolina, respectivamente, onde observa-se o resíduo Pro139 em região desfavorável (Fote: Procheck 3,0). (D) Gráfico de energia obtido pelo ANOLEA das cadeias laterais de aminoácidos (-21,4170 E/kT ; + 5,8440 E/kT "32,93%"). Em destaque região do sítio ativo. 65

**Figura 21.** (A) Gráfico de Ramachandran do modelo dv3 com 55,1% dos aminoácido da cadeia principal nas regiões favoráveis (A e B) e 31,9% contidos nas regiões permitida (a,b,l,p). (B) Gráfico de energia das cadeias laterais de aminoácidos (-11,2180 E/kT ; + 7,0350 E/kT "54,97%"). Em destaque regioao do sítio ativo. 66

**Figura 22.** Gráfico do RMS da cisteína protease do *T. cacao*. Estabilidade estrutural alcançada a partir de 500ps. 68

**Figura 23.** (A) Gráfico de Ramachandran do modelo 3D da cisteína protease do *T. cacao*. (B, C) Ramachandran da Glicína e Ramachandran da Prolina onde observa-se o resíduo Gly94 e 38 em região desfavorável. 69

**Figura 24.** Gráficos Chi-1 e Chi-2 dos resíduos. Os números de resíduos são mostrados entre parêntese. As regiões sombreadas representam áreas favoráveis, coloração vermelha representa o resíduo que está com o ângulo desfavorável. 71

**Figura 25.** Qualidade da cadeia principal do modelo estrutural da cisteína protease do *T. cacao*. O 72

eixo das abscissas está descrito em graus, e o das coordenadas está descrito em Å: (A) Avaliação do Gráfico de Ramachandran; (B) Planaridade da ligação peptídica; (C) Interações ruins; (D) Distorção dos carbonos alfa; (E) Energia das ligações de hidrogênio.

**Figura 26.** Desvio padrão para os ângulos diedros Chi1 do modelo estrutural da cisteína protease do *T. cacao*. (A) Conformação gauche menos; (B) Conformação trans; (C) Conformação gauche mais; (D) Somatório de todos os desvios padrões para Chi1; (E) Conformação trans para o ângulo torcional Chi2. O eixo das abscissas está descrito em graus, e o das coordenadas está descrito em Å. 73

**Figura 27.** Representação gráfica de cada átomo na cadeia de proteína durante a DM. As projeções em verde referem-se a um valor de energia favorável e, as projeções em vermelho a valores desfavoráveis de energia. 74

**Figura 28.** Estrutura 3D da cisteína protease do *T. cacao*. 75

**Figura 29.** Estrutura 3D da cisteína protease do *T. cacao*. Em destaque a região do sítio catalítico da enzima. 76

**Figura 30.** (A) Modelo estrutural dv1 representado em wireframe definido por QM/MM. (B) Em destaque o sítio ativo da molécula representado sob a forma de ballstick com 3,4Å entre os átomos de N e S da histidina e cisteína respectivamente. 77

**Figura 31.** (A) Modelo estrutural do mutante analisado por QM/MM. (B) Em destaque os aminoácidos glicína e cisteína com 3,4Å entre os átomos de H e S respectivamente. 79

**Figura 32.** (A) Gráfico de Ramachandran do modelo mutante da cisteína protease do *T. cacao*. (B e C) em destaque os resíduos Gly e Pro. 80

**Figura 33.** Representação gráfica de cada átomo na cadeia de proteína mutante da cisteína protease do *T. cacao* durante a DM. As projeções em verde referem-se a um valor de energia favorável e as projeções em vermelho a volores desfavoráveis de energia. 80

**Figura 34.** Gráficos Chi-1 e Chi-2 dos resíduos da estrutura mutante. Os números de resíduos são mostrados entre parêntese. As regiões sombreadas representam áreas favoráveis coloração vermelha representa o resíduo que está com o ângulo desfavorável. 82

**Figura 35.** Qualidade da cadeia principal do modelo mutante da cisteína protease do *T. cação*: (A) Avaliação do gráfico de Ramachandran; (B) Planaridade da ligação peptídica; (C) Interações ruins; (D) Distorção dos carbonos alfa; (E) Energia das ligações de hidrogênio. 83

|  |    |
|--|----|
| <b>Figura 36.</b> Desvio padrão para os ângulos diedros Chi1 do modelo mutante da cisteína protease do <i>T. cacao</i> . (A) Conformação gauche menos; (B) Conformação trans; (C) Conformação gauche mais; (D) Somatório de todos os desvios padrões para Chi1; (E) Conformação trans para o ângulo torcional Chi2. O eixo das abscissas está descrito em graus, e o das coordenadas está descrito em Å. | 84 |
| <b>Figura 37.</b> Gráfico do RMS do modelo mutante da cisteína protease do <i>T. cacao</i> . Estabilidade estrutural alcançada a partir de 837ps.  | 85 |
| <b>Figura 38.</b> Modelo mutante da cisteína protease do <i>T. cacao</i> .   | 85 |
| <b>Figura 39.</b> Equação da obtenção da energia de ligação ligante-proteína.  | 86 |

## **LISTA DE ABREVIATURAS E SIGLAS**

AIA - Auxina

CEPLAC - Comissão Executiva do Plano da Lavoura Cacaueira

CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico

DFT - Teoria do Funcional de Densidade

EMBRAPA - Empresa Brasileira de Pesquisa Agropecuária

FAPESB - Fundação de Amparo à Pesquisa do Estado da Bahia

FINEP - Financiadora de Estudos e Projetos

fs. - Fentossegundos

KDa – Kilodaltons

MD – Dinâmica Molecular

MM – Mecânica Molecular

QM/MM – Quantum-Mecanical/Molecular-Mecanical

RMS - Rout Mean Square

SEP – Superfície de Energia Potencial

UEFS - Universidade Estadual de Feira de Santana

UESC - Universidade Estadual de Santa Cruz

UNICAMP - Universidade Estadual de Campinas

# SUMÁRIO

ABSTRACT

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

|                |  |    |
|----------------|--|----|
| <b>1</b>       | <b>INTRODUÇÃO</b>  | 16 |
| <b>1.1</b>     | Objetivo geral   | 18 |
| <b>1.1.2</b>   | Objetivos específicos  | 18 |
| <b>2</b>       | <b>REVISÃO DA LITERATURA</b>   | 19 |
| <b>2.1</b>     | A cultura cacauera   | 19 |
| <b>2.1.2</b>   | Vassoura-de-bruxa  | 20 |
| <b>2.2</b>     | Mecânismo de resistência do <i>Theobroma cacao</i> à vassoura-de-bruxa | 23 |
| <b>2.3</b>     | Cisteína protease  | 24 |
| <b>3</b>       | A bioinformática   | 27 |
| <b>3.1</b>     | Métodos em bioinformática  | 28 |
| <b>3.2</b>     | Mecânica molecular   | 32 |
| <b>3.2.1</b>   | Amber  | 33 |
| <b>3.3</b>     | Dinâmica molecular   | 35 |
| <b>3.4</b>     | Métodos quânticos  | 36 |
| <b>3.5</b>     | Método mecânico quântico/mecânica molecular                            | 38 |
| <b>4</b>       | <b>MATERIAIS E MÉTODOS</b>   | 41 |
| <b>4.1</b>     | Identificação, alinhamento e seleção da proteína molde                 | 42 |
| <b>4.1.2</b>   | Determinação da estrutura 3D   | 43 |
| <b>4.1.3</b>   | Construção dos modelos   | 43 |
| <b>4.1.3.4</b> | Construção e refinamento do modelo elucidativo                         | 46 |
| <b>4.1.4.5</b> | Cálculos de QM/MM  | 47 |
| <b>5</b>       | Determinação do modelo mutante   | 50 |
| <b>5.1</b>     | Interação do modelo mutante com íons metálicos                         | 48 |

|              |  |    |
|--------------|--|----|
| <b>6</b>     | <b>RESULTADOS E DISCUSSÕES</b>                 | 51 |
| <b>6.1</b>   | Seleção e determinação dos moldes              | 51 |
| <b>6.1.2</b> | Análises das estruturas alinhadas              | 53 |
| <b>6.2</b>   | Construção, identificação e seleção de modelos | 56 |
| <b>6.3</b>   | Otimização                                     | 64 |
| <b>6.4</b>   | Dinâmica molecular                             | 67 |
| <b>6.5</b>   | Cálculos de QM/MM                              | 76 |
| <b>6.6</b>   | Determinação do modelo mutante                 | 78 |
| <b>6.7</b>   | A interação do modelo mutante com metais       | 86 |
| <b>7</b>     | <b>CONSIDERAÇÕES FINAIS</b>                    | 88 |
| <b>8</b>     | <b>REFERÊNCIAS</b>                             | 90 |



## 1. INTRODUÇÃO

A chamada "bioglobalização" tornou o Brasil altamente vulnerável ao ingresso de novas pragas e doenças. Desde que se instalou na região sul da Bahia em 1989, o *Moniliophthora perniciosa* (Stahel) Singer, fungo causador da vassoura-de-bruxa, é uma das enfermidades que mais afeta a produção do cacaueteiro, reduzindo drasticamente o rendimento das lavouras nas diferentes regiões produtoras do continente americano, sendo, atualmente, considerada a mais importante praga da lavoura cacaueteira da Bahia e da Amazônia brasileira (BASTOS, 1987; PEREIRA, 1990).

Devido ao pouco conhecimento da biologia deste patógeno e dos danos que tem causado às lavouras cacaueteiras em 2000, foi estruturada a Rede de Genômica do Estado da Bahia e implantado o Projeto Genoma do *M. perniciosa* com o objetivo de fornecer dados para o desenvolvimento e controle da vassoura-de-bruxa. Inicialmente formada por quatro instituições (UNICAMP, UESC, CEPLAC e EMBRAPA) e atualmente com a inclusão de outras duas instituições (UEFS e UFBA), formando uma rede com sete laboratórios, dos quais seis estão situados no Estado da Bahia. O Laboratório de Pesquisa em Microbiologia (LAPEM) da UEFS é um dos seis laboratórios desta Rede e foi implantado com recursos do CNPq, FINEP e FAPESB. Mais de 151.000 seqüências já foram obtidas e inúmeros genes de interesse têm sido identificados, sendo, atualmente, o maior banco de dados público de seqüências gênicas de um fungo basidiomiceto (CEPLAC/BRASIL, 2000).

Existem poucos estudos que relatam os mecanismos bioquímicos e fisiológicos de resistência do cacaueteiro (*Theobroma cacao* L.). Dentre estes, destaca-se a atividade de cisteínas proteases durante o período necrofítico da doença e um acentuado desequilíbrio hormonal presente no cacaueteiro (KONO, 2004).

As cisteínas proteases são encontradas em uma larga diversidade de fontes como plantas, animais e microrganismos. Estas possuem um importante papel na degradação de proteínas armazenadas durante a germinação. Em mamíferos, elas atuam na degradação lisossomal de proteínas, na apresentação de antígenos pelas células do sistema imune, no mecanismo de stress oxidativo, apoptosi e invasão de células cancerígenas (SIEGEL, 1993; SÁNCHEZ, 2000).

As proteases representam uma importante classe de alvos moleculares para o desenvolvimento de compostos bioativos (KLEBE, 2000). As proteases tornam-se ativas por catálise e hidrólise das ligações peptídicas. E devido a esta característica elas são classificadas como serina, cisteína, aspártico ou metalo-proteinases, de acordo com o seu mecanismo de ação.

As cisteínas proteases são amplamente utilizadas nas indústrias. Na balança comercial mobilizam cerca de US\$ 1 bilhão de dólares por ano (KLEBE, 2000). Desta forma há uma grande perspectiva biotecnológica na obtenção e determinação estrutural (3D) por modelagem comparativa da cisteína protease do *Theobroma cacao*, pois a predição estrutural e o emprego de estudos de mutagênese dirigido *in silico* poderá propor novas estratégias para o emprego industrial, contribuindo para maior agregação de valor a cultura cacaeira no Brasil.

## 1.1. OBJETIVOS

### 1.1.1. OBJETIVO GERAL

Determinar a estrutura 3D da cisteína protease do *Theobroma cacao* através de Modelagem Comparativa para uma provável aplicação tecnológica. E deste modo:

1. Determinar a estrutura da cisteína protease do *Theobroma cacao L.* por Homologia Comparativa;
2. Propor a estrutura do mutante His/Gly159;
3. Estudar a interação da estrutura do mutante com os íons metálicos:  $Zn^{+2}$ ,  $Cu^{+1}$ ,  $Cu^{+2}$  e  $Cd^{+2}$ .

## 2. REVISÃO DA LITERATURA

### 2.1. A CULTURA CACAUEIRA

O *Theobroma cacao L.*, popularmente chamado de cacaueteiro é uma árvore diplóide da família Malvacea, cujo porte pode atingir até 6 metros de altura. Possui duas fases de produção: temporão (março a agosto) e safra (setembro a fevereiro). A propagação é por sementes, em solo argilo-arenoso. Por ser uma planta umbrófila, vegeta bem em sub-bosques e matas raleadas, sendo uma cultura extremamente conservacionista de solos, fauna e flora (BASTOS, 1987; PEREIRA, 1990).

Historicamente, o cacaueteiro foi inicialmente denominado pelos astecas e maias como cacahuati, na qual suas sementes representavam a matéria prima para produção do chacoatlera (chocolate), uma bebida oferecida aos deuses em atos religiosos e destinada as mais altas castas sociais da época. Com o passar dos anos, o cacau e seus derivados foram rapidamente difundidos entre povos, sendo utilizados como moeda pelos pipiles, povo indígena pré-colombiano de El Salvador, que com ele compravam qualquer tipo de mercadoria e pagavam tributos (ALVARENGA, 1941).

Até o século XVI o cultivo do cacau permaneceu restringido a América Central, e embora outros estudos apontem como origem do cacau as regiões de florestas pluviais da América Tropical (CEPLAC, 2006). A partir do século XVII houve uma ampla expansão comercial do produto no sul da América Central, África, Europa e Ásia. Devido à rápida ascensão comercial dos produtos do cacau com o processo de industrialização da amêndoa do fruto em chocolate em barras. No século XIX houve uma disseminação da cacauicultura para as regiões de florestas tropicais na América Latina e África (ALVARENGA, 1941).

As condições edafoclimáticas do território brasileiro favoreceram o estabelecimento da cultura cacaueteira nas várzeas dos principais rios da região amazônica no estado do Pará. Em 1746, foram plantadas sementes do cacau no município de Canavieiras, no Estado da Bahia, Brasil, sendo as primeiras variedades do tipo forasteiro (RANGEL, 1982). Entretanto, até meados do século XIX, o cultivo do cacau no sul da Bahia apresentava pouco destaque à

economia do nordeste brasileiro, pois a atividade econômica predominante era a cana-de-açúcar em função do elevado preço desse produto no mercado. No entanto a lavoura açucareira frente à pesada concorrência de produtores internacionais sofreu uma grave crise, fato que favoreceu maior remanejamento de capital para o sul da Bahia, estimulando assim a rápida expansão dos cacauais nessa região, levando a nação brasileira a ocupar lugar de destaque na exportação de produtos do cacau em 1890 (MARINHO, 2001).

O desempenho produtivo do Brasil nas exportações de amêndoas do cacau na segunda metade do século XX atingiu o patamar de 3.633 toneladas entre 1976 e 1977. A alta lucratividade caracterizou aquele período da cultura cacauaieira como cultura de “ouro” da região sul do nordeste baiano (MENEZES, 1993). Devido à alta lucratividade, o Governo brasileiro e agências internacionais incentivaram o aumento da produtividade da lavoura cacauaieira através da concessão de linhas de crédito e assistência técnica. No entanto, ao final dos anos 1980, a cacauicultura no Brasil sofreu uma grande crise com a ocorrência da vassoura-de-bruxa no sul da Bahia, causando mudanças político-econômica nacional e internacional, onde as exportações entre o período de 1991-2000 tiveram uma queda substancial, na ordem de US\$ 360 para US\$ 110 milhões fato que conduziu o Brasil de 2º para o 5º maior produtor de cacau no mundo. Além disso, a entrada da vassoura-de-bruxa no sul da Bahia provocou a redução da produção de 350 mil para menos de 100 000 toneladas de cacau, o desaparecimento de mais de 300 mil postos de trabalhos relacionados com a cadeia produtiva do cacau; o êxodo rural, o surgimento de favelas e conseqüente o desmatamento dos fragmentos de Mata Atlântica, antes protegida pelo cultivo do cacauaieiro (DUFUMIER, 2004; CEPLAC, 2006).

### **2.1.2. VASSOURA-DE-BRUXA**

A vassoura-de-bruxa é uma praga causada pelo fungo basidiomiceto *Moniliophthora perniciosa* (Stahel) Singer. A sua fitopatologia abrange todas as partes em crescimento do cacauaieiro, particularmente os brotos mais jovens e os frutos. Esta praga foi descoberta pela

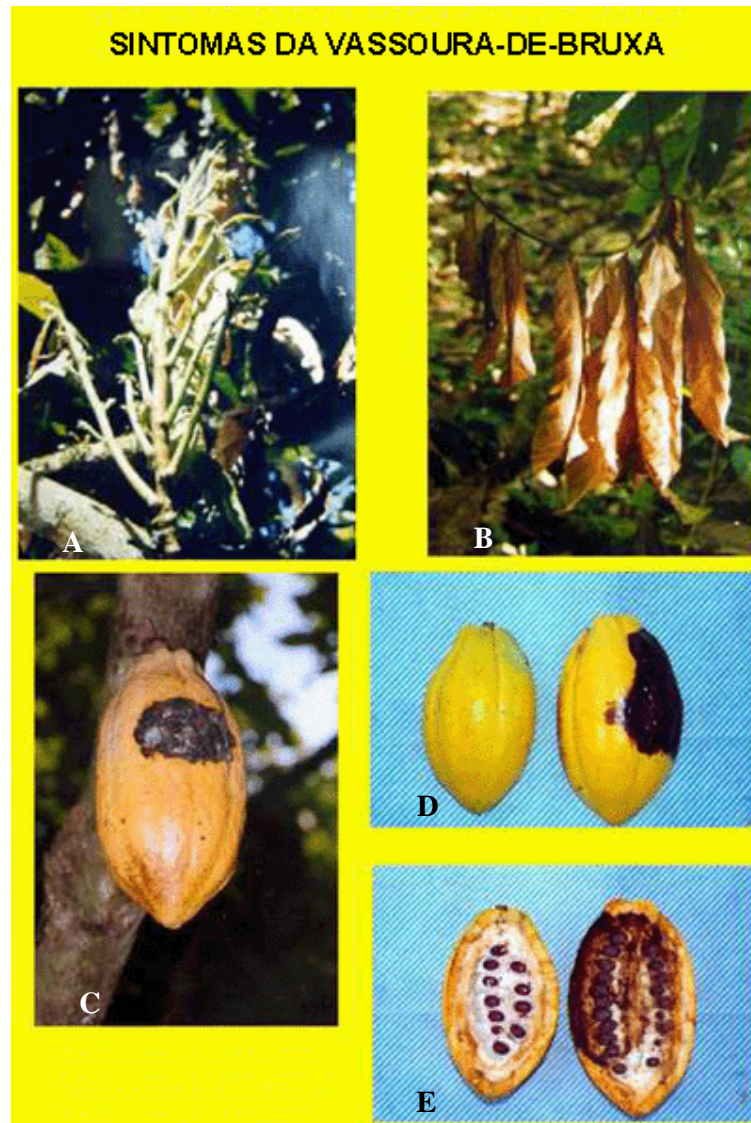
primeira vez no Suriname em 1895, e atualmente ela é responsável pelo grande impacto econômico nos países produtores de cacau como o Brasil, Colômbia, Equador, Granada, Guiana, Panamá, São Vicente, Trindade, Tobago, Ilhas do Caribe e Bolívia. (STAHEL, 1915; WHEELER, 1988; PEREIRA, 1990; WENT, 1904; PURDY, 1996).

A vassoura-de-bruxa foi inicialmente observada no Brasil na região amazônica (BAKER, 1957). Porém a praga rapidamente expandiu-se entre todos os quadrantes cacauero do território brasileiro, até que em maio 1989 o cacau baiano foi contaminado pela vassoura-de-bruxa. O primeiro local de infecção confirmado na Bahia ocorreu no Município de Uruçuca, sendo que a praga foi rapidamente disseminada entre o território baiano devido aos periódicos índices pluviométricos sem períodos de secas bem definidas naquela região (PEREIRA, 1990).

Atualmente, a vassoura-de-bruxa constitui o maior problema fitopatológico de toda a cadeia produtiva de cacau no Brasil, como nos Estados do Pará, Rondônia, Amazonas, Mato Grosso, Acre e especialmente Bahia. O agente etiológico da praga possui um ciclo de vida dividido em duas fases principais, uma parasítica e outra saprofítica. A fase parasítica é constituída pelo micélio monocariótico, sem grampos de conexão, apresentando crescimento intercelular. Seu micélio é mais encorpado que o saprofítico e dicariótico, que ao contrário da fase parasítica, apresenta grampos de conexão e pode crescer tanto inter quanto intracelularmente (EVANS, 1982; WHEELER, 1988; KILARU, 2005; SILVA, 2005). Os basidiocarpos representam a fonte primária de inóculo liberando basidiósporos, que são as principais unidades infectivas do patógeno, os quais se apresentam sob a forma de vassouras secas sobre o fruto infectado (BASTOS, 1986; KILARU, 2005).

A liberação dos basidiósporos ocorre preferencialmente, durante a noite devido à baixa umidade do ar e queda da temperatura (EVANS, 1982; RODRIGUES, 1983; ROCHA, 1985; LAWRENCE, 1991; SOLORZANO, 1977). De início é evidente alterações no aspecto fenotípico da planta como demonstrado na Figura 1 (A e C), o desenvolvimento da doença é rápido, porém depois de 5 a 6 semanas as folhas e frutos começam a secar (B, C, D e E), aparentemente, ocorre a dicarionização do micélio fúngico e a formação de um micélio secundário, dicariótico e fibulado, com hifas mais estreitas que invadem as células dos tecidos do hospedeiro onde se encontram, culminando com a morte dos ramos e pedúnculo (das antigas vassoura verde). As vassouras, agora necróticas, de coloração amarronzada, podem permanecer presas à planta ou podem se destacar e cair no solo (B, D e E). O micélio presente

nessas vassouras produz então basidiomas onde são, por sua vez, gerados os basidiósporos, fechando o ciclo de vida do fungo (PEREIRA, 1990).



**Figura 1.** Comparação entre os galhos e frutos do cacaueiro com a presença dos sintomas da vassoura-de-bruxa. Em A e C está representado o aspecto fenotípico do processo de dicarionização e fibulação do micélio fúngico secundário com a invasão das hifas nos ramos dos tecidos do hospedeiro culminando com a morte dos ramos e necrose do fruto. Em B, C, D e E está mostrado a presença das vassouras necróticas de coloração amarronzada no fruto. Fonte: CEPEC/CEPLAC, 2008.

Devido a este impacto socioeconômico o governo Federal, Estadual, e produtores discutem medidas que possam solucionar ou amenizar o problema causado pelo *M. pernicioso*. Atualmente, uma das recomendações da Comissão Executiva do Plano da Lavoura Cacaueira (CEPLAC), para a reabilitação de plantas suscetíveis à vassoura-de-bruxa, é o uso de variedades clonais resistentes, por meio da enxertia, combinado com aplicação de fungicidas cúpricos, como o óxido cuproso, para o controle da doença (ROSA, 1998; DIAS, 2001; SILVA, 2005). No entanto, esta metodologia não mostra resultados muito satisfatórios, pois estes fungicidas não protegem tecidos em crescimento, necessitando de inúmeras pulverizações (CRONSHAW, 1979; CONRATH, 2002).

Vários compostos químicos vêm sendo testados com o objetivo de prevenir ou erradicar a vassoura-de-bruxa, porém os resultados têm sido insatisfatórios, pois o rápido crescimento da superfície dos frutos durante os dois ou três meses de desenvolvimento faz com que o fungicida tenha que ser aplicado freqüentemente. Contudo, isto é especialmente difícil em árvores muito altas (SOBERANIS, 2000).

## **2.2. MECÂNISMO DE RESISTÊNCIA DO *THEOBROMA CACAO* À VASSOURA-DE-BRUXA**

Existem poucos trabalhos na literatura sobre os mecanismos bioquímicos e fisiológicos de resistência do *T. cacao* a vassoura-de-bruxa. Os primeiros trabalhos tentando elucidar a natureza bioquímica e fisiológica dos distúrbios causados pelo *M. pernicioso* foram realizados na década de 1960. Naquele período foi constatado um desbalanço hormonal no cacaueiro ante a oxidação de auxina (AIA) pela ação de enzimas oxidativas do fungo, sendo uma lacase (polifenoloxidase) e outra uma peroxidase (KRUPASAGAR, 1968; HAMMOND-KOSAK, 2000; ELLIOT, 2003). Este fenômeno favorece a quebra da dominância apical, estímulo das gemas laterais, superbrotamento ocasionado por uma hipertrofia e hiperplasia celular favorecendo o estabelecimento parasítico do fungo com o hospedeiro, dando origem às vassouras (KRUPASAGAR, 1969; CONRATH, 2002; KILARU, 2005).



A indução de resistência sistêmica adquirida em plantas contra fitopatógenos constitui uma alternativa tecnológica bastante promissora. Esta se caracteriza pela ativação dos mecanismos de resistência de caráter multigênico que se encontra em estado latente nas células vegetais capazes de proteger a planta contra vírus, bactérias, fungos e nematóides, durante um longo período, ou talvez, por toda a vida da mesma (ROMEIRO, 1999; DIAS, 2001; KUC, 2001). O mecanismo de ação se processa após a formação de lesões necróticas causada por um fitopatógeno, na qual células vizinhas às lesões necróticas são induzidas no reforçamento das paredes celulares via lignificação e produção de proteínas ricas em hidroxiprolina, como também a síntese de fitoalexinas, ácido salicílico e proteínas relacionadas à patogênese para interrupção do processo infeccioso. (VAN, 1982; RYALS, 1996; SCHNEIDER, 1996).

Estudos realizados na década de 1960 destinaram-se a selecionar compostos fenólicos constitutivos na planta possivelmente relacionados com a resistência de cultivares do cacau à *Ceratocystis fimbriata* (CAPRILES, 1968). Compostos fenólicos, como o ácido clorogênico em plantas resistentes sempre estiveram abaixo dos níveis requeridos para toxicidade *in vitro*. Capriles & Reys (1968) encontraram outros compostos fenólicos em maior nível, como o éster do ácido gentísico, detectado em níveis fungitóxicos para *C. fimbriata* no caule. A resistência do cacau à murcha-de-verticillium também foi estudada como parte do sistema de resistência induzida. Como resultado, quatro compostos pós-infeccionais do tipo fitoalexinas foram isolados do cacau pela primeira vez (CAPRILES, 1968; SCHNEIDER, 1996; MARTÍNEZ, 2003).

### 2.3. CISTEÍNA PROTEASE

A atividade de enzimas representa parte do mecanismo de defesa induzida em resposta à ação parasitária do *M. pernicioso* (SIEGEL, 1993; SÁNCHEZ, 2000). As cisteínas proteases são encontradas em uma larga diversidade de fontes como plantas, animais e microrganismos (RAWLINGS, 2000; MOTTRAM, 2003; BARRETT, 2004). Estas representam um importante papel na degradação de proteínas armazenadas durante a germinação da semente (POMPE-NOVAK, 2002). São relatadas como enzimas primordiais no sistema de defesa das

plantas contra insetos herbívoros e fungos (KONO, 2004). Em mamíferos, estas desempenham uma extraordinária função na degradação lisossomal de proteínas, na apresentação de antígenos pelas células do sistema imune, no mecanismo de estresse oxidativo, apoptosi e invasão de células cancerígenas (MAZAL, 1999). Os mecanismos de ação das proteases se caracterizam conforme seus processos catalíticos, sendo elas classificadas como serina, cisteína, aspártico ou metalo-proteinasas (BARRETT, 1980; MAJERLE, 2003).

As proteases ocupam uma posição principal no que diz respeito a sua aplicação nos campos fisiológicos e comerciais. O valor das vendas mundiais de enzimas industriais é de US\$1 bilhão, sendo que 75% dessas enzimas são hidrolíticas e as proteases representam um dos três maiores grupos correspondendo por aproximadamente 60% da venda total de enzimas no mundo (RAO, 1998; SAID, 2002). As enzimas proteolíticas são de grande valor comercial, e são amplamente utilizadas nas indústrias. Devido às recentes tecnologias e tendências ambientais em desenvolvimento, as proteases estão sendo empregadas em vários processos de biorremediação. Elas possuem uma ampla variedade de aplicações principalmente nas indústrias de alimentos, detergentes e farmacêuticas (RAO, 1998; SAID, 2002).

Na indústria alimentícia o uso de proteases vem desde a antiguidade (WORKMAN, 1986; BARRETT, 2004). Elas são usadas em várias aplicações, como nos processos de fermentação e produção de alimentos orientais, na produção de gelatina hidrolisada e leite de soja. São usadas na clarificação de sucos através da hidrólise das proteínas solúveis neles contidos em altas concentrações, as quais provocam turbidez e formação de sedimentos indesejáveis durante a estocagem (WORKMAN, 1986; DAWES, 1994). Na indústria leiteira são úteis na fabricação de queijos, onde as proteases são empregadas para maturação e desenvolvimento da textura e do sabor (POOLMAN, 1995). Na indústria de panificação, as proteases bacterianas são utilizadas na modificação do glúten através de hidrólise. O glúten é uma proteína insolúvel que está presente na farinha de trigo, e o conseqüente tratamento enzimático da massa facilita a sua manipulação e aumenta o volume dos pães (LYONS, 1988). Proteases alcalinas e neutras de origem de fungos têm um papel importante no processamento de molhos de soja. A modificação proteolítica de proteínas da soja ajuda a melhorar as suas propriedades funcionais. Na indústria de carnes, as proteases são utilizadas para tenderização e condicionamento, também para o melhoramento de alimentos para animais (LYONS, 1988; BILINSKI, 1988).

As proteases são utilizadas na indústria de detergentes, sendo estas enzimas um dos principais ingredientes presentes em todos os tipos de detergentes que são utilizados para lavagem de roupas, pratos, lentes de contato, dentaduras, entre outros (WORKMAN, 1986). Todas as proteases de detergentes empregadas atualmente no mercado são serinas proteases produzidas por *Bacillus sp.* No entanto, as proteases alcalinas de fungo são mais vantajosas devido à facilidade de processamento de uma enzima livre de micróbios (PHADATARE, 1993). Na indústria de couro, o uso de proteases vem como alternativa para a substituição de substâncias químicas antes utilizadas e tem mostrado um melhoramento na qualidade do couro, reduzindo dessa forma, a poluição ambiental (LYONS, 1988; BILINSKI, 1988; PHADATARE, 1993)

### 3. A BIOINFORMÁTICA

A Bioinformática tem suas origens na década de 1960, quando computadores foram empregados como ferramentas importantes para o desenvolvimento da Biologia Molecular. A partir da vinculação direta entre a biologia e a computação, é o estudo, aplicação de técnicas e ferramentas computacionais em problemas de modelados a partir da biologia, realizando análises de dados de cunho biológico. Este surgimento sucedeu-se: i) devido ao crescente número de seqüências protéicas disponíveis; ii) a necessidade de organização dos dados biológicos para melhor compreensão de macromoléculas; iii) pela disponibilidade de computadores mais velozes nas universidades e centros de pesquisas. Dessa forma, a Bioinformática possibilita manusear, analisar e interpretar grandes quantidades de dados biológicos (HAGEN, 2000; OLIVEIRA, 2002; INOUE, 2002).

O final da década de 1960 caracterizou-se pelo desenvolvimento de algoritmos, programas e técnicas computacionais destinadas à análise de estruturas, função e evoluções moleculares, bem como bancos de dados rudimentares de seqüências protéicas (HAGEN, 2000). Com isso no final das décadas seguintes foram desenvolvidos algoritmos para o alinhamento de seqüências (OUZOUNIS, 2002). Técnica importante para a definição da similaridade entre duas ou mais seqüências, comparação esta que pode ser feita de modo global, através de seqüências completa ou local, através de regiões internas de duas seqüências específicas (CARVALHO, 2003). O armazenamento de dados biológicos em computadores possibilitou a criação de banco de dados de acesso ao público, à implementação de sistemas rápidos de busca em bancos de dados, o desenvolvimento de sistemas sofisticados para a predição de estrutura de proteínas, anotação e comparação de genomas e de sistemas para análise funcional de genomas (OUZOUNIS, 2002).

Foi somente na década de 1980 que a Bioinformática pode ser também aplicada na análise de outros tipos de dados como árvores filogenéticas, relações entre vias metabólicas, correlação com dados estatísticos da genética de populações e suas patologias, sendo a primeira vez em que os programas desenvolvidos foram disponibilizados para toda a comunidade científica (OUZOUNIS, 2003; GLAUCIUS, 2008). Ainda nesta década surgiram programas para análise de seqüências nucleotídicas e protéicas, como Staden (STADEN, 1977), Pustell (PUSTELL, 1982) e GCG (DEVEREUX, 1984), assim como bases de dados públicas servindo de repositórios para as seqüências e resultados de análises das mesmas,

como o GenBank (GENBANK, 2009), EMBL (EMBL, 2009), DDBJ (DDBJ, 2009) e Swiss-Prot (SWISS-PROT/UNIPROT, 2009). Com o surgimento dos projetos genoma, transcriptoma e proteoma na década de noventa, a Bioinformática obteve grande destaque industrial com a produção de supercomputadores para o processamento de dados biológicos em larga escala.

A era genômica marcou o século XXI devidos aos avanços na Biologia Molecular, como o seqüenciamento do *Arabidopsis thaliana*, *Drosophila melanogaster* e espécies de interesse agrícola como o arroz, a cana-de-açúcar e o projeto genoma do cacau (CEPLAC 2006). A disponibilidade dos genomas seqüenciados tem provocado um aumento no volume de informação, gerando dificuldades de armazenamento e manipulação, exigindo recursos computacionais cada vez mais eficientes e pessoas cada vez mais capacitadas na análise de dados. Tradicionalmente, o enfoque das pesquisas consistia em conhecer uma determinada função e buscar o gene responsável. Atualmente, a situação é inversa, dispõe-se de um enorme número de genes desconhecidos que demandam caracterização para o conhecimento de biomoléculas e elucidação de estruturas químicas que favorecerá a compreensão molecular de estruturas para elucidação de patologias (SANTOS-FILHO, 2003).

### **3.1. MÉTODOS EM BIOINFORMÁTICA**

A Bioinformática é uma ciência interdisciplinar com aplicações nas mais diversas áreas da Biologia. É uma área do conhecimento baseada no paradigma fundamental da biologia molecular que postula que a informação genética está armazenada nas seqüências de bases de DNA, que são transcritas e traduzidas em seqüências de aminoácidos que constituem as proteínas (SÁNCHEZ, 1998). Estes polímeros de aminoácidos são as principais ferramentas biológicas responsáveis pela manutenção de diversos processos bioquímicos e celulares *in vivo*, portanto, estão envolvidas direta ou indiretamente em todos os processos moleculares existentes dentro e fora da célula. Estudar a função, estrutura e modo de ação das

proteínas em processos celulares são etapas essenciais da biotecnologia (SÁNCHEZ, 1998; KUNDROTAS, 2006).

A Bioinformática usa técnicas computacionais para criar modelos de sistemas químicos reais baseados no domínio estrutural de proteínas. Estes domínios são formados em geral por não mais que 300 aminoácidos. Métodos *ab initio*<sup>1</sup>, enovelamento, e modelagem comparativa (ou modelagem por homologia) concentram-se na análise e comparação de elementos estruturais em proteínas (domínios estruturais) contra seqüências de aminoácidos (HEHRE, 1986; FORESMAN, 1996). Para efeito de maior precisão deste cálculo, seqüências muito longas (com mais de 300 aminoácidos) devem ser subdivididas em fragmentos menores. Alguns métodos de predição *ab initio* utilizam-se de simulações de dinâmica molecular (DM), simulações Monte Carlo, construção de algoritmos genéticos, entre outros (FORESMAN, 1996). O modelo molecular resultante é o produto de simulações parametrizadas e automatizadas como uma tentativa de simular alguns aspectos químicos das propriedades moleculares de uma forma macroscópica, a qual é mais facilmente compreendida e manuseada para a modelagem molecular (FORESMAN, 1996; LINDAHAL, 2001).

A Modelagem Comparativa é uma ferramenta computacional empregada para solucionar problemas de elucidação de seqüências onde a forma tridimensional de moléculas e suas propriedades são importantes. A sua expansão encontra-se intimamente relacionado ao projeto genoma que traz em si inovações tecnológicas que revolucionou a ciência e inúmeros bancos de dados e ferramentas computacionais de acesso público ou privado, os quais têm sido criados na tentativa de organizar e permitir acesso eficiente e rápido às anotações gênicas bem como permitir a análise comparativa dessa quantidade maciça de informações disponíveis em banco de dados de anotações gênicas (Tabela 1) (FORESMAN, 1996; HIGGINS, 2000).

A Modelagem Comparativa de proteínas é uma metodologia empregada para determinação estrutural de seqüências a partir de uma ancestral em comum devido a similaridade protéica entre a função e estrutura. A seqüência protéica é menos conservada do que a estrutura. No entanto, o sucesso desta metodologia depende do quanto à seqüência entre as proteínas são similares entre si (SANTOS-FILHO, 2003).

A realização da modelagem por homologia freqüentemente se processa pela localização inicial de uma proteína conhecida para realização de alinhamento global possível

entre a seqüência desconhecida e o modelo. As sobreposições de ambas as seqüências resultam em um modelo tendo o arcabouço da estrutura determinada experimentalmente com referência, sendo denominada de molde. O processo de modelagem molecular possui também a finalidade de otimizar as posições das cadeias laterais, refinar a geometria obtida e validar o modelo construído (GIBAS, 2001; GOLDSMITH-FISCHMAN, 2003; SANTOS-FILHO, 2003; PATNY, 2006).

**Tabela 1. Lista dos principais bancos de dados disponíveis para análise comparativa de genomas.**

| <b>Banco de Dados</b>   | <b>Descrição</b>  | <b>Referência(s)</b> | <b>URL</b>  |
|---|---|----------------------|---|
| <b>Banco de dados para armazenamento de todas as seqüências gênicas</b> |   |                      |   |
| GenBank   | Banco de dados de seqüências gênicas  | BENSON, 2007.        | <a href="http://www.ncbi.nlm.nih.gov/GenBank/">www.ncbi.nlm.nih.gov/ GenBank/</a> |
| EMBL  | European Molecular Biology Laboratory   | KULIKOVA, 2007.      | <a href="http://www.ebi.ac.uk/embl/">www.ebi.ac.uk/embl/</a>                      |
| DDBJ  | DNA Databank of Japan   | SUGAWARA, 2007.      | <a href="http://www.ddbj.nig.ac.jp/">www.ddbj.nig.ac.jp/</a>                      |
| <b>Banco de dados para realização de busca por similaridade</b>         |   |                      |   |
| BLAST   | O programa BLAST executa buscas por similaridade local de uma ou mais seqüências em bancos de dados de ácidos nucléicos ou proteínas. | ALTSCHUL, 1990.      | <a href="http://www.ncbi.nlm.nih.gov/BLAST/">www.ncbi.nlm.nih.gov/BLAST/</a>      |

<sup>1</sup> O método *ab initio* referido neste capítulo trata-se de um método de predição de estrutura sem ter nenhuma informação prévia experimental. Não está de forma alguma relacionado aos métodos *ab initio* da Química Teórica, na qual será discutido no item 3.2.3 Métodos Quânticos.

---

**Banco de dados para alinhamentos múltiplos de seqüências gênicas**

|           |  |                 |  |
|-----------|--|-----------------|--|
| Clustal W | Programa para alinhamento múltiplo de seqüências nucleotídicas ou protéicas. (global)  | GIBSON, 1996.   | <a href="http://www.ebi.ac.uk/clustalw/">www.ebi.ac.uk/clustalw/</a>                                 |
| T-Coffee  | Programa para alinhamento múltiplo de seqüências nucleotídicas ou protéicas. (global)  | NOTREDAME, 2000 | <a href="http://www.ebi.ac.uk/t-coffee">www.ebi.ac.uk/t-coffee</a>                                   |
| MUMmer    | Sistema para alinhamento múltiplo e visualização de genomas inteiros, ou grandes segmentos de DNA ou proteína, de forma computacionalmente rápida e eficiente.   | KURTZ, 2004.    | <a href="http://www.mummer.sourceforge.net/">www.mummer.sourceforge.net/</a>                         |
| SeaView   | Editor gráfico de alinhamentos múltiplos de seqüências.  | GALTIER, 1996.  | <a href="http://www.pbil.univ-lyon1.fr/software/seaview">www.pbil.univ-lyon1.fr/software/seaview</a> |
| FASTA     | Este programa executa buscas baseadas no algoritmo de Pearson e Lipman, visando a detecção de similaridade local entre uma determinada seqüência e um grupo de seqüências do mesmo tipo (ácidos nucléicos ou proteínas). | ALTSCHUL, 1990. | <a href="http://www.ebi.ac.uk/fasta/">www.ebi.ac.uk/fasta/</a>                                       |
| HMMER     | Programa para buscas por similaridade em bancos de dados que utiliza descrições estatísticas do consenso obtido para uma família de seqüências ( <i>Profile Hidden Markov Models</i> ).                                  | DURBIN, 1998.   | <a href="http://www.hmmer.janelia.org/">www.hmmer.janelia.org/</a>                                   |

**Programas computacionais para realização de Modelagem Molecular por Homologia.**

|             |  |             |  |
|-------------|--|-------------|--|
| Swiss Model | Servidor de modelagem por homologia desenvolvido pelo <i>Swiss Institute of Bioinformatics</i> . | GUEX, 2001. | <a href="http://www.swissmodel.expasy.org//SWISS-MODEL">www.swissmodel.expasy.org//SWISS-MODEL</a> . |
|-------------|--|-------------|--|

---



---

|            |  |                              |  |
|------------|--|------------------------------|--|
| Rasmol     | Servidor para visualização de estruturas.                          | BERNSTEIN, 2000.             | <a href="http://www.umass.edu/microbio/rasmol/">www.umass.edu/microbio/rasmol/</a>                               |
| UniProt    | Universal Protein Resource.  | THE UNIPROT CONSORTIU, 2007. | <a href="http://www.uniprot.org/">www.uniprot.org/</a>   |
| InterProt  | Para análise de domínios de proteínas e regiões funcionais         | MULDER, 2007.                | <a href="http://www.ebi.ac.uk/interpro/">www.ebi.ac.uk/interpro/</a>   |
| SWISS-PROT | Swiss-Prot Database  | BOECKMANN, 2003              | <a href="http://www.ebi.ac.uk/swissprot">www.ebi.ac.uk/swissprot</a>   |
| TrEMBL     | Translated EMBL  | BOECKMANN, 2003              | <a href="http://www.ebi.ac.uk/uniprot/">www.ebi.ac.uk/uniprot/</a>   |
| PROSITE    | Banco de dados de domínios de proteínas, famílias e sítios ativos. | HULO, 2004.                  | <a href="http://www.expasy.ch/prosite/">www.expasy.ch/prosite/</a>   |
| PRINTS     | Banco de dados de domínio de proteínas.                            | ATTWOOD, 1999.               | <a href="http://www.bioinf.manchester.ac.uk/dbbrowser/PRINTS/">www.bioinf.manchester.ac.uk/dbbrowser/PRINTS/</a> |
| PDB        | Protein Data Bank  | BERMAN, 2000.                | <a href="http://www.rcsb.org/pdb/">www.rcsb.org/pdb/</a>   |

---

### 3.2. MECÂNICA MOLECULAR

A modelagem de proteínas por homologia tem sido importante ferramenta em estudos relacionados com os genomas para a elucidação de estruturas protéicas em curto espaço de tempo e a custos reduzidos (SANTOS-FILHO, 2003). Entretanto, durante o desenvolvimento do modelo 3D, algumas “anomalias” estruturais podem ser introduzidas na proteína problema. Desta forma, é imprescindível que o modelo 3D recém construído seja cuidadosamente

refinado, antes de ser utilizado para estudos de interação proteína-receptor. Algumas “incongruências” que podem ocorrer após a construção incluem: substituição de uma cadeia grande por uma pequena ou vice-versa; ligações peptídicas “esticadas” e *loops* com conformações ruins. Entretanto, a maioria destes problemas podem ser resolvidos por meio de protocolos de minimização por mecânica molecular (MM) seguido de simulações de dinâmica molecular (DM) (PATNY, 2006).

A Mecânica Molecular (MM) também chamado de cálculos de campo de força, permitem calcular a energia e otimizar geometrias moleculares, explorar superfícies de potencial (FUJTISU, 2003). Neste método assim como na dinâmica molecular o movimento das estruturas moleculares basea-se em cálculos de energia potencial usando um campo de força e as equações de Newton para o movimento, assumindo cada átomo como uma partícula conforme equação descrita na Figura 2 abaixo (FORESMAN, 1996; FUJTISU, 2003).

$$E_{\text{total}} = \sum E_{\text{ligação}} + \sum E_{\text{angular}} + \sum E_{\text{torcional}} + \sum E_{\text{int. não ligadas}}$$

**Figura 2.** Equação clássica dos métodos empíricos.

Algumas das vantagens da mecânica molecular são a rapidez e a economia de tempo de computação, a facilidade de compreensão em relação aos métodos de mecânica quântica e a disponibilidade de vários algoritmos disponíveis nos softwares HyperChem, Quanta, Sybyl, Alchemy, BioMedCache, Pmodel, Spartan, Titan, Gaussian etc, podendo destacar-se o MMX, MM2, AMBER, MM3, SYBYL, MMFF94, UFF. Sendo cada um desses programas

possuidor de um campo de força específico à geometria de uma molécula especificada em termos de suas coordenadas atômicas (FORESMAN, 1996).

### 3.2.1. AMBER

Existem vários métodos de MM, também chamados de campos de força, os quais diferem-se pela natureza das equações, assim como detalhes das suas parametrizações. O AMBER é um pacote de programas utilizado para conduzir cálculos de Mecânica e Dinâmica Molecular de biomoléculas em determinados campos de força, os quais se diferem pela natureza das equações, assim como detalhes das suas parametrizações. O campo de força representa um ambiente adequado para a determinação estrutural dos potenciais de proteínas juntamente com o conjunto de parâmetros adotados. Deve-se determinar um conjunto de parâmetros que descrevam da melhor forma possível às propriedades e interações químicas de uma molécula, bem como a energia potencial intramolecular. Vários campos de força foram desenvolvidos nas últimas décadas para o estudo de moléculas orgânicas e biomoléculas, como por exemplo, o AMBER, o OPL e o CHARMM (PERLMAN, 1995; MACKERREL, 1995; JORGENSEN, 1998). Entretanto, em algumas situações, o termo AMBER também é empregado para descrever alguns campos de força empíricos, como ocorre no Gaussian 03 (CASE, 2004).

O software AMBER (*Assisted Model Building with Energy Refinement*, Construção Assistida de Modelos com Refinamento de Energia) refere-se a um programa utilizado para conduzir simulações de Mecânica Molecular e Dinâmica Molecular. Este programa está composto basicamente de um banco de dados de estruturas biológicas, um conjunto de campos de força parametrizados representados pelo parm99, usado para o tratamento eletrostático em aminoácidos e algumas moléculas orgânicas; o Gaff, campo de força para moléculas orgânicas; o frcmod, para modificações nos ângulos phi e psi e o all\_amino03 para determinação de carga e *atom types* em proteínas; o ff99 e o ff03 apresentam parametrização

padrão que utiliza campos de força que fixam a carga parcial e são centrados no átomo (CASE, 2004).

O AMBER opera a partir das coordenadas cartesianas dos átomos contidos em um arquivo no formato *Protein Data Bank* (PDB). E, através do programa LEaP é realizada uma análise da proteína inserida no sistema quanto à valência, proximidades entre os átomos, erros estruturais e carga total. Além disso, este programa é responsável pela geração do arquivo topológico e o arquivo de coordenada específico para operacionalização do AMBER (CASE, 2006). Assim, ao submeter à molécula ao AMBER, o programa LEaP prossegue com a “preparação” da molécula corrigindo átomos não descritos anteriormente e neutralizando a carga total da molécula através do potencial Coulombiano que neutraliza a molécula com a adição monoatômica de íons por meio de uma rede de 1 Å, podendo essa rede ser estendida para 4 Å quando se utiliza solvente para os programas de cálculos como o SANDER (CASE, 2004; PEARLMAN, 1995). Em situações cujas informações topológicas de moléculas não constem nos bancos de dados do programa, tais como ligantes, faz-se necessária a utilização do programa *antechamber* que usa algoritmos para classificar os tipos de átomos, tipos de ligações e determina a carga atômica da molécula para deste modo calcular os parâmetros de campo de força (KERRIGAN, 2006).

O SANDER representa a principal plataforma operacional do AMBER, pois neste módulo se conduz os principais parâmetros operacionais para a realização de cálculos de minimização, Dinâmica Molecular e deslocamentos químicos para RMN. Através do algoritmo *Steepest Descent* para cálculo da energia de minimização seguido de cálculos de Gradiente ambos necessários para refinamentos por MM e determinação do mínimo local (CASE, 2006). Assim como, o algoritmo Verlet é empregado para cálculos de trajetória atômica em DM (PEARLMAN, 1995).

Assim, os métodos de MM utilizam algoritmos para otimizar a geometria encontrando o conformero de menor energia na superfície de energia potencial (SEP). Enquanto que cálculos de energia potencial (SEP) asseguram uma análise conformacional da molécula (PATRICK, 2001). Os algoritmos aplicados para otimização de geometria normalmente encontram regiões de mínimo na SEP próximo as coordenadas iniciais (geometria de partida). Na análise conformacional por busca sistemática são feitos incrementos nos valores dos ângulos diedro e de todas as ligações passíveis de rotação para explorar o espaço conformacional da molécula. Desta forma o espaço conformacional de moléculas pode ser varrido (FORESMAN, 1996). Dentre os algoritmos de minimização de energia, os mais frequentemente utilizados são:

*Steepest Descent* e Gradiente Conjugado, ambos utilizam técnicas que calculam a primeira derivada da energia. (LYBRAND, 1990; PATRICK, 2001).

Os cálculos de dinâmica molecular simulam a inserção, colocação da molécula em um banho de temperatura e viscosidades definidas. É dado um intervalo de tempo em que ocorre a transferência de calor entre o banho e a molécula (da ordem de 1 fs) e um intervalo de tempo em que a molécula equilibra a sua energia (da ordem de 300 fs). O processo é repetido o número de vezes especificado. A dinâmica molecular é capaz de ultrapassar pequenas barreiras energéticas e, portanto, é mais eficiente na localização de um mínimo local mais profundo que a minimização simples (CASE, 2004).

### **3.3. DINÂMICA MOLECULAR**

De modo geral, Dinâmica Molecular (DM) consiste em uma simulação do comportamento de várias moléculas ao longo de uma trajetória de pontos representativos do espaço de fase através da solução numérica das equações do movimento (RONCHETTI, 1993). O método da DM, assim como a de Monte Carlo, é observar a evolução do sistema dado através da determinação do movimento das partículas individuais. Devido às interações entre partículas, o sistema é capaz de manter tanto o equilíbrio mecânico quanto térmico, e no caso de perturbações externas o sistema pode atingir uma nova configuração de equilíbrio (RINO, 2001).

Tipicamente as simulações por DM representam um sistema newtoniano que estima o movimento dos átomos ligados entre si considerando as interações electrostáticas e os parâmetros relativos às energias de alongamento, de torção e de dobragem da ligação química. As simulações DM determinam aproximadamente o movimento real do sistema por meio de algum campo de força cuja energia cinética do sistema é derivada da velocidade atômica, essa por sua vez é influenciada pela temperatura da simulação. Desta forma, as simulações DM

tem por objetivo estudar o movimento do sistema, sendo por isso empregados para a validação de modelos 3D (FUJTISU, 2003).

A dinâmica clássica usa campos de forças devidamente parametrizados em um campo de força empírico. Muitos programas computacionais voltados para simulações de DM usam campos de força ligados a métodos de Mecânica Molecular (ALLEN, 1989). O Verlet é o algoritmo mais utilizado para simulações DM (RONCHETTI, 1993). Este representa uma integração numérica da equação Newtoniana do movimento. Esta simulação conserva a energia total e o volume do sistema (FUJTISU, 2003). Neste algoritmo, podemos determinar o coeficiente de difusão, o espectro de estados vibracionais, o fator de estrutura dinâmico. Com respeito à sua termodinâmica, a capacidade térmica, a energia livre, a entropia e pressão, entre outras grandezas, podem ser explicitamente calculadas por meio de uma relação com a velocidade (PEARLMAN, 1995).

### 3.4. MÉTODOS QUÂNTICOS

Em princípio, a Química Quântica deve ser capaz de fornecer descrições quantitativas precisas da estrutura eletrônica da molécula e suas propriedades físico-químicas (ANKERI, 1990). Os métodos quânticos são considerados como uma das maiores realizações intelectuais do século (HEHRE, 1986; FORESMAN, 1996). E são de extrema importância na parametrização dos campos de forças clássicos. A Química Quântica descreve matematicamente uma estrutura molecular conforme o núcleo atômico e distribuição eletrônica em torno dele. Os seus modelos são fundamentados em várias aproximações da equação de Schrödinger,  $H\Psi = E\Psi$ , onde o  $H$  é uma expressão matemática hamiltoniano que engloba a energia cinética dos elétrons e potencial das partículas do sistema, e  $\Psi$  é uma função de onda molecular descrita em termos das coordenadas espaciais das partículas que constituem o sistema em um determinado estado (HEHRE, 1986; FORESMAN, 1996).

Os métodos quânticos dividem-se em *ab initio* e semi-empíricos. Os métodos *ab initio* são úteis na determinação de estruturas complexas em termos das interações entre os núcleos e os elétrons e a geometria molecular em termos da energia mínima do arranjo dos núcleos

(HEHRE, 1986). Este método demanda um altíssimo tempo computacional para simulações realísticas. Dentre os vários métodos *ab initio*, destaca-se a teoria do funcional de densidade (DFT)<sup>2</sup>, desenvolvido por Kohn et al (1977), que consiste na solução numérica aproximada da equação de Schrödinger para N elétrons fazendo uso do fato de que o estado fundamental do sistema é uma função apenas da densidade eletrônica (POPLE, 1970).

O método semi-empírico resulta de uma simplificação na resolução da equação de Schrödinger, na qual algumas integrais são substituídas por parâmetros cujos valores são ajustados para reproduzir propriedades moleculares obtidas por dados experimentais ou calculadas por métodos *ab initio*. Neste método são analisados valores realísticos de propriedades eletrônicas e geométricas de uma molécula e suas interações com o ambiente adquirido por modelos da mecânica quântica em tempos razoáveis de computação. Os métodos semi-empíricos têm características interessantes para o estudo de mecanismos enzimáticos, pois diferente dos métodos de campo empírico, a metodologia semi-empírica pode avaliar a formação e a quebra de ligações covalentes, por considerar explicitamente os elétrons de valência através da resolução da equação de Schrödinger (POPLE, 1970). Devido às dificuldades encontradas na aplicação de métodos *ab initio* para moléculas médias e grandes, vários métodos semi-empíricos foram desenvolvidos. Os métodos semi-empíricos mais comuns são: RM1 (*Recife Model 1*), MNDO (*Modified Neglect Diatomic Overlap*), AM1 (*Austin Model 1*), PM3 (*Parametric Method 3*), PM5 (*Parametric Method 5*) e o MOZYME (FORESMAN; FRISCH, 1996; FREIRE, 2006), implantados sobretudo nos programas MOPAC, AMPAC, HyperChem, Gaussian, BioMedCahe (HEHRE, 1986).

Em comparação com os métodos *ab initio*, o método semi-empírico torna-se mais viável para a realização de cálculo em computadores de baixo custo (POPLE, 1970). Isto porque sistemas moleculares com várias centenas de orbitais, tais como os aminoácidos que compõe o sítio ativo, poderão ser analisados levando-se em consideração apenas os elétrons de valência. (POPLE, 1970; SEGAL, 1977). O objetivo fundamental dos métodos semi-empíricos é o desenvolvimento de um tratamento quantitativo de propriedades moleculares com precisão, confiabilidade e custo computacional suficiente para ser de valor prático em química.

---

<sup>2</sup> A Teoria do Funcional de Densidade não é considerada como método *ab initio* por alguns autores por surgir muito depois do desenvolvimento da Química Quântica.

### 3.5. MÉTODO MECÂNICO QUÂNTICO/MECÂNICA MOLECULAR

A evolução tecnológica promoveu o desenvolvimento de computadores com maior capacidade de armazenamento de dados e processamento, houve o crescimento do tamanho dos sistemas químicos possíveis de serem analisados através de cálculos teóricos. Entretanto, quando as propriedades de interesse dependem de técnicas *ab initio* com a inclusão de correlação eletrônica, fazendo-se uso de funções de base mais extensas e de boa qualidade, o custo computacional pode ser tão dispendioso a ponto de tornar-se inviável. A melhor alternativa para contornar essas limitações é a utilização de métodos híbridos (SVENSSON, 1996; DAPPRICH, 1999).

Uma das estratégias da química computacional para análise de macromoléculas baseia-se no emprego do método híbrido. O método híbrido é uma ferramenta computacional que utiliza a combinação de duas ou mais técnicas computacionais em uma simulação para investigar processos químicos de sistemas grandes, como proteínas. Este método representa uma nova estratégia de cálculos, em decorrência do entrelaçamento entre métodos de Química Quântica (estrutura eletrônica) e Mecânica/Dinâmica Molecular, conhecida por: *Quantum Mechanics/Molecular Mechanics* - QM/MM (DAPPRICH, 1999; VREVEN, 2006;).

A aproximação QM/MM envolve a separação do sistema em regiões QM e MM, O desenvolvimento de algoritmos eficientes tem ajudado na rotina de cálculos mecânico-quântico. Quando um tratamento mais refinado é requerido, a geometria otimizada pela MM pode ser usada como ponto de partida para cálculos quanto-mecânicos de orbitais moleculares. Esta separação dá-se em razão de vários fatores: tamanho dos fragmentos e natureza dos átomos/moléculas presentes no sistema onde existam átomos não parametrizados, enquanto o restante da molécula, constituída por átomos parametrizados e interações consideradas simples, é tratado por métodos rápidos, baseados na mecânica molecular (MASERAS, 1995; WOO, 1998; ATAUALPA, 2006).

O início dos métodos híbridos é estudar a molécula “real” aplicando-se o cálculo de maior nível possível para o sistema modelo (parte ativa), enquanto cálculos menos sofisticados são aplicados no restante da molécula. Existem alternativas de métodos híbridos na literatura (MASERAS, 1995; WOO, 1998; ATAUALPA, 2006; VREVEN, 2006).



Entretanto, a maior distinção entre estes métodos está na forma como a ligação covalente é tratada entre as duas regiões. Uma das formas é a quebra da ligação e conseqüentemente introdução de um átomo de ligação denominado de átomo *link* que pode ser parametrizado para os orbitais atômicos p e d, descrevendo assim uma maior acurácia na mudança de densidade do átomo de ligação (VREVEN, 2006).

O método QM/MM apresenta um sistema onde a MM descreve os efeitos estéricos e eletrostáticos da parte exterior da molécula em um nível intermediário de M que se representa sob os efeitos eletrônicos de grupos funcionais ou ligantes próximos ao centro. A interação eletrostática é avaliada, como uma interação de cargas parciais M com cargas parciais atribuídas a átomos na região QM. Desta forma consegue-se uma maior reprodutibilidade do comportamento eletrostático da camada MM e obtenção da correlação eletrônica sobre o centro de ação mais importante do sistema. (VREVEN, 2006).

A análise das interações eletrostáticas por QM/MM geralmente pode ser ilustrada como uma superposição de camadas de uma “cebola”, método que ficou conhecido como ONIOM (*Our own n-layered integrated molecular orbital and molecular mechanics*) (SVENSSON, 1996). Segundo o método ONIOM, qualquer sistema molecular pode ser dividido em diferentes níveis ligados segundo uma ordem mais conveniente ao problema em questão. Cada nível pode ser tratado sobre qualquer método e ao integrar-se os resultados obtidos sobre os níveis, produz-se uma extrapolação atingindo valores de energia mais precisos sobre todo o sistema (DAPPRICH, 1999; VREVEN, 2006). Em um cálculo realizado pelo ONIOM. A energia para separação do sistema em três partes é dada segundo o esquema da Figura 3 em duas camadas (QM/MM), a energia total do sistema é obtida por meio de três cálculos independentes.

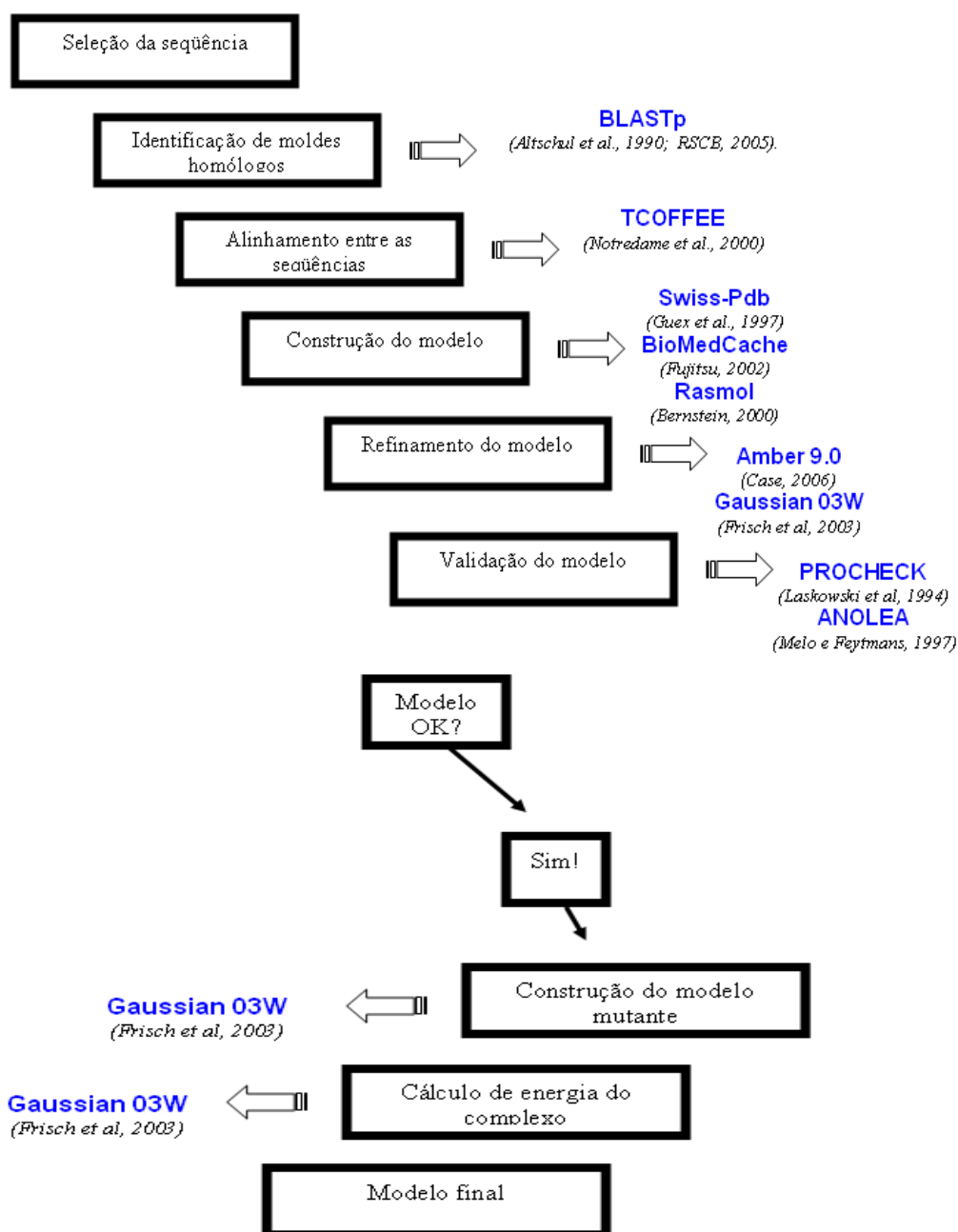
$$E^{\text{ONIOM}} = E^{\text{real,MM}} + E^{\text{modelo,QM}} - E^{\text{MODELO,MM}}$$

**Figura 3.** Cálculo da energia total resultante pela aplicação do método ONIOM.

Vários estudos vêm sendo feitos sobre a confiabilidade e aplicabilidade do método híbrido (MASERAS,1995; MATSUBBARA, 1996; DAPPRICH, 1999; KARADAKOV, 2000). As principais motivações destes estudos estão na escolha do sistema modelo para a região de catálise de enzimas que representem além da geometria, fatores como energia livre dos reagentes e estado de transição. Atualmente o método mais eficiente para a representatividade desses fatores teve sua determinação feita de maneira mais eficiente com a utilização do método QM/MM (HENSEN, 2004).

#### 4. MATERIAIS E MÉTODOS

Para completa elucidação da estrutura 3D da cisteína protease do *T. cacao* foram empregados alguns métodos computacionais para realização da Modelagem Comparativa conforme a Figura 4.



**Figura 4.** Etapas da Modelagem Comparativa para a construção do modelo 3D da cisteína protease do *T. cacao*. Em azul encontram-se os programas utilizados em cada etapa.

#### 4.1. IDENTIFICAÇÃO, ALINHAMENTO E SELEÇÃO DA PROTEÍNA MOLDE

A seqüência primária de aminoácidos da cisteína protease do *Theobroma cacao* foi obtida em colaboração com a Universidade Estadual de Santa Cruz - Laboratório de Genômica e Expressão Gênica (Figura 5), sendo gentilmente cedida pela Profª. Fabienne Florence Lucienne Micheli. A referida seqüência de aminoácidos foi submetida ao *Basic Local Alignment and Search Tool* (BLAST), empregando o algoritmo de busca *Programs Search Protein Databases Using a Protein Query* - BLASTp, uma ferramenta que realiza uma busca contra as proteínas presentes no *Protein Data Bank* (PDB) (ALTSCHUL, 1990; RSCB, 2005). Dessa forma foram selecionadas proteínas conhecidas estruturalmente como moldes com similaridade estrutural igual ou superior a 25% (GOLDSMITH-FISCHMAN, 2003). As proteínas que apresentaram o maior grau de homologia possível foram submetidas ao *Swiss Institute of Bioinformatics* (SIB), um servidor de modelagem por homologia automática (GUEX, 1997).

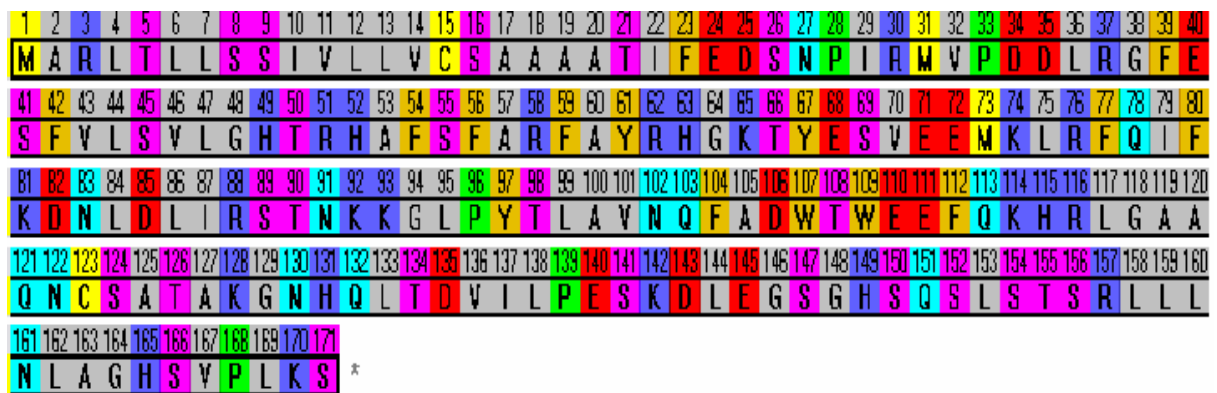


Figura 5. Seqüência de aminoácidos da cisteína protease do *T. cacao*.

#### 4.1.2. DETERMINAÇÃO DA ESTRUTURA 3D

Duas metodologias foram utilizadas para a construção e determinação estrutural da cisteína protease do *T. cacao* por Modelagem Comparativa. Os métodos computacionais de cálculos Quantum Mecânicos (QM) e Mecânica Molecular (MM) foram empregados para construção do modelo (FRISCH, 2003; CASE, 2006). Os procedimentos iniciais consistiram na identificação e seleção das proteínas moldes, seguido do alinhamento entre as seqüências de resíduos de aminoácidos; refinamento; construção do modelo e validação do modelo (GOLDSMITH-FISCHMAN, 2003; SANTOS-FILHO, 2003).

Inicialmente, o modelo foi construído utilizando o software SWISS MODEL (SM) (GUEX, 1997). O SM usa um processo automatizado de modelagem de proteínas entre o modelo experimental e seus respectivos moldes. Igualmente foi utilizado o processo automatizado de modelagem de proteínas por homologia oferecido pelo software Threading (PACHECO, 2000). No final de todo processo, vários modelos foram comparados entre si conforme uma validação inicial realizada no programa *Atomic Non-Local Environment Assessment* (ANOLEA) (MELO, 1997).

Foram selecionados três modelos estruturais cujos cálculos de mecânica e dinâmica molecular, e subsequente validação foram realizadas nos programas Amber 9.0 e Procheck (LASKOWSKI, 1994; CASE, 2006) respectivamente. Estes softwares encontram-se implementados no servidor Intel Xeon Dual-Core presente no Laboratório de Modelagem Molecular da Universidade Estadual de Feira de Santana.

#### 4.1.3. CONSTRUÇÃO DO MODELOS

A partir da identificação, alinhamento e seleção da proteína molde foram construídos três modelos utilizando ferramentas de Modelagem Comparativa constituintes dos programas Swiss Model e BioMedCache 6.1 (GUEX, 1997; FUJITSU, 2005). As regiões

estruturalmente conservadas e regiões variáveis foram localizadas por intermédio dos algoritmos disposto no programa Tcoffee (NOTREDAME, 2000).

Os modelos desenvolvidos foram identificados pelas siglas dv1, dv2 e dv3 respectivamente. O primeiro modelo (dv1) refere-se ao alinhamento nativo entre os resíduos de aminoácidos (Figura 6) (GUEX, 1997). Para a construção dos demais modelos foram aplicadas algumas modificações entre o alinhamento dos resíduos de aminoácidos da cisteína protease do *T. cacao*. A primeira alteração consistiu na subtração de *gaps* (dv2) (Figura 7); na segunda foram incluídos aminoácidos correspondentes aos *gaps* existentes (Figura 8). Concomitantemente, foi utilizada a metodologia de corpos rígidos para construção dos modelos estruturais (3D) da proteína problema (GUEX, 1997; SANTOS-FILHO, 2003).



Figura 6. Sequência de aminoácidos da cisteína protease com a permanência de *gaps* no modelo dv1.

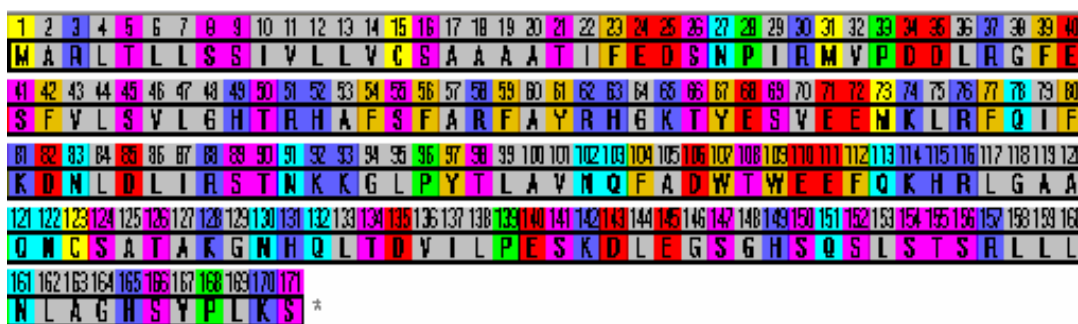


Figura 7. Sequência de aminoácidos da cisteína protease com a deleção de *gaps* no modelo dv2.

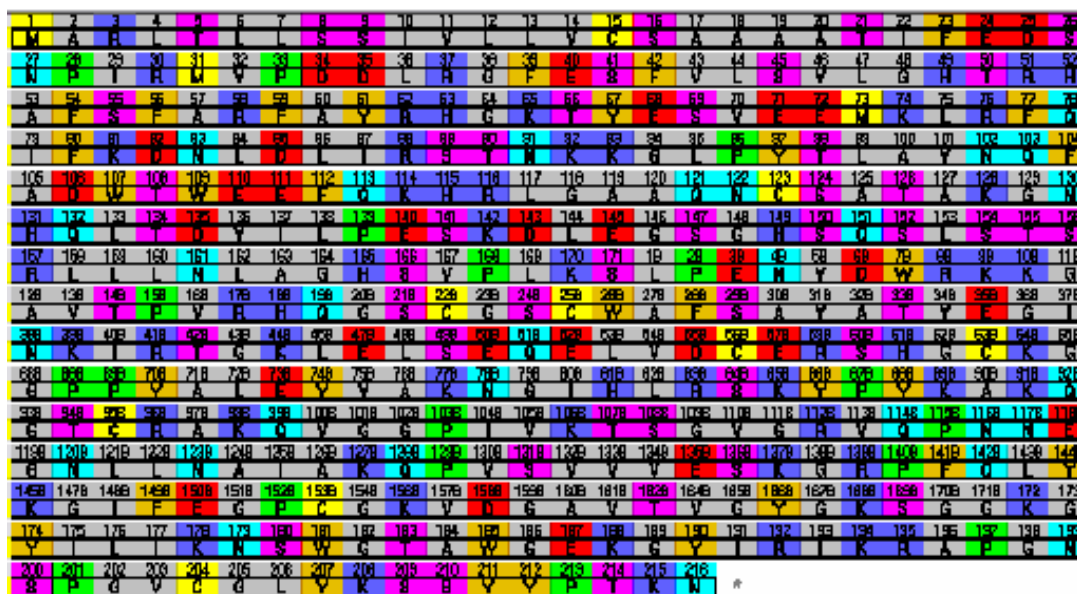


Figura 8. Sequência primária da cisteína protease com a inserção de aminoácidos aos gaps no modelo dv3.

Os modelos gerados foram inicialmente submetidos ao programa AMBER 9.0 para uma otimização (CASE, 2006). Na primeira etapa, a molécula foi preparada usando-se o programa tLeap por meio do campo de força ff03. Através deste procedimento foram gerados os arquivos topológicos e de coordenadas da proteína. Posteriormente os modelos foram neutralizados com íons de Cl<sup>-</sup> e a estrutura resultante foi utilizada como input em todos os cálculos de otimização. Esta etapa foi realizada no programa SANDER, sendo utilizadas as seguintes palavras-chaves: imin=1, realiza uma otimização da estrutura; maxcyc=1000, número máximo de ciclos; nyc=500, número de ciclos necessários para a mudança de *Steepest Descent* para o Gradiente Conjugado; cut=14, valor de cutoff para a especificação da distância das interações de Van der Walls para átomos não ligados; ntb=0, manter volume constante; igb=1, incluir modelo de solvatação (PEARLMAN, 1995; CASE, 2006).

Vinte cálculos foram realizados para determinar o melhor valor de cutoff a ser empregado para a otimização da cisteína protease do *T. cacao*. Os resultados indicaram que

após a otimização com o valor de cutoff de 14Å ocorreu uma queda brusca na energia de formação da molécula e o valor de energia permaneceu constante não havendo alteração na geometria da proteína. Entretanto, constatamos que a variação no tempo computacional (de 5000 para 10000 maxcyc ) e/ou aumento da temperatura do sistema (de temp0=300.0 para 600.0Kelvins) a qualidade esteroquímica avaliada no gráfico de Ramachandran decaiu significativamente com o aumento do tempo computacional, assim como o aumento da temperatura proveu a desnaturação da estrutura molecular. Com base nestes aspectos pode-se inferir que o melhor parâmetro de otimização molecular ocorreu a 5000 maxcyc devido a baixa energia de formação quando é utilizado este valor para sua otimização.

A qualidade esteroquímica da cadeia principal e a energia de interação de cada átomo da cadeia lateral dos modelos foram analisadas, calculadas e avaliadas respectivamente conforme os comprimentos de ligação, ângulos de ligação, ligação peptídica, planaridade de anéis, quiralidade, ângulos torsionais da cadeia principal e lateral, contato estérico entre átomos não ligados através dos dados e gráficos gerados pelo programa PROCHECK e ANOLEA (LASKOWSKI, 1994; MELO, 1997).

No programa Gaussian 03W, foi empregado o campo de força UFF utilizando as seguintes palavras-chave: opt (requere otimização de geometria), UFF (especificação do campo de força a ser utilizado) e geom=connectivity (determina a ligação entre os átomos) (FRISCH, 2003; CASE, 2006;). Nesta metodologia, o algoritmo de Berny foi utilizado durante o processo de otimização, sendo o valor de cutoff determinado pelo *default* do programa (FRISCH, 2003).

#### 4.1.3.4. CONSTRUÇÃO E REFINAMENTO DO MODELO ELUCIDATIVO

A construção do modelo 3D da cisteína protease do *T. cacao* foi desenvolvida no programa Amber 9.0 de acordo com os parâmetros metodológicos de Case et al 2006. Em seguida o modelo foi sucessivamente refinado através de ciclos de otimização *Steepest Descent*, Gradiente Conjugado, seguido de simulações de Dinâmica Molecular (DM) através das palavras-chave: imin=0, realiza uma dinâmica molecular da estrutura; nstlim=1000000,



número máximo de ciclos;  $dt=0.001$ , determina o número de passos do procedimento;  $ntc=1$ , interação molecular padrão para átomos de hidrogênio;  $ntr=100$ , energia de informação;  $ntwx=100$ , define os intervalos de velocidade existente na coordenada do arquivo;  $cut=14$ , valor de cutoff para a especificação da distância das interações de Van der Walls para átomos não ligados;  $ntb=0$ , manter volume constante;  $igb=1$ , incluir modelo de solvatação;  $ntt=3$ , temperatura padrão do sistema;  $gamma\_ln=1.0$ , eficiente conformacional;  $tempi=0.0$ , controle de temperatura;  $temp0=300.0$ , definição da temperatura máxima do sistema (PEARLMAN, 1995; CASE, 2006). Todas as simulações foram executadas a 1 nanosegundo. Estas simulações permitem estudar o movimento atômico, observar as estruturas de menor energia na superfície de energia potencial (SEP), e também avaliar o comportamento termodinâmico da proteína (ŠALI, 1993; FUJITSU, 2005). O modelo resultante foi analisado sobre a qualidade do efeito estérico dos resíduos de aminoácidos da cadeia primária da proteína através do gráfico de Ramachandran (LASKOWSKI, 1994).

#### 4.1.4.5. CÁLCULOS DE QM/MM

Para a localização e determinação do sítio ativo do modelo da cisteína protease do *T. cacao* foi realizado um alinhamento entre as seqüências de aminoácidos no programa SWISS MODEL (GUEX, 1997). Com os aminoácidos que fazem parte do sítio ativo da proteína molde e a distância em Å existente entre eles foram utilizadas como referência na proteína problema para localização dos resíduos que compõe o sítio ativo da enzima. Os resíduos constituintes da região do sítio ativo da molécula foram submetidos a cálculos de Mecânica Quântica (QM) e Mecânica Molecular (MM) no programa Gaussian 3.0W, campo de força UFF,  $geom=connectivity$  através do método de DFT (B3LYP-LanL2DZ) para alta camada, e o campo de força UFF para as demais partes da molécula (HEAD, 1988; FRISCH, 2003; SAKIYAMA, 2008).

A partir da caracterização do sítio ativo do modelo da cisteína protease do *T. cacao*, obtido através do alinhamento estrutural entre o modelo e 1PCI|A, realizado no programa SWISS MODEL (GUEX, 1997), foi possível determinar quais seriam as regiões a serem

tratadas no formalismo mecânico-quântico (QM) e mecânica molecular (MM). Assim, os resíduos constituintes da região do sítio ativo da molécula foram definidos na região Mecânica Quântica (QM) e o “resto” da proteína no formalismo Mecânica Molecular (MM), construindo o método híbrido DFT (B3LYP-LanL2DZ)/UFF no programa Gaussian 03W (HEAD, 1988; FRISCH, 2003; SAKIYAMA, 2008).

## 5. DETERMINAÇÃO DO MODELO MUTANTE

O modelo mutante da cisteína protease do *T. cacao* foi construído através do método de substituição de resíduo de aminoácido constituinte da região do sítio ativo da enzima. O resíduo de histidina 159 foi substituído por um resíduo de glicina. Este procedimento foi realizado de acordo com a metodologia do programa Swiss Model (GUEX, 1997). Em seguida o modelo mutante foi submetido à otimização molecular, dinâmica molecular e cálculos de QM/MM e validação respectivamente nos programas Amber 9.0, Gaussian 03W e Procheck (LASKOWSKI, 1994; CASE, 2006; FRISCH, 2003) de acordo com os parâmetros experimentais anteriormente empregados na construção e refinamento do modelo inicial, localização e determinação do sítio ativo da enzima.

### 5.1. INTERAÇÃO DO MODELO MUTANTE COM ÍONS METÁLICOS

A interação entre o modelo mutante e íons metálicos foi estudada por métodos QM/MM. O hidrogênio (H244) ligado ao átomo de enxofre (S243) do resíduo Cys25 foi substituído pelos íons  $Zn^{+2}$ ,  $Cu^{+1}$ ,  $Cu^{+2}$  e  $Cd^{+2}$  através do método ONIOM implementado no Gaussian 03W. Para avaliação da atividade da interação química foi empregado estudos QM/MM de acordo com os parâmetros metodológicos estabelecidos no programa Gaussian 03W (FRISCH, 2003).

## 6. RESULTADOS E DISCUSSÕES

### 6.1. SELEÇÃO E DETERMINAÇÃO DOS MOLDES

A ferramenta mais bem sucedida de predição de estruturas tridimensionais de proteínas é a modelagem por homologia. Esta abordagem baseia-se no princípio de que homologia entre seqüências de aminoácidos implica em semelhança estrutural e funcional; proteínas homólogas apresentam regiões internas conservadas (principalmente constituídas de elementos de estrutura secundária:  $\alpha$ -hélices e folhas- $\beta$ ). As principais diferenças estruturais entre proteínas homólogas ocorrem nas regiões externas, constituídas principalmente por alças (*loops*), que ligam os elementos de estruturas secundárias (BRANDEN, 1991; ALENCASTRO, 2003).

A seleção e determinação do molde para modelagem comparativa constituem uma etapa crucial para a construção de uma nova estrutura. Ao submeter à seqüência primária da cisteína protease do *T. cacao* ao servidor BLASTp foi possível identificar 12 (doze) seqüências com relativo grau de homologia (Figura 10). No entanto, somente 08 (oito) seqüências apresentaram grau de homologia superior a 30% (Tabela 2) (ALTSCHUL, 1990; RSCB, 2005).



**Figura 9.** Resultado do alinhamento do BLASTp da seqüência de aminoácidos da cisteína protease do *T. cacao* com seus respectivos moldes.

A partir da seqüência de aminoácidos é possível identificar proteínas com domínios estruturais homólogos aos da proteína em estudo. A existência de semelhanças entre as seqüências de aminoácidos permite prever, com alguma confiança, o melhor modelo estrutural da proteína em estudo (BERMAN, 2000). De acordo com este princípio as estruturas molde fornecidas pelo servidor BLAST foram analisadas quanto ao grau de identidade apresentado, número de *gaps* e *e-value*. Analisando-se o arquivo pdb da *Carica papaya* (pdb 1PCI|A) foi constatado que a mesma apresenta razoável grau de identidade (36%), melhor *score* (64.3), *e-value* (2e-11), e reduzido número de *gaps* (3%) (Tabela 2; Figura 10). Semelhança entre as seqüências de aminoácidos em proteínas homólogas se expressa pelo grau (percentual) de identidade igual ou superior a 25%. Adicionalmente, tanto a estrutura tridimensional como a função se conservam durante o processo evolutivo (SANDER, 1991; HÖLTJE, 1997).

Os graus de identidade obtidos nas demais seqüências do arquivo pdb foram próximos aos valores encontrados na estrutura 1PCI|A. No entanto, de acordo com Pearson (1998), quanto maior o valor do *score* melhor é o alinhamento e quanto menor o *e-value* (< 0.02) mais semelhantes são as duas seqüências alinhadas. Desse modo existe grande probabilidade destas identidades terem sido alinhadas ao acaso resultando em seqüências não homólogas, fato correspondente com o baixo percentual do alinhamento entre as cadeias de aminoácidos e variável percentual de *gaps*. Como resultado, a 1PCI|A foi selecionada para a construção do modelo da cisteína protease do *T. cacao*.

**Tabela 2.** Resultado do alinhamento do BLASTp da seqüência primária da cisteína protease do *T. cacao* com seus respectivos moldes.

| Estrutura (PDB)                                  | Identidade | Gaps | Score | E-value |
|--|------------|------|-------|---------|
| <a href="#">1PCI A</a> (Pickersgill et al. 1991) | 36%        | 3%   | 64.3  | 2e-11   |
| <a href="#">1DQKIN</a> (Aleshin et al 1998).     | 35%        | 8%   | 30.4  | 0.24    |
| <a href="#">1HKCIA</a> (Aleshin et al 1996).     | 35%        | 8%   | 30.4  | 0.24    |
| <a href="#">1HKBIA</a> (Aleshin et al 1996).     | 35%        | 8%   | 30.4  | 0.24    |
| <a href="#">1CZAIN</a> (Aleshin et al 1998).     | 35%        | 8%   | 30.4  | 0.24    |
| <a href="#">1CJLJ</a> (Coulombe et al 1996).     | 30%        | 8%   | 28.1  | 1.2     |
| <a href="#">1CSSIA</a> (Coulombe et al 1996).    | 30%        | 8%   | 28.1  | 1.2     |
| <a href="#">2NUP B</a> (Mancias et al 2007).     | 34%        | 2%   | 26.6  | 3.6     |



**Figura 10.** Modelo estrutural da 1PCI|A.

### 6.1.2. ANÁLISES DAS ESTRUTURAS ALINHADAS

A 1PCI|A é uma caricaína, previamente conhecida como protease Omega oriunda do látex da *Carica papaya* (E.C. 3.4.22.6), constituída primariamente por 216 resíduos de aminoácidos (24 kDa) (PICKERSGILL, 1991; GROVES, 1996; BARRETT, 2004). Analisando o arquivo pdb da 1PCI|A foi constatado que a mesma apresenta-se sobre a forma trimérica representada pelas cadeias A, B e C constituída por 7341 átomos, 7518 pontes de hidrogênio e 933 resíduos de aminoácidos distribuídos entre 30  $\alpha$ -hélices e 48 folhas  $\beta$  (Figura 10) (GROVES, 1996).

Proteínas triméricas caracterizam-se pela similaridade monomérica (PICKERSGILL, 1991). Foi constatado que cada monômero da 1PCI|A apresentam uma seqüência complexada de 106 resíduos associados aos 216 aminoácidos da cadeia principal B e C. Portanto, sucessivos alinhamentos foram realizados entre as seqüências primárias da cisteína do *T. cacao* e a seqüência do molde, assim como com a seqüência complexada. Como pode ser visto nas Figuras 11 e 12, o alinhamento da seqüência primária da 1PCI|A apresentou um *score* (51) menor que a média obtida entre o alinhamento da seqüência de 106 resíduos de aminoácidos (*score*: 65) respectivamente. A partir de uma análise quanti/qualitativa, foi constatado que 45 resíduos dos 216 aminoácidos da seqüência primária do 1PCI|A alinharam-se com a seqüência primária da cisteína protease do *T. cacao* conforme resultado do programa TCOFFEE um melhor alinhamento local (45 alinhamentos classificados como *good*), contra 26 resíduos alinhados com a seqüência excedente.

```

SCORE=51
*
  BAD AVG GOOD
*
CISTPROT : 51
IPCIA    : 51
CISTPROT M-----ARLTLSSIVLLVCSAAAATIFEDSNPIR-----
IPCIA    LPENVDRKKGAVTPVRHOGSCGSCWAFSAVATV-EGINKIRTGKLVELSE
cons     : : * : * : * : * : * : *
CISTPROT --MVPDDL-----CFESFVLSVLGHTRHAFSFARFAYRHGKTYESVEEM
IPCIA    QELVDCERRSHGCKGYPYALEYVARNC---IHLRSKYPYKAKOGTCRAK
cons     : * : * : * : * : * : * : * : * : * : *
CISTPROT KLRFOIFKDN-LDLIRSTNKKGLPYTLA---VNOFADWTWEEFOKHRLGAA
IPCIA    OVGGPVKTSGVGRVOPNNEGMLLNALAKQPWSVVVESKGRPFOLYKGGIF
cons     : : * * : : : * : * : * : * : * : * : *
CISTPROT ON-CSATAKG-----NHOLDVILPESKDLEGSCHSOSLSTSRLLNL
IPCIA    EGPCGTRKVDCAVAVGYGKSCGKGYILIKNSWGTAWGKGYIRIKRAPGNS
cons     : : * : * : * : * : * : * : * : * : *
CISTPROT AGHSVPLKS-----
IPCIA    PGVCGLYKSSYYPTKN
cons     : * : *

```

Figura 11. Alinhamento entre a seqüência primária da proteína IPCIA dissociada da molécula complexada.

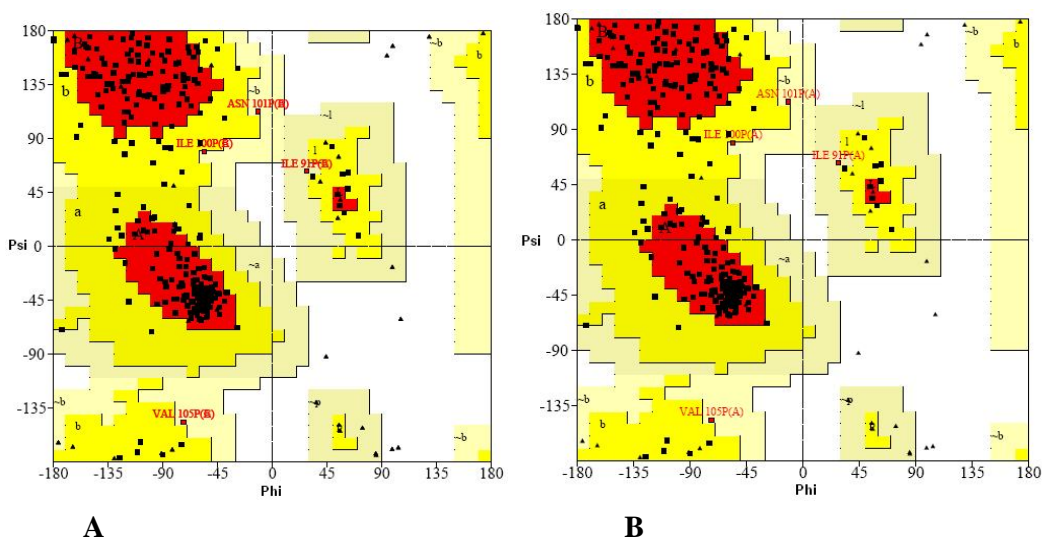
```

SCORE=65
*
  BAD AVG GOOD
*
CISTPROT : 65
IPCIA    : 65
CISTPROT MARLTLSSIVLLVCSAAAATIFEDSNPIRMVPDDLRFESFVLSVLGHTR
IPCIA    -----DFSIVGYSODDLTSTERLIOL-----
cons     : * : * : * : * : * : * : * : * : *
CISTPROT HAFSFARFAYRHGKTYESVEEMKLRFOIFKDNLDLIRSTNKKGLPYTLAVN
IPCIA    ----FNSWMLNHNKIFYENVDEKLYRFEIFKDNLNIDETNKKNSYWLGLN
cons     : * : * : * : * : * : * : * : * : * : * : *
CISTPROT OFADWTWEEFOKHRLGAAQNCSATAKGMHQLTDVILPESKDLEGSCHSQSL
IPCIA    EFADLSNDEFNEKYVGS-----LIDATIEOSYDEEFINED---
cons     : * * : : * * : : * : * : * : * : *
CISTPROT STSRLLNLNLAGHSVPLKS
IPCIA    -----IVN-----
cons     : : *

```

Figura 12. O alinhamento da molécula complexante demonstraram melhor disposição entre os resíduos de aminoácidos da seqüência primária do molde e a seqüência da cisteína protease de *T. cacao*.

Gráfico de Ramachandran é o melhor indicativo da qualidade estereoquímica do modelo de uma proteína (LASKOWSKI, 1993). Assim sendo, um estudo comparativo entre os 322 e 216 resíduos de aminoácidos da 1PCI|A foi realizado como forma de visualizar ângulos dihedrais  $\phi$  contra  $\psi$  de sua estrutura protéica. Análises da qualidade estrutural entre as seqüências estudadas apresentaram equivalentes indicativos da posição estereoquímica dos resíduos não revelaram diferença significativa de qualidade estereoquímica entre os aminoácidos avaliados. Ambos os resíduos apresentaram 98,4% de aminoácidos em regiões energeticamente favoráveis, sendo que 78% dos resíduos encontram-se distribuídos em regiões favoráveis, 20,2% em regiões permitidas e 1,5% em regiões aceitáveis e 0,0% em regiões desfavoráveis (Figura 13).



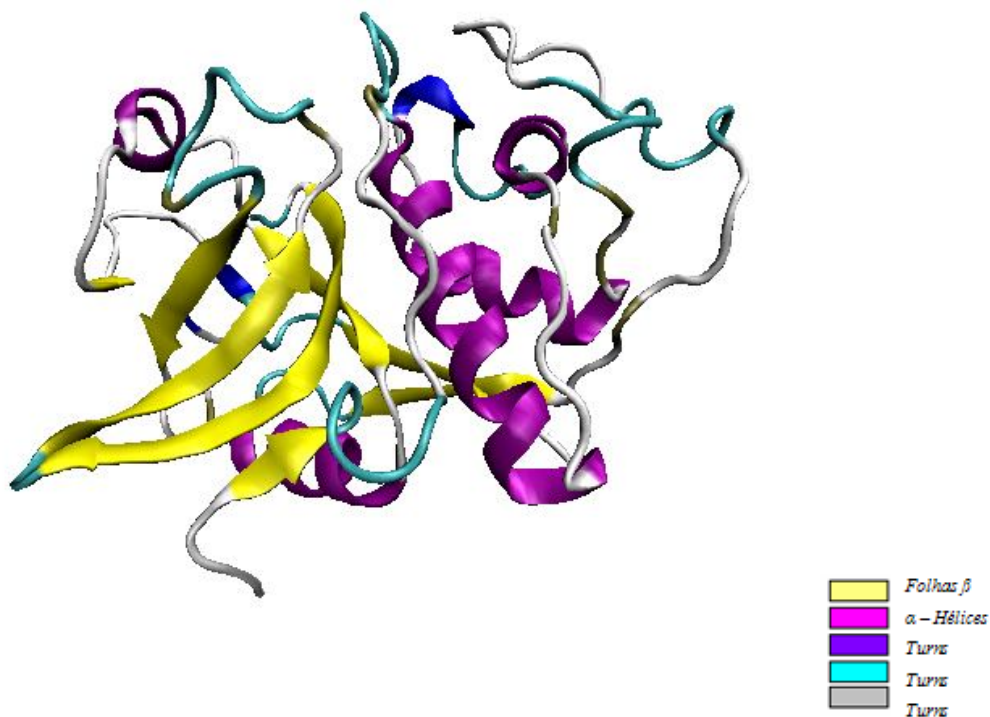
**Figura 13.** Validação representada pelo gráfico de Ramachandran para os 322 (A) e 216 (B) resíduos de aminoácidos da 1PCI|A. As regiões marcadas em vermelho e amarelo representam regiões favoráveis, permitidas e aceitáveis, respectivamente e as regiões em branco como área desfavorável.



Similaridade de seqüência primária indica um processo conservativo durante a evolução e um papel fisiológico importante na atividade enzimática (GROVES, 1996; BARRETT, 2004). O mesmo tipo de análises foram realizadas nos demais moldes alinhados com a nossa seqüência, contudo não foi constatado resultados significativos que alterassem os dados obtidos na Tabela 2. Dessa forma, a cadeia primária da enzima foi selecionada para realização da Modelagem Molecular da cisteína protease do *T. cacao*.

## **6.2. CONSTRUÇÃO, IDENTIFICAÇÃO E SELEÇÃO DE MODELOS**

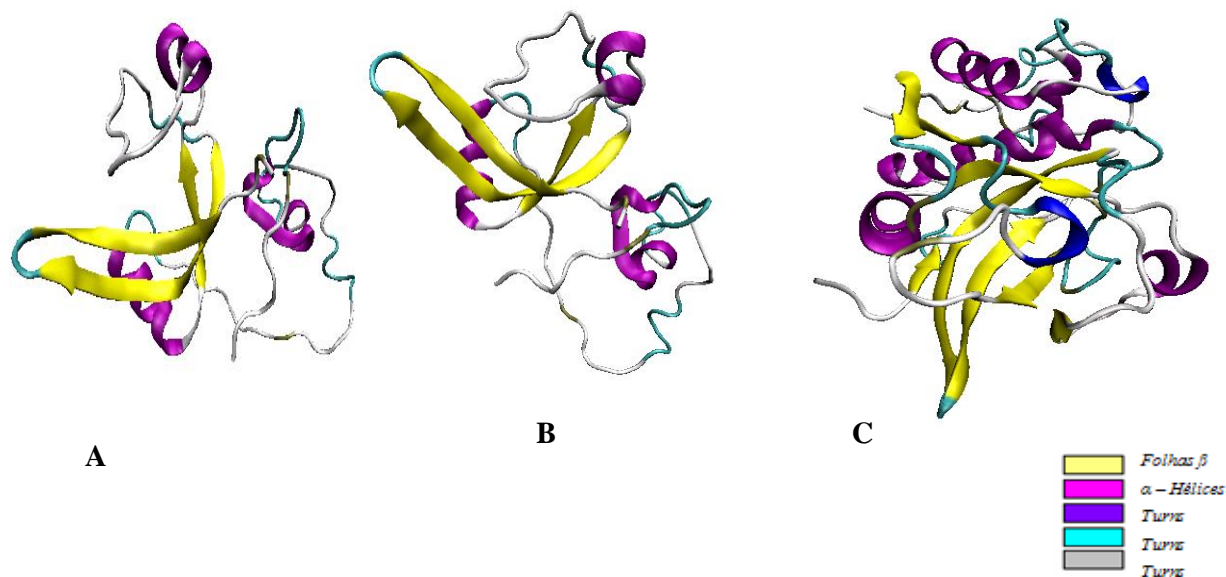
O fundamento do método de modelagem por homologia consiste na hipótese de que uma seqüência de aminoácidos similar implica em estrutura e função similares entre diferentes proteínas isoladas de diferentes organismos ou não. Deste modo, havendo a informação disponível a respeito da estrutura terciária de uma proteína, é possível prever a estrutura terciária de outra proteína homóloga cuja estrutura é desconhecida (GINALSKI, 2005; REEVES, 2006). Assim sendo, os 216 resíduos, 1634 átomos com 1675 ligações químicas da estrutura primária tridimensional da proteína procaricaina 1PCI|A foi empregada para a construção dos modelos tridimensionais denominados de dv1, dv2 e dv3 por modelagem comparativa com a seqüência primária da cisteína protease do *T. cacao* (Figuras 14 e 15).



**Figura 14.** Representação estrutural da sequência primária da 1PCI|A constituída por 216 aminoácidos, 08  $\alpha$ -hélices, 18 folhas  $\beta$ , 22 *gaps*, 1634 átomos, 1675 ligações químicas.

O alinhamento e predição da estrutura tridimensional (3D) entre proteínas consideradas de uma mesma família, mesmo que seja apenas na região do sítio ativo, pode ser um importante método para se inferir a conformação e função da sequência em estudo (LASKOWSKI, 2003). Neste contexto, três modelos (dv1, dv2 e dv3) da cisteína protease do *T. cacao* foram construídos. Em análise realizada pelo VDM (Visual Molecular Dynamics - Versão 1.8.6) (HUMPHREY, 1996), os modelos dv1 e dv2 apresentaram um número constante de 171 resíduos de aminoácidos, 1345 átomos com 1439 ligações químicas. Entretanto, o resultado obtido pela construção do modelo dv3 apresentou-se similar a 1PCI|A (216 resíduos, 1634 átomos com 1675 ligações atômicas). Uma ampla variabilidade estrutural foi verificada em todos os modelos (Figuras 15A, B e C). O mesmo número de  $\alpha$ -hélices (6) foi encontrado nos modelos dv1 e dv2. Estes modelos diferenciaram-se de acordo com o

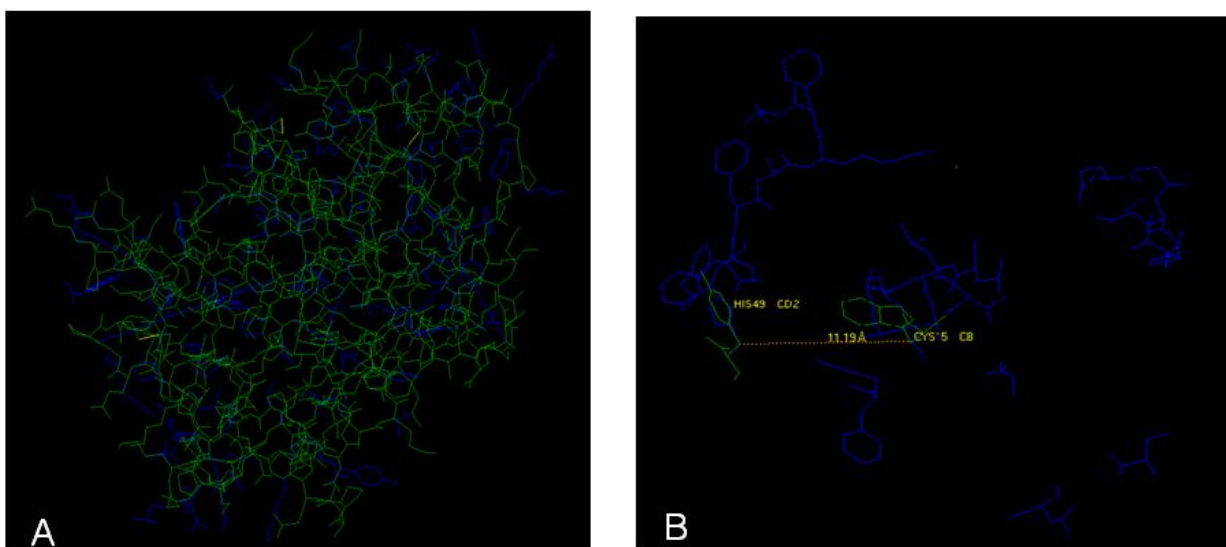
numero de folhas- $\beta$  e *turns*. Entretanto, foi encontrada similaridade estrutural entre o molde e o modelo dv3 (Figura 14 e 15C).



**Figura 15.** (A) Estrutura 3D do modelo dv1 com 6  $\alpha$ -hélices, 17 *loops* e 14 folhas- $\beta$ ; (B) Estrutura do modelo dv2 com 6  $\alpha$ -hélices, 16 *loops* e 10 folhas- $\beta$ . (C) Estrutura do modelo dv3 com 8  $\alpha$ -hélices, 22 *loops* e 22 folhas- $\beta$ .

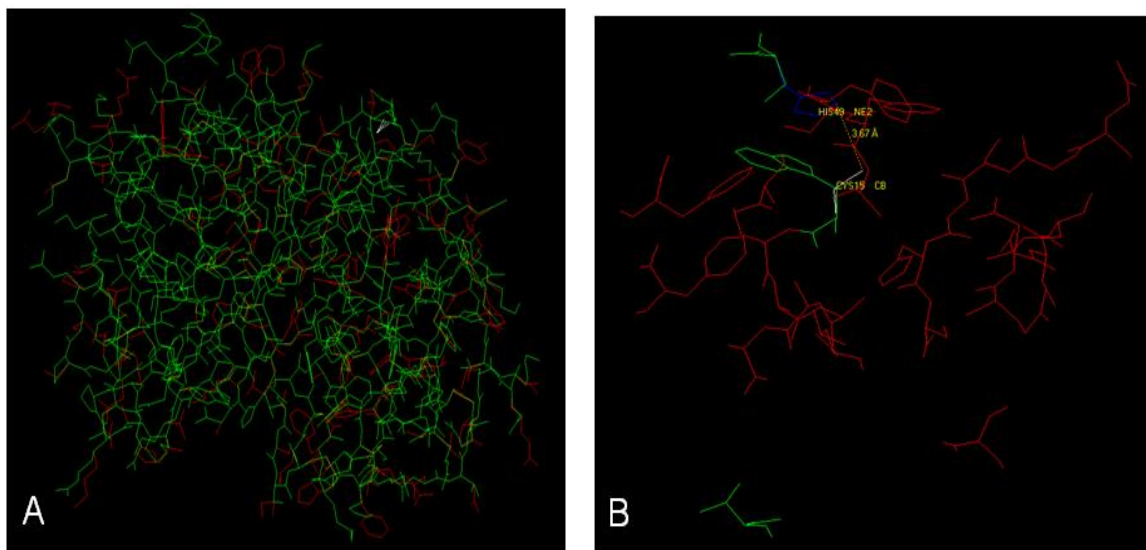
A 1PCI|A é uma cisteína protease onde os aminoácidos cisteína (Cys25) e histina (His159) estão separados por apenas 3,4 Å e representam a região de catálise enzimática (GINALSKI, 2005; REEVES, 2006; GUL, 2008). De modo similar, foi determinada a região do sítio ativo nos modelos dv1, dv2 e dv3. Os resíduos que compõem esta região para o modelo dv1 correspondem a Leu12, Leu13, Val14, Cys15, Ser16, Ala17, Ala18, Ala19, Ala20, Phe39, Phe42, His49, Phe54, Ser55, Phe56, Ala57 Arg58, Phe59, Ala60, Ala125, Glu145, Glu146, Ser147, Gly148, His149, Ser150, Ser166, Leu169. Neste modelo o RMS

entre o alinhamento dos resíduos Cys25 e Glu24 foi de 14.16 Å e para os resíduos His159 e Leu95 foi de 11.04 Å (Figura 16A e B).



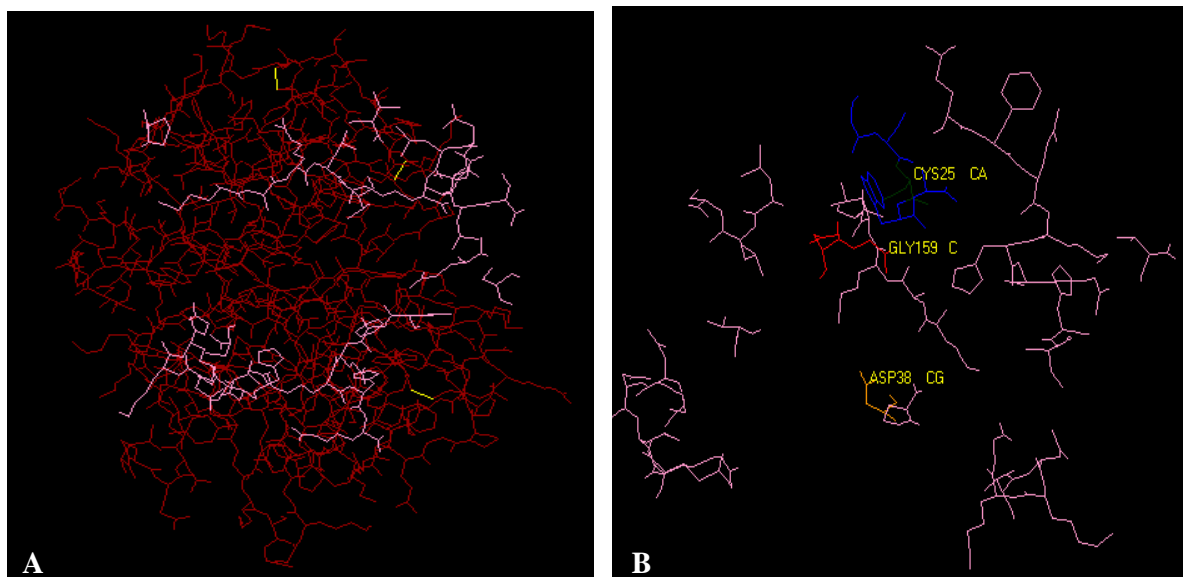
**Figura 16.** (A) Modelagem das estruturas moleculares para obtenção do modelo dv1 visualizado na cor azul e na cor verde 1PCI|A. (B) Em destaque modelagem da região do sítio ativo.

Resíduos similares foram identificados no modelo dv2, sendo eles Leu12, Leu13, Val14, Cys15, Ser16, Ala17, Ala18, Ala19, Ala20, Phe39, Phe42, His49, Phe54, Ser55, Phe56, Ala57, Ala60, Gln121, Asn122, Cys123, Ser124, Ala125, Glu145, Gly146, Ser147, Gly148, His149, Ser150, Ser166, Leu169. O alinhamento entre os resíduos Cys25 e Val14 resultou em um RMS de 16.98 Å, para os resíduos His159 e Gly148 o RMS calculado foi de 27.99 Å (Figura 17A e B).



**Figura 17.** (A) Modelagem das estruturas moleculares para obtenção do modelo dv2 (cor vermelha). (B) Em destaque os aminoácidos Cys15 e His149 do modelo sobreposto ao molde.

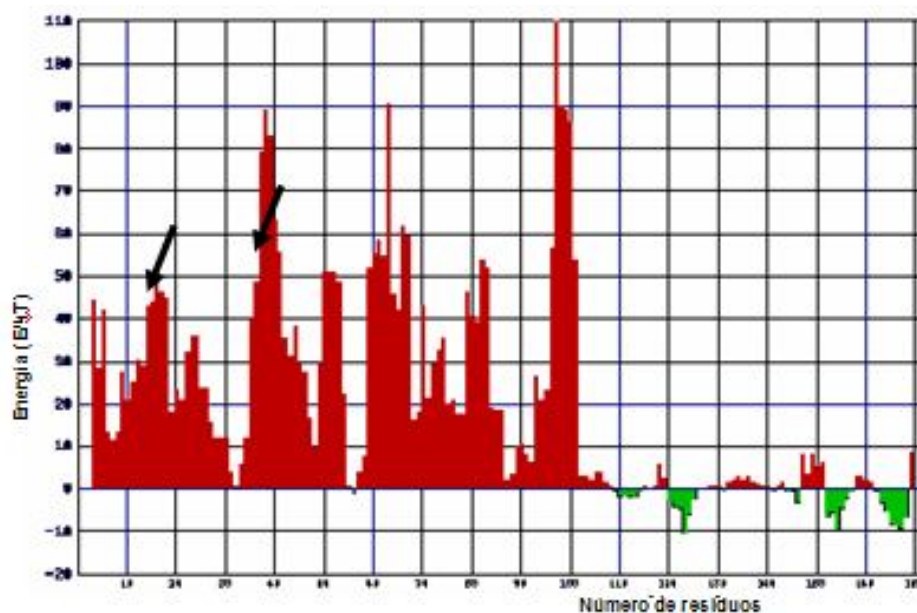
No modelo dv3 os aminoácidos que representam a região do sítio ativo está descrito por Met12, Ala13, Arg14, Leu15, Thr16, Leu18, Leu19, Ser20, Pro42, Val49, Asp54, Asp55, Leu56, Arg57, Glu60, Cys65, Thr121, Asn122, Lys123, Lys124, Gly125, Glu145, Phe146, Gln147, Lys148, His149, Arg150, Cys166, Thr169, His174. O RMS encontrado entre o alinhamento dos resíduos Cys 25 e Asp38 foi de 20.67 Å. Neste modelo não houve alinhamento com a His159 do molde (Figura 18A e B).



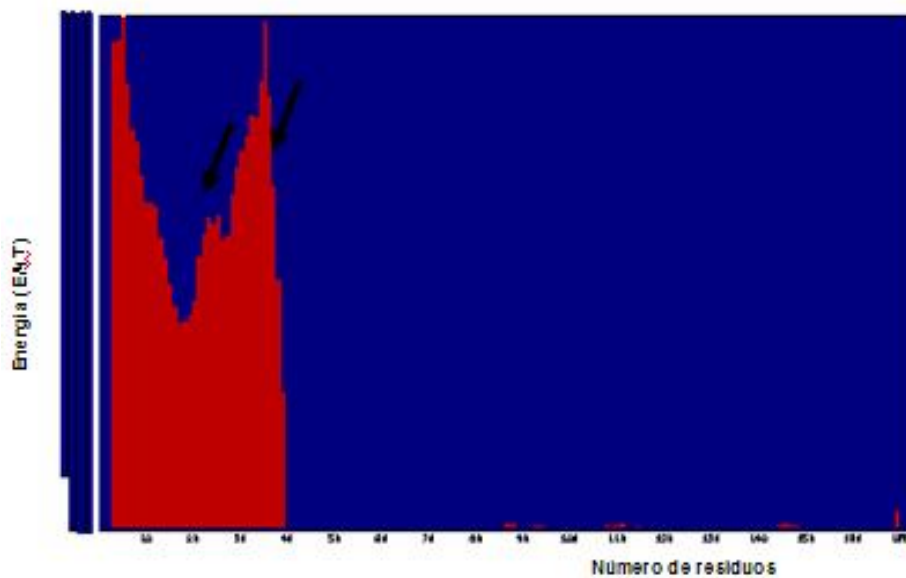
**Figura 18.** (A) Modelagem das estruturas moleculares para obtenção do modelo dv3 (cor vermelha). (B) Em destaque o alinhamento da Cys 25 e com ausência de alinhamento entre as seqüências com a His159.

A qualidade da cadeia lateral dos modelos propostos foi avaliada pelo programa ANOLEA, disponível no Swiss Model. Este programa calcula a energia de conformação de cada resíduo da cadeia lateral da proteína (MELO, 1998). Entre os modelos propostos, dv1 e dv3 apresentaram favorável qualidade estereoquímica nos aminoácidos da cadeia lateral. Os valores de energia obtidos com a validação destes modelos foram de -10,215 E/kT e -9,358 E/kT respectivamente. Por esta análise foi verificado que os resíduos Phe56 (-0,733), Trp109 (-0,208), Glu110 (-1,687), Glu111 (-1,345), Phe112 (-1,750), Gln113 (-1,843), Lys114 (-0,574), Ala120 (-2,508), Gln121 (-4,013), Asn122 (-4,360), Cys123 (-10,215), Ser124 (-5,889), Ala125 (-2,571), His131 (-0,518), Ser141 (-0,482), Leu144 (-0,377), Glu145 (-0,305), Gly146 (-3,072), Ser152 (-6,581), Leu153 (-5,273), Ser154 (-9,337), Thr155 (-4,500), Ser156 (-2,293), Arg157 (-0,448), Leu162 (-0,093), Ala163 (-3,198), Gly164 (-4,977), His165 (-8,356), Ser166 (-8,410), Val167 (-9,054), Pro168 (-6,516), no modelo dv1 apresentaram valores negativos de energia. Referências similares foram encontradas no modelo dv3: Glu46

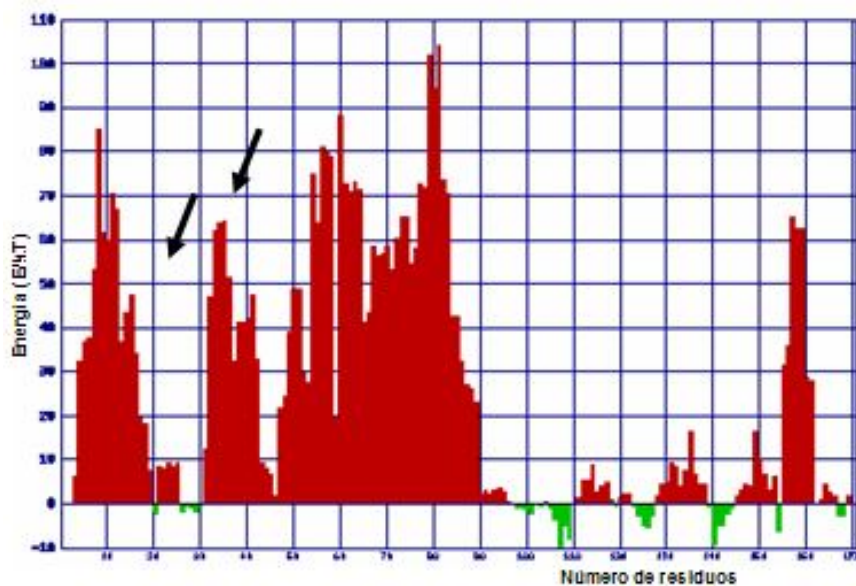
(-0,206), Thr103 (-0,777), Asn104 (-1,837), Lys105 (-1,611), Lys106 (-1,265), Gly107 (-0,892), Val114 (-4,559), Asn115 (-1,755), Pro116 (-1,193), Phe120 (-2,908), Ala121 (-1,461), Asp122 (-1,362), Thr124 (-0,038), Thr142 (-4,156), Asp145 (-0,044). O modelo dv2 apresentou 92,98% dos resíduos com elevada energia, com apenas -2,1840 E/kT, resultando em uma estrutura com conformação química inadequada com elevado grau de instabilidade nos resíduos das cadeias laterais. A região dos resíduos constituintes do sítio ativo da molécula, também foi avaliada nos três modelos propostos, e os resíduos de Cys25 e His159 nos modelos dv1 e dv3 encontram-se representados nas regiões com baixa energia potencial como indicado nas Figuras 19A, B e C para os modelos dv1, dv2, dv3 respectivamente.



A



B



C

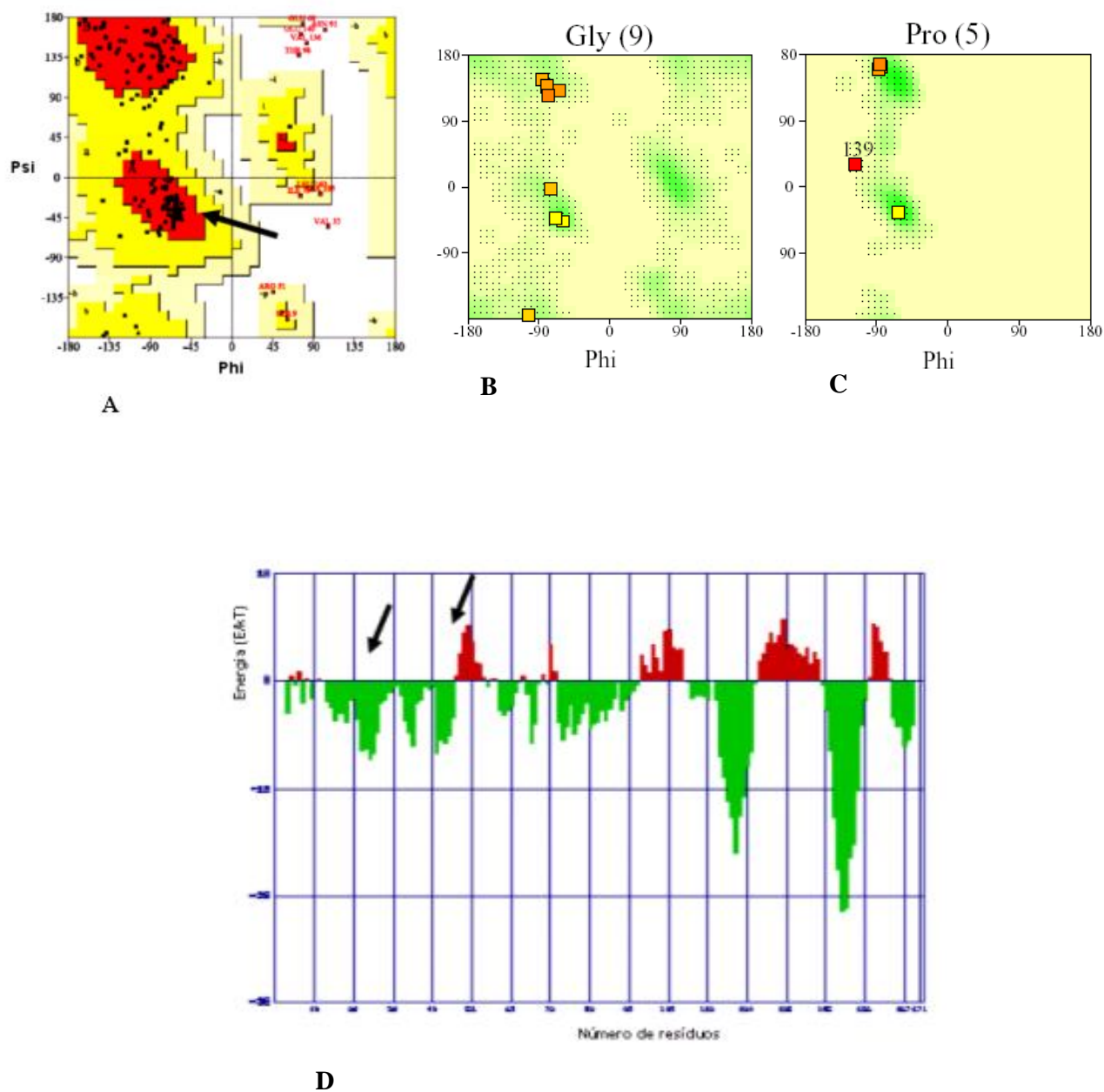
**Figura 19.** Representação da energia potencial de cada átomo na cadeia da proteína através do gráfico do ANOLEA para a validação dos modelos dv1(A), dv2(B) e dv3(C). O valor de energia é dado em  $E/kT$ , onde  $k$  corresponde a constante de Boltzman (0,582 Kcal/mol) e  $T$  corresponde a temperatura absoluta. As projeções em vermelho referem-se a um valor positivo de energia potencial para cada resíduo, e as projeções em verde identificam os valores de energia negativa, ou seja, valores satisfatórios. Em destaque as regiões do sitio ativo.



Após análise da validação dos resíduos da cadeia polipeptídica dos modelos propostos, foi constatado que o modelo dv2 devido ao elevado grau de energia entre os resíduos não alcançou uma energia potencial satisfatória estabelecida advinda de interações físico-químicas, as quais estabilizam e mantêm as estruturas enoveladas em uma proteína. Portanto a metodologia para cálculos de MM e DM para minimização de energia molecular para obtenção de uma melhor conformação espacial entre os resíduos de aminoácidos da cisteína protease do *T. cacao* foi empregada para os modelos dv1 e dv3 (CASE, 2006).

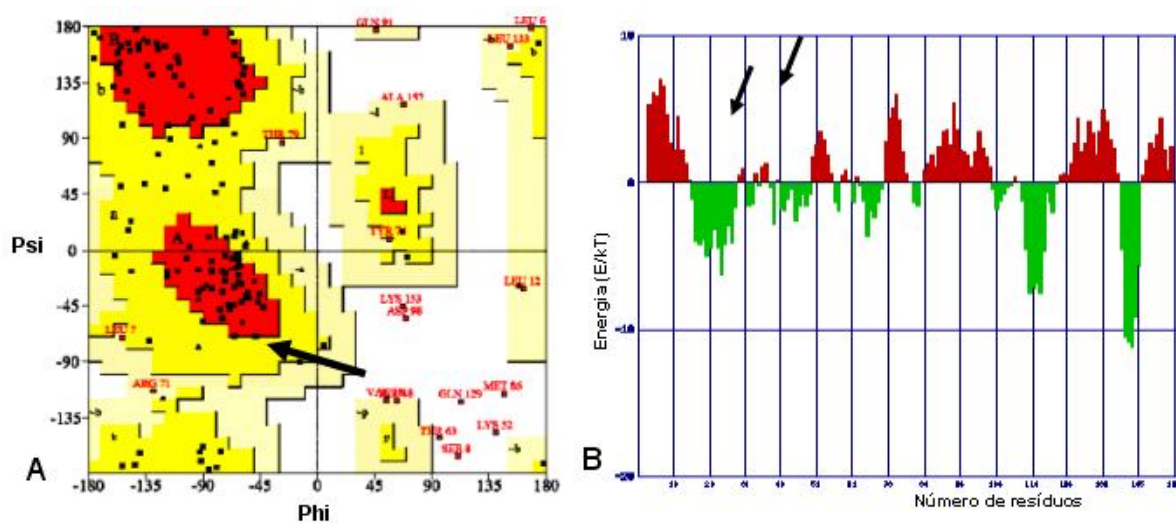
### 6.3. OTIMIZAÇÃO

Otimização molecular consiste principalmente em ajustar a molécula, para que assumam menor estado de energia (minimização de energia) (CASE, 2006). Os modelos dv1 e dv3 foram submetidos à otimização a vácuo e solvente intrínseco a uma resolução de 2Å. E, a qualidade geométrica dos modelos foi avaliada pelo programa PROCHECK (Versão 3,0) e pelo *software* ANOLEA, disponível no Swiss Model. Considerando os dados obtidos através do refinamento a vácuo não houve diferenças significativamente favoráveis entre os modelos analisados. Entretanto, baseado na análise das estruturas na presença de solvente previamente determinado pelo campo de força ff09 do programa AMBER (CASE, 2006), foi constatado que 72,9% dos resíduos do modelo dv1 apresentam-se inseridos em regiões energeticamente favoráveis, 20,0% dos resíduos estão em regiões permitidas, 3,9% em regiões concedidas e 3,2% em regiões desfavoráveis. Na área desfavorável consta apenas os resíduos de Glu68, Asn91, Glu140, Val136, Thr98, Leu160, Ile79, Ala105, Val32, Arg51, Ser9 (Figura 20A, B, C e D ). O Procheck analisa separadamente os resíduos de glicina e prolina, conforme é mostrado pelas Figuras 20B-C, respectivamente. Esta separação é necessária, pois estes aminoácidos apresentam diferenças estereoquímicas, e isso requer diferentes características em suas regiões favoráveis ou desfavoráveis (LASKOWSKI, 1994).



**Figura 20.** (A) Gráfico de Ramachandran do modelo dv1 com 72,9% dos aminoácido da cadeia principal nas regiões favoráveis (A e B) e 20,0% nas regiões permitida (a,b,l,p). Em destaque região do sítio; B e C. Ramachandran da Glicína e da Prolina, respectivamente, onde observa-se o resíduo Pro139 em região desfavorável (Fote: Procheck 3,0). (D) Gráfico de energia obtido pelo ANOLEA das cadeias laterais de aminoácidos (-21,4170 E/kT ; + 5,8440 E/kT "32,93%"). Em destaque região do sítio ativo.

Segundo a literatura, um bom modelo deve possuir pelo menos 90% dos resíduos entre as regiões energeticamente favorável e energeticamente permitida no Gráfico de Ramachandran (LASKOWSKI, 1994). Neste modelo a qualidade geométrica da estrutura aferida pelo Gráfico de Ramachandran foi de aproximadamente 93% e a energia de formação validada pelo ANOLEA foi de  $-21,4170$  E/kT. Similar qualidade estutural não foi encontrada na validação do modelo dv3. A qualidade estrutural aferida para este modelo pelo gráfico de Ramachandran demonstrou que 55,1% dos resíduos da cadeia principal estão contidos em regiões energeticamente favoráveis e 31,9% em regiões permitidas, 8,0% em regiões concedidas com 5,1% em regiões desfavoráveis. A energia de formação aferida pelo ANOLEA foi de  $-11,2180$  E/kT (Figura 21A e B).

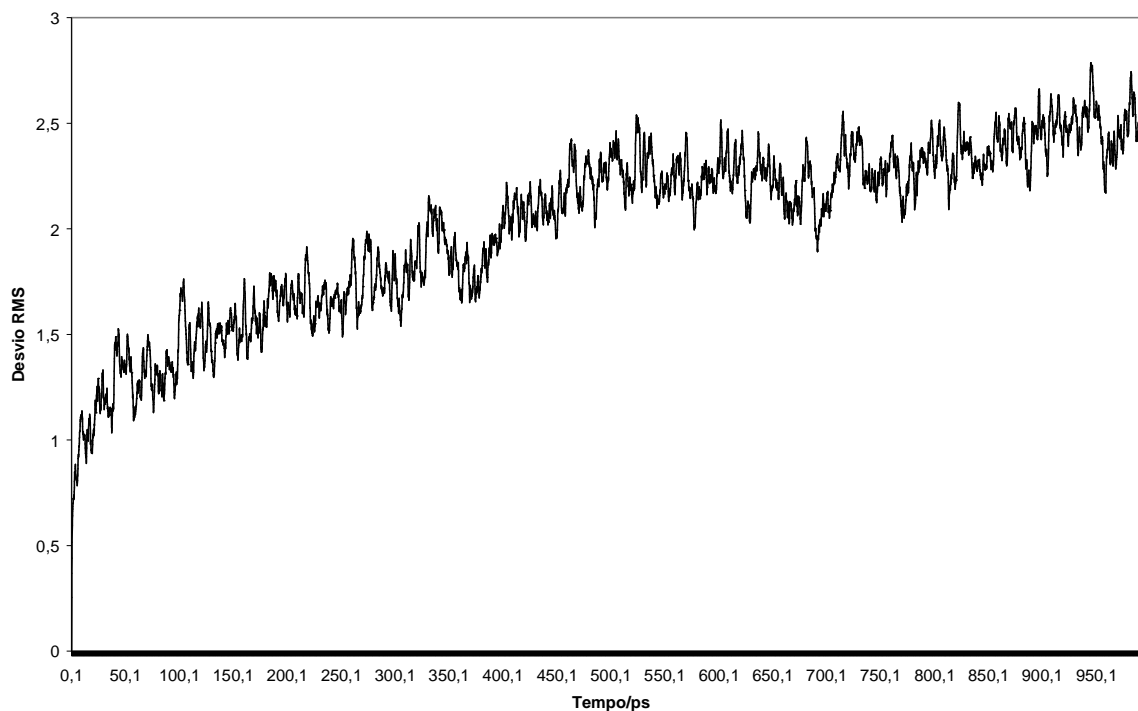


**Figura 21.** (A) Gráfico de Ramachandran do modelo dv3 com 55,1% dos aminoácido da cadeia principal nas regiões favoráveis (A e B) e 31,9% contidos nas regiões permitida (a,b,l,p). (B) Gráfico de energia das cadeias laterais de aminoácidos ( $-11,2180$  E/kT ;  $+7,0350$  E/kT "54,97%"). Em destaque região do sítio ativo.

O modelo dv3 resultante da otimização apresentou incongruência na posição de alguns resíduos demonstrados pela validação do Procheck e também do ANOLEA. Logo, este modelo não possui uma boa qualidade geométrica. No entanto, o modelo dv1 resultou em um melhor arranjo espacial da cadeia principal no Gráfico de Ramachandran (93%), comitantemente houve um aumento da qualidade dos parâmetros nas cadeias laterais.

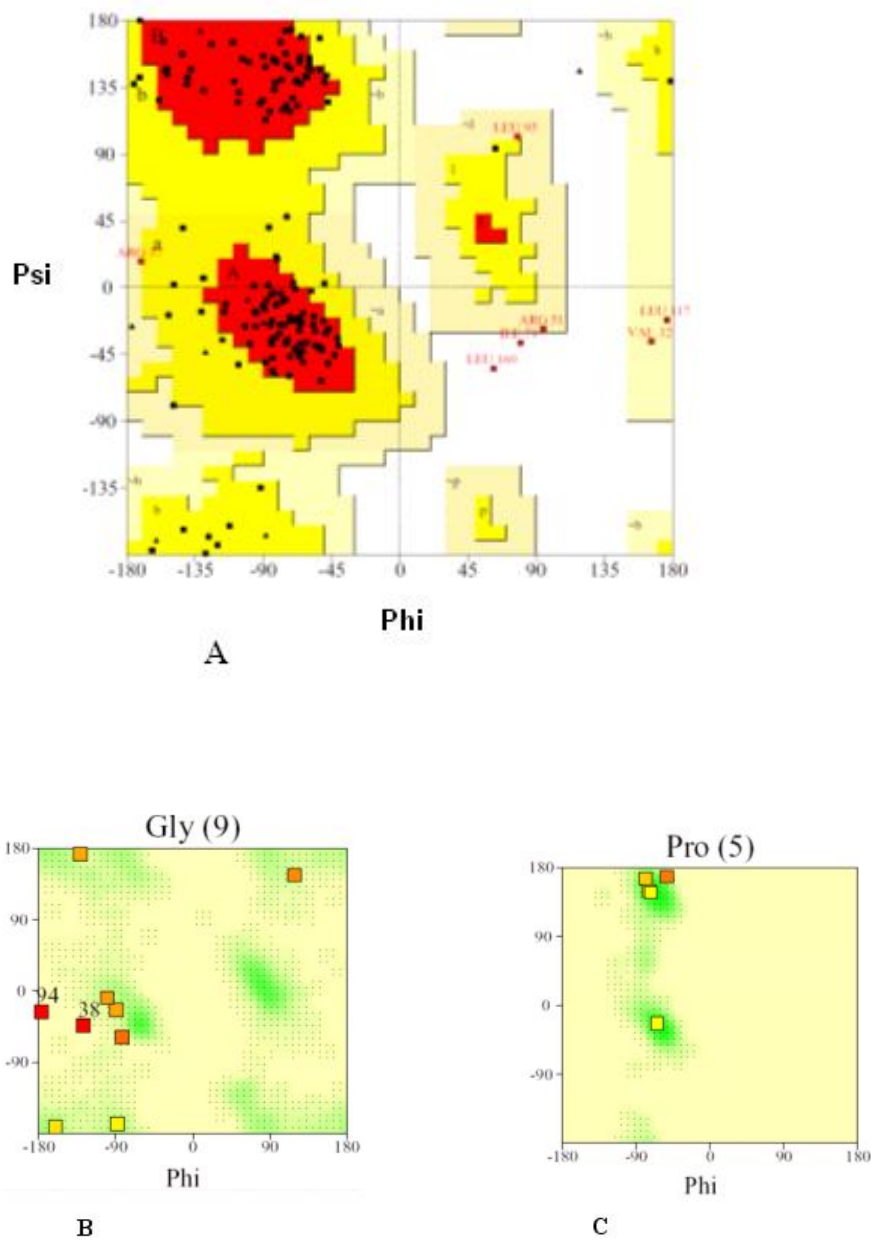
#### **6.4. DINÂMICA MOLECULAR**

O modelo dv1 foi submetido à dinâmica molecular (DM) para verificação de sua estabilidade termodinâmica e busca de um conformero de menor energia na superfície de energia potencial. Simulações de DM são de fundamental importância para se conhecer o potencial de interação entre os átomos e as interações que regem seu movimento em função do tempo. A estratégia desenvolvida para alcançar a estrutura 3D da cisteína protease do *T. cacao* consistiu em submeter o modelo dv1 à DM. Ao final das simulações, com duração de 1 picosegundo a estrutura apresentou estabilidade com o decorrer do tempo, como é possível observar no gráfico de desvio RMS a partir de aproximadamente 500ps (Figura 22). A estrutura torna-se praticamente constante após atingir 600ps, esse comportamento aponta para uma estabilização geométrica do modelo proposto.

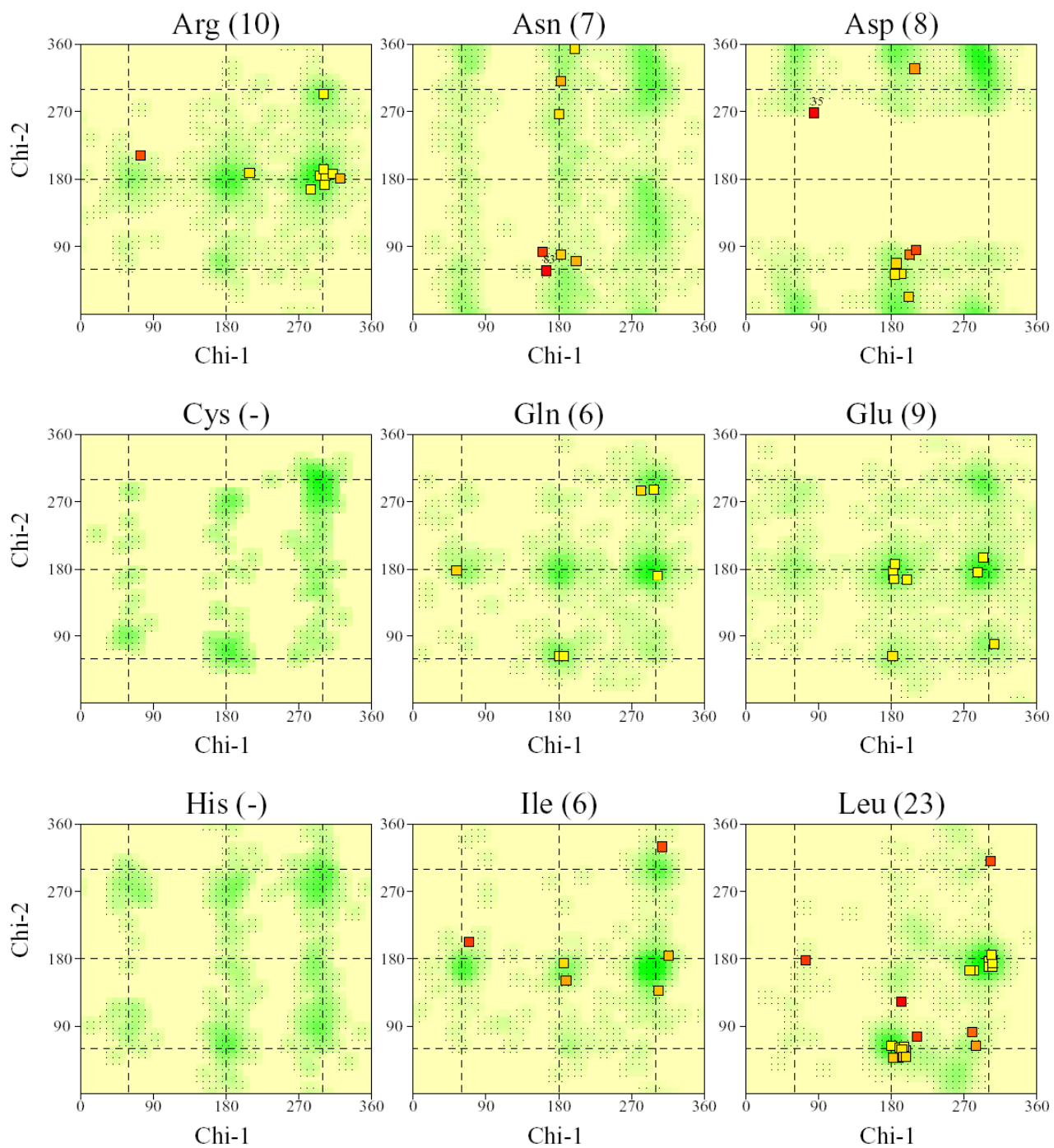


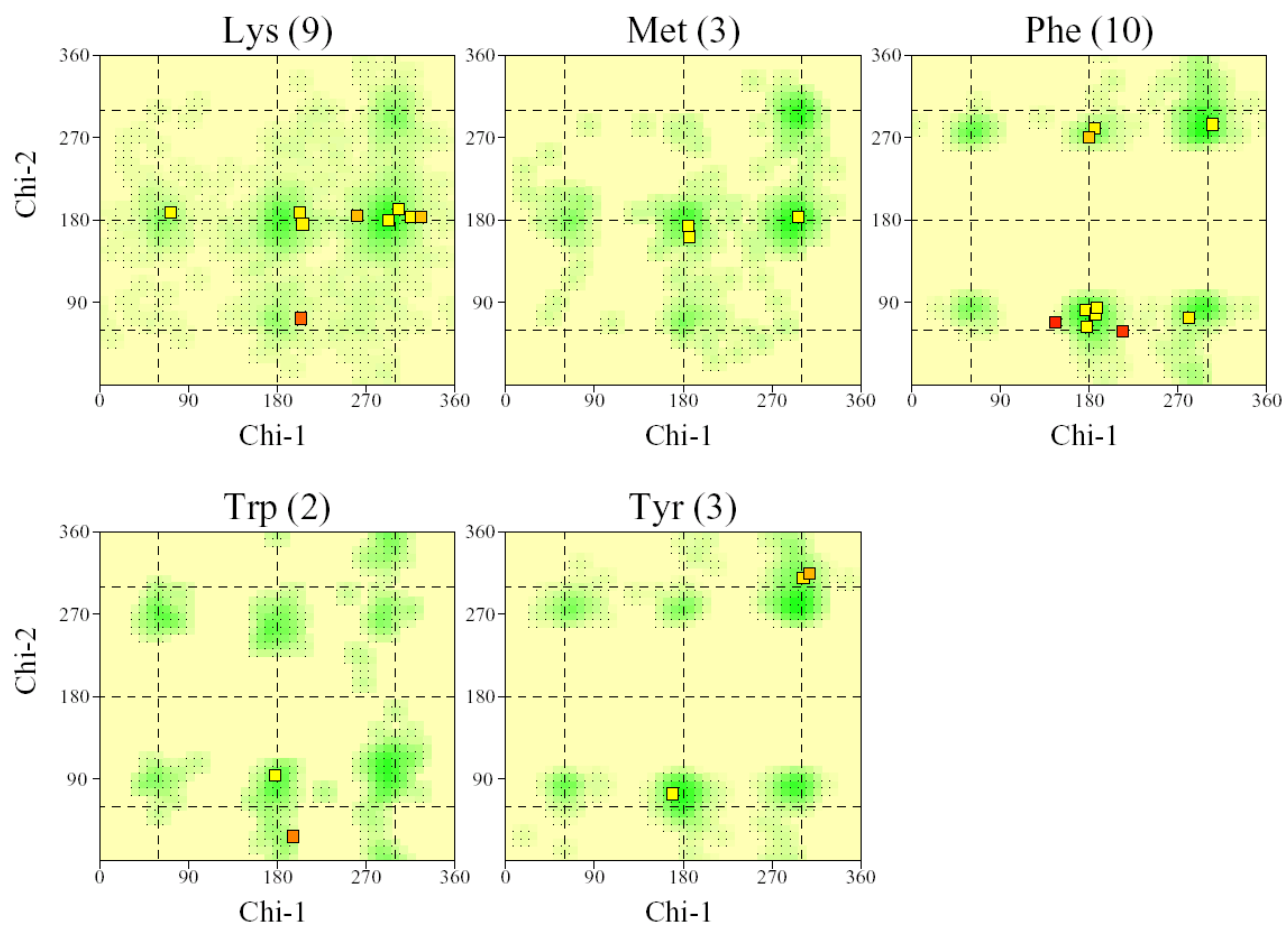
**Figura 22.** Gráfico do RMS pelo tempo de simulação da cisteína protease do *T. cacao*.

Por meio do Gráfico de Ramachandran foi verificado que 97,58% dos aminoácidos encontram-se em regiões energeticamente favoráveis (*core region*), sendo que apenas 1,3% estão localizados em regiões desfavoráveis. Desse modo, observa-se que ao final da DM o modelo dv1 apresentou uma melhor qualidade estereoquímica em relação aos modelos otimizados (Figuras 20 e 21). Apenas os resíduos de Ile7 e Leu160 não alcançou uma geometria favorável. A Figura 23A, B e C, mostram a qualidade estereoquímicas dos ângulos de torção da cadeia lateral de todos os aminoácidos. A Figura 24, 25 e 26 avalia a qualidade do Gráfico de Ramachandran obtido, planaridade das ligações peptídicas, má interação entre os átomos não-ligados e energia de ligação de hidrogênio na cadeia principal. Nestas representações o desejado é que o ponto alcance à reta da faixa sombreada.



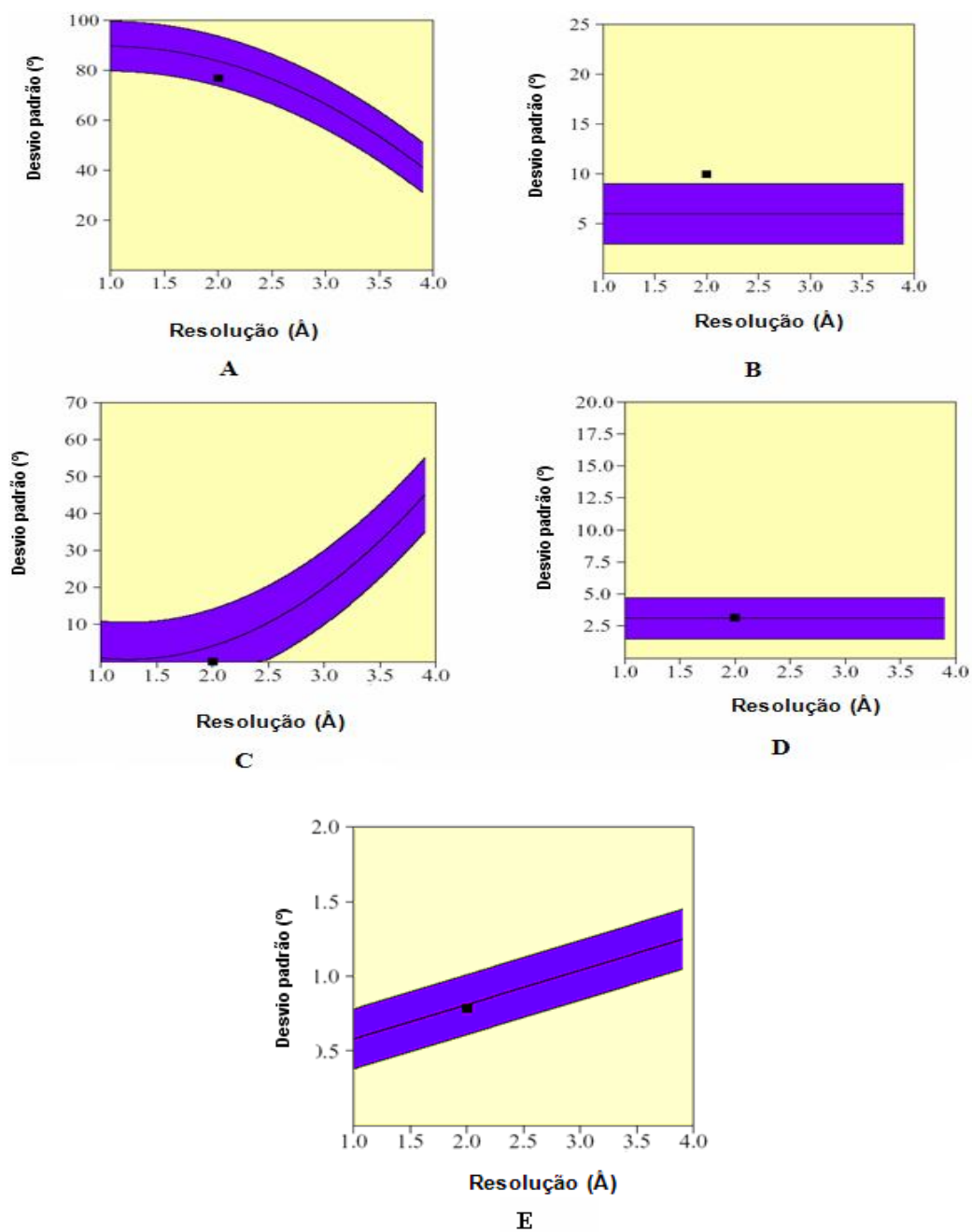
**Figura 23.** (A) Gráfico de Ramachandran do modelo 3D da cisteína protease do *T. cacao*. Respectivamente B e C Ramachandran da Glicína e Ramachandran da Prolina onde observa-se o resíduo Gly94 e 38 em região desfavorável.



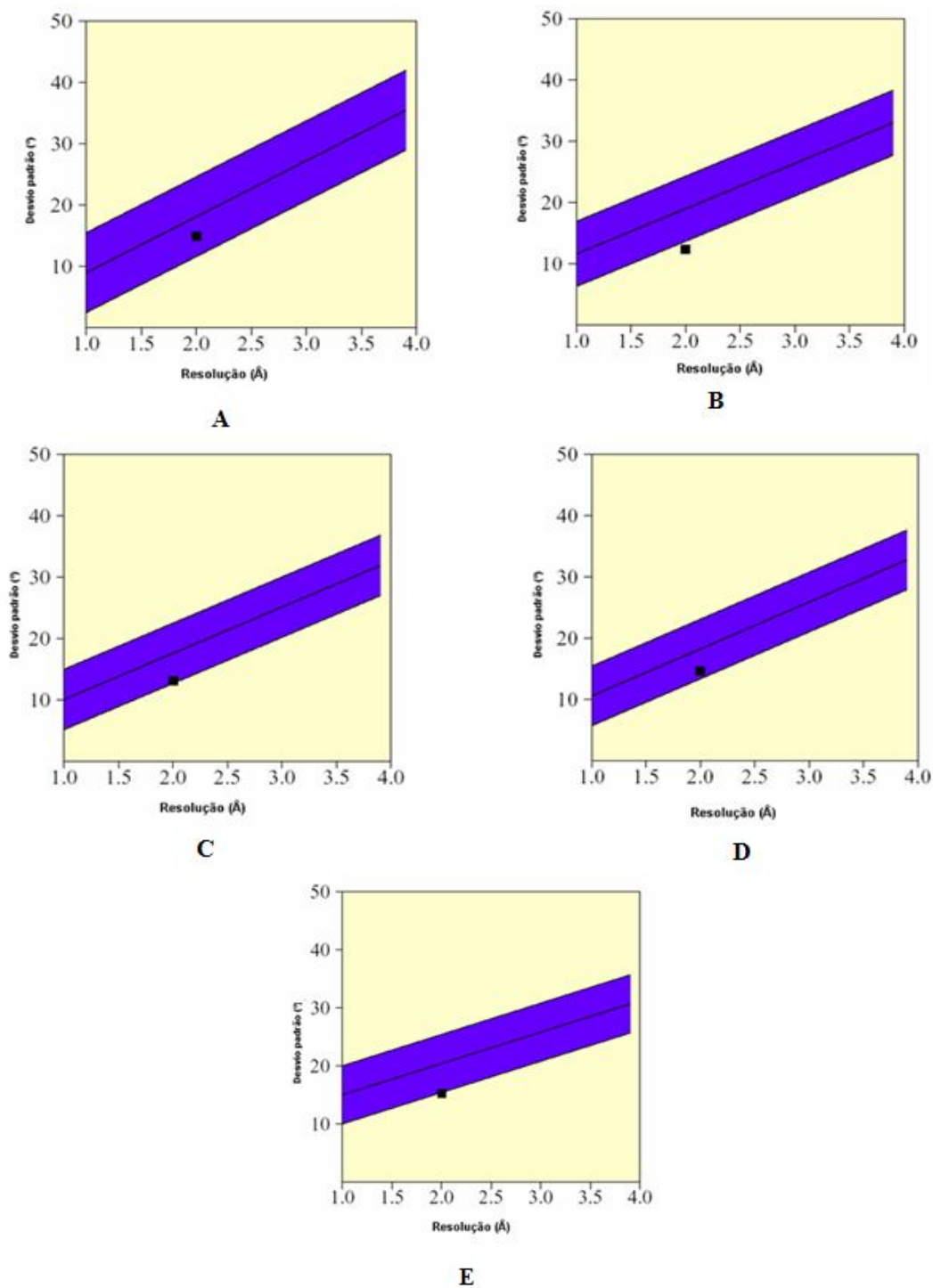


**Figura 24.** Gráficos Chi-1 e Chi-2. Os números de resíduos são mostrados entre parêntese. As regiões sombreadas representam áreas favoráveis, com exceção às áreas com coloração vermelha, estas demonstram resíduos com ângulo desfavorável.



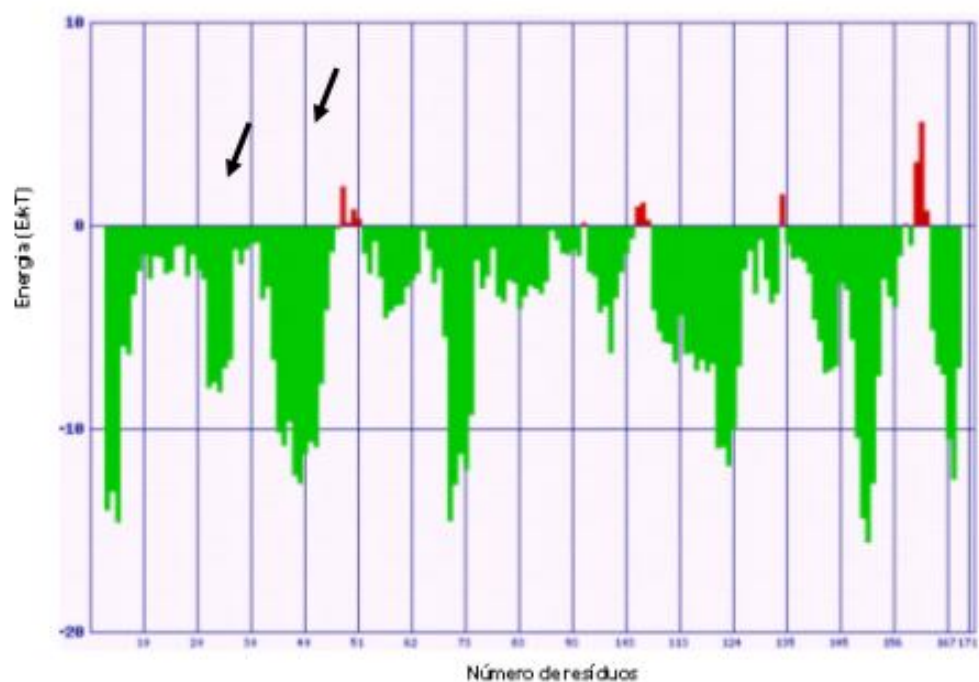


**Figura 25.** Qualidade da cadeia principal do modelo estrutural da cisteína protease do *T. cacao*. O eixo das abscissas está descrito em graus, e o das coordenadas está descrito em Å: (A) Avaliação do Gráfico de Ramachandran; (B) Planaridade da ligação peptídica; (C) Interações ruins; (D) Distorção dos carbonos alfa; (E) Energia das ligações de hidrogênio.



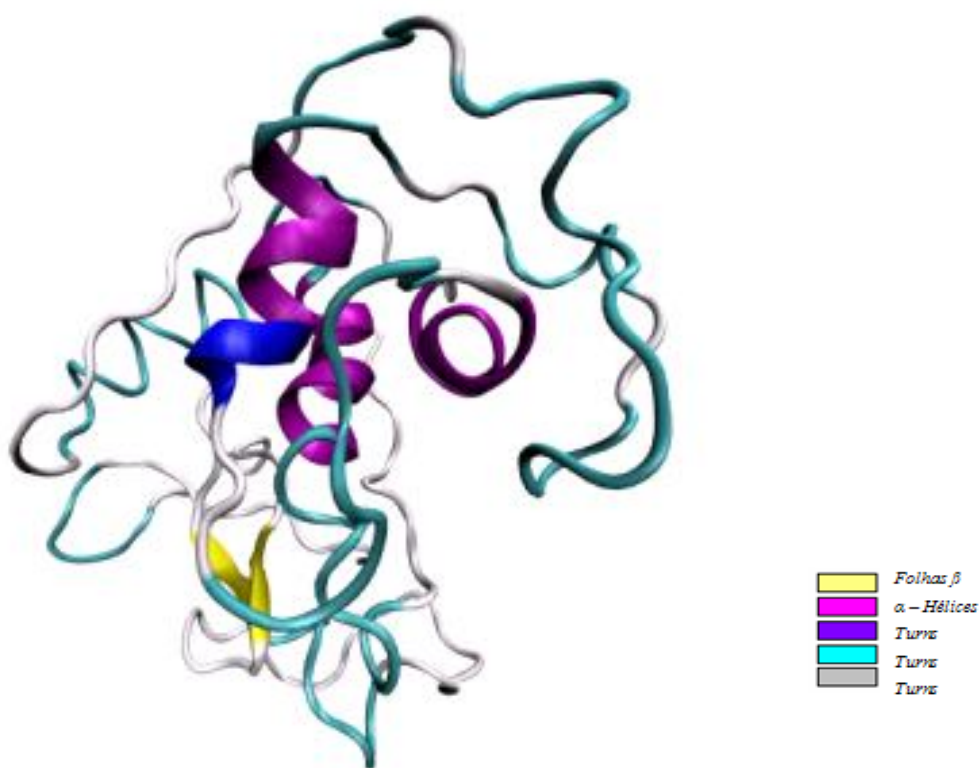
**Figura 26.** Desvio padrão para os ângulos diedros Chi1 do modelo estrutural da cisteína protease do *T. cacao*. (A) Conformação gauche menos; (B) Conformação trans; (C) Conformação gauche mais; (D) Somatório de todos os desvios padrões para Chi1; (E) Conformação trans para o ângulo torcional Chi2. O eixo das abscissas está descrito em graus, e o das ordenadas está descrito em Å.

Em todas as validações através do Procheck foi verificado um resultado satisfatório para o modelo dv1. A melhoria da qualidade estrutural do modelo pela análise de superfície da força atômica (*Atomic Mean Force Potential – AMPF*), realizada pelo ANOLEA, apresentaram resultados satisfatório com energia de  $-750,740$  E/kT ficando apenas os resíduos Leu47, Gly48, Thr50, Arg51, Ala105, Asp106, Gly129, Asn130, Gln132, Leu133, Pro139, Glu140, Ser141, Leu158, Asn161, Leu162 com energia positiva (Figura 27). Dessa forma, o modelo dv1 é o melhor modelo proposto para representação estrutural da cisteína protease do *T. cacao*.



**Figura 27.** Representação gráfica de cada átomo na cadeia de proteína após a DM. As projeções em verde referem-se a um valor de energia favorável e as projeções em vermelho a valores desfavoráveis de energia.

O modelo 3D da cisteína protease do *T. cacao* foi construído com 171 aminoácidos, formados por 2693 átomos, unidos por 2719 ligações químicas. Através de uma análise estrutural do modelo foi identificado que o mesmo está composto por 7  $\alpha$ -hélices, 23 *turns* e 2 folhas- $\beta$  (Figura 28). Os aminoácidos Leu13, Cys15, Ser16, Ala17, Ala19, Ala20, Thr21, Ile22, Glu24, Ser25, Ile29, Glu40, Ser41, Val43, Leu44, Arg58, Phe59, Ala60, Try61, Gly64, Thr108, Glu110, Glu113, His115, Cys123, Ala127, Gly129 e Ala163 assumem a conformação de alfa hélices e os aminoácidos Leu7, Ser9, Leu13, Asp35, Leu36, Arg37, Gly38, Val70, Glu71, Gln121, Asn122, Gly148, Ser150, Ser152, Ser154, Thr155, Ser156, Arg157, Leu158, Leu160, Asn161, Leu162, Ala163, Gly164, Ser166 e Pro168 correspondem as folhas- $\beta$ . Como pode ser observado na Figura 29, estas estruturas secundárias localizam-se essencialmente em uma região mais externa da proteína. Estes estão direcionados para a parte externa da proteína enquanto que os aminoácidos que compõem as  $\alpha$ -hélices estão direcionadas para o interior. Esta conformação estrutural da molécula poderá favorecer interações com prováveis ligantes em um ambiente enzimático.



**Figura 28.** Estrutura 3D da cisteína protease do *T. cacao*.

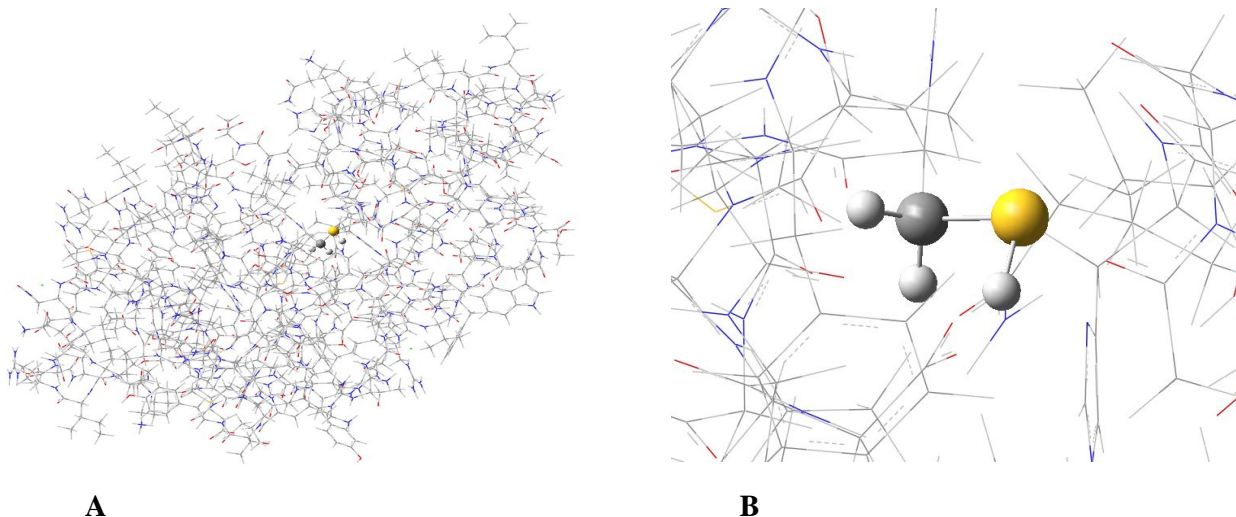


**Figura 29.** Estrutura 3D da cisteína protease do *T. cacao*. Em destaque a região do sítio catalítico da enzima.

## 6.5. CÁLCULOS DE QM/MM

O método de mecânica molecular não permite descrever propriedades onde há a necessidade explícita da participação de elétrons. No entanto, são extremamente úteis em sistemas moleculares grandes, para os quais existam parâmetros. Desse modo, torna-se necessário descrever partes do sistema por um método quântico, por exemplo DFT (MASERAS, 1995; WOO, 1998; ATAUALPA, 2006). Em nosso estudo, o resíduo Cys que compõe o sítio ativo da cisteína protease do *T. cacao* encontra-se na posição 742 com a distância de 1,8Å entre o átomo de S 762 da Cys e N 742 da His. Os aminoácidos que compõe

a região do sítio ativo desta molécula esta representado pelo Leu13, Leu12, Ser16, Cys15, Ala17-20, Phe39, Phe42, His49, Phe54, Ser55, Phe56, Ala57, Ala60, Gln121, Asn122, Cys123, Ser124, Ala125, Glu145, Gly146, Ser147, Gly148, His 149, Ser150, Ser 166 e Leu169. As Figuras 29 e 30 mostra detalhes do arranjo espacial do sítio ativo, o qual os resíduos Leu12, Leu13, Val14, Cys15, Ser16, Ala17, Ala18, Ala19, Ala20, Phe39, Phe42, Arg58, Phe59 e Ala60 formam  $\alpha$ -hélices e, os aminoácidos His49, Phe54, Ser55, Phe56, Ala57, Ala125, Glu145, Ser147, Gly148, His149, Ser150, Ser66, Leu169 formam folhas- $\beta$ . De pose dessas coordenadas foi empregado o método QM/MM implementado no programa Gaussian 03W. Dessa forma, a Cys742 ficou definida como alta camada, sendo refinado pelo método B3LYP com a base Lanl2DZ, e as demais parte da molécula ficaram definidas como baixa camada, empregando o método UFF (Figuras 30, 31A, B e C) (HEAD, 1988; FRISCH, 2003; SAKIYAMA, 2008). E o valor de energia encontrado para esta região foi de  $-274,0210 \times 10^3$  kcal/mol.



**Figura 30.** (A) Modelo estrutural dvl representado em wireframe definido por QM/MM. (B) Em destaque o sítio ativo da molécula representado sob a forma de *ballstick* com 1,8 Å entre os átomos de N e S da histidina e cisteína respectivamente.

## 6.6. DETERMINAÇÃO DO MODELO MUTANTE

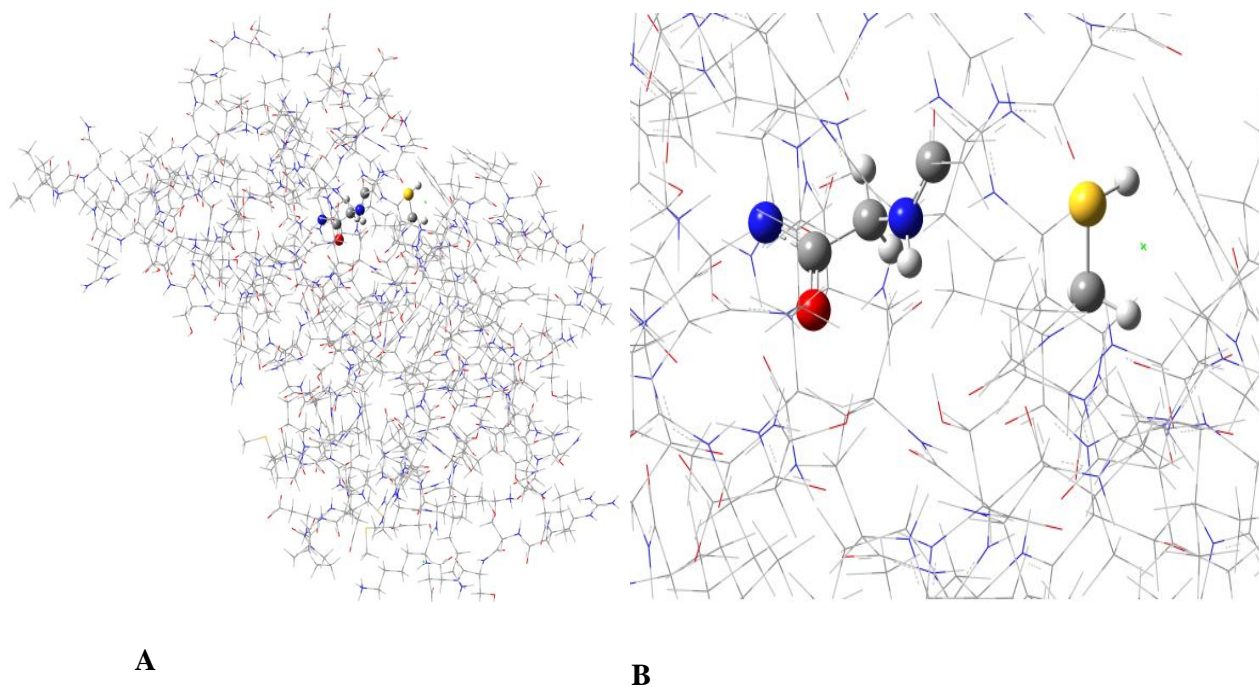
De acordo com a literatura, o resíduo His159 da seqüência primária de aminoácidos que compõe o sítio ativo da procaricana 1PCI|A sofreu um processo de mutação ao ser substituído por uma Gly. Semelhantemente, o modelo mutante da cisteína protease do *T. cacao* foi desenvolvido (GROVES, 1997). A glicina é um aminoácido que apresenta cadeia lateral curta e a sua presença inibe a formação de alfa hélices. No entanto facilita a formação de folhas beta na estrutura secundária de proteínas, devido ao seu alto grau de flexibilidade quando integrado numa cadeia polipeptídica (NELSON, 2004).

Além da determinação estrutural da cisteína protease do *T. cacao*, optou-se pela construção do modelo mutante da proteína, objetivando que a enzima em estudo pudesse apresentar atividade quelante de metais pesados que são substâncias presentes no meio ambiente devido principalmente ao processo de industrialização (Figura 31A, B, 38 e 39). Neste contexto, existe uma necessidade de desenvolvimeto de um produto químico capaz de se ligar a íons por ligações covalentes e a estruturas heterocíclica de compostos orgânicos como aminoácidos, peptídeos ou polissacarídeos que apresenta grande importancia biotecnologica e ambiental, pois a quelação funciona como o sistema de defesa natural de sulfuração onde pequenas moléculas se ligam aos metais tóxicos para sequestrá-los e/ou eliminá-los (BOBTELSKY, 1945).

O modelo mutante foi construído pelas ferramentas do Swiss PDB Viewer, sendo submetido ao programa AMBER 9.0 para cálculos de otimização molecular (CASE, 2003) e ao Gaussian 3.0W para cálculos de QM/MM. Nossos resultados demosntraram que a distância entre os átomos de S (762) da Cys 742 em relação ao átomo de H 7625 da Gly é 3,4Å (Figura 31). Sendo o total de energia calculada para o sistema por QM/MM de  $-274,0398 \times 10^3$  kal/mol (Tabela 3). O modelo proposto foi validado pelo Procheck 3.0, o qual gerou o gráfico de Ramachandran apresentado na Figura 32A, B e C. Nos parâmetros do Ramachandran 95% dos resíduos estão em locais favoráveis, caracterizando o modelo proposto como uma estrutura de boa qualidade estereoquímica. Considerando-se a análise feita no ANOLEA, obteve-se uma energia potencial de superfície de -812.502 E/kT. Apenas os resíduos Leu47, Gly48, Thr50, Arg51, Val101, Thr126, Ala127, Lys128, Gly129, Asn130, Gln132, Leu133,

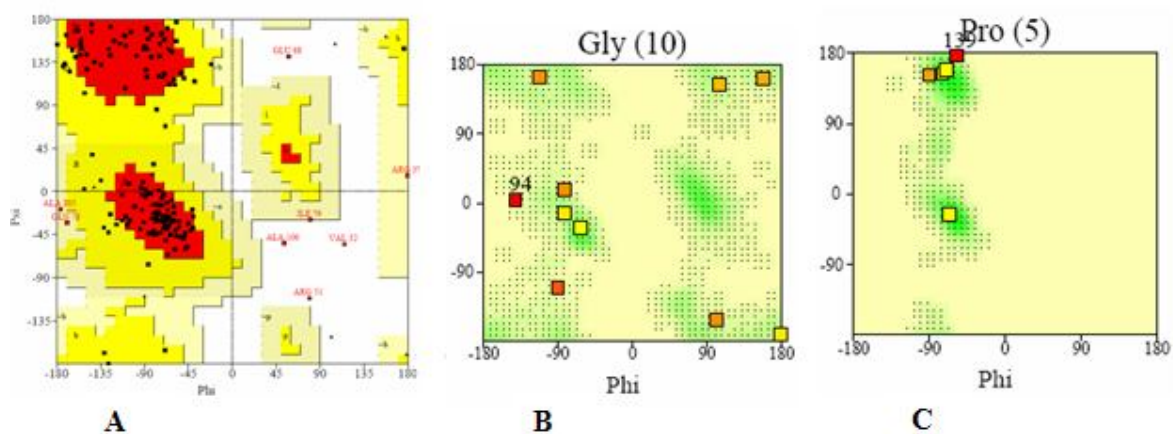


Thr134, Leu160, Asn161, Leu162 apresentaram um total de energia desfavorável de 32,241 E/kT correspondendo a apenas 10,91% dos resíduos do modelo mutante determinado (Figura 33). Os resultados obtidos mostraram um valor de energia global ideal para posteriores cálculos da estrutura resultante com íons metálicos (Figuras 35 a 39)

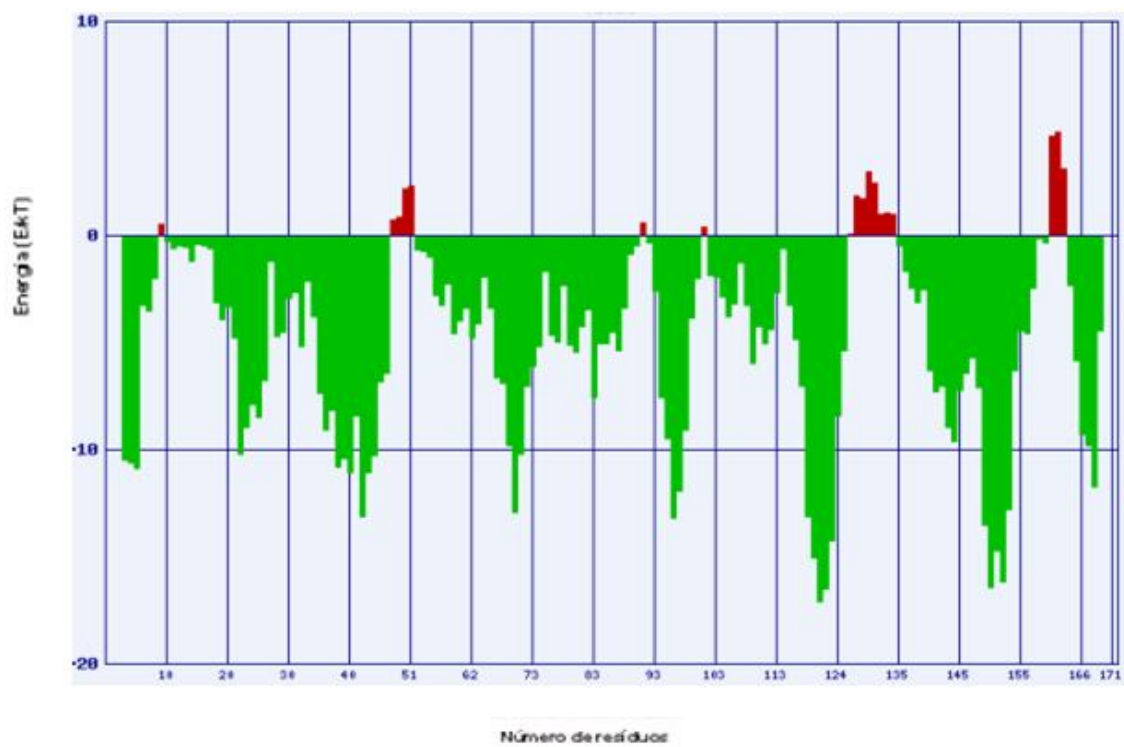


**Figura 31.** (A) Modelo estrutural do mutante analisado por QM/MM. (B) Em destaque os aminoácidos glicina e cisteína com 3,4Å entre os átomos de H e S, respectivamente.

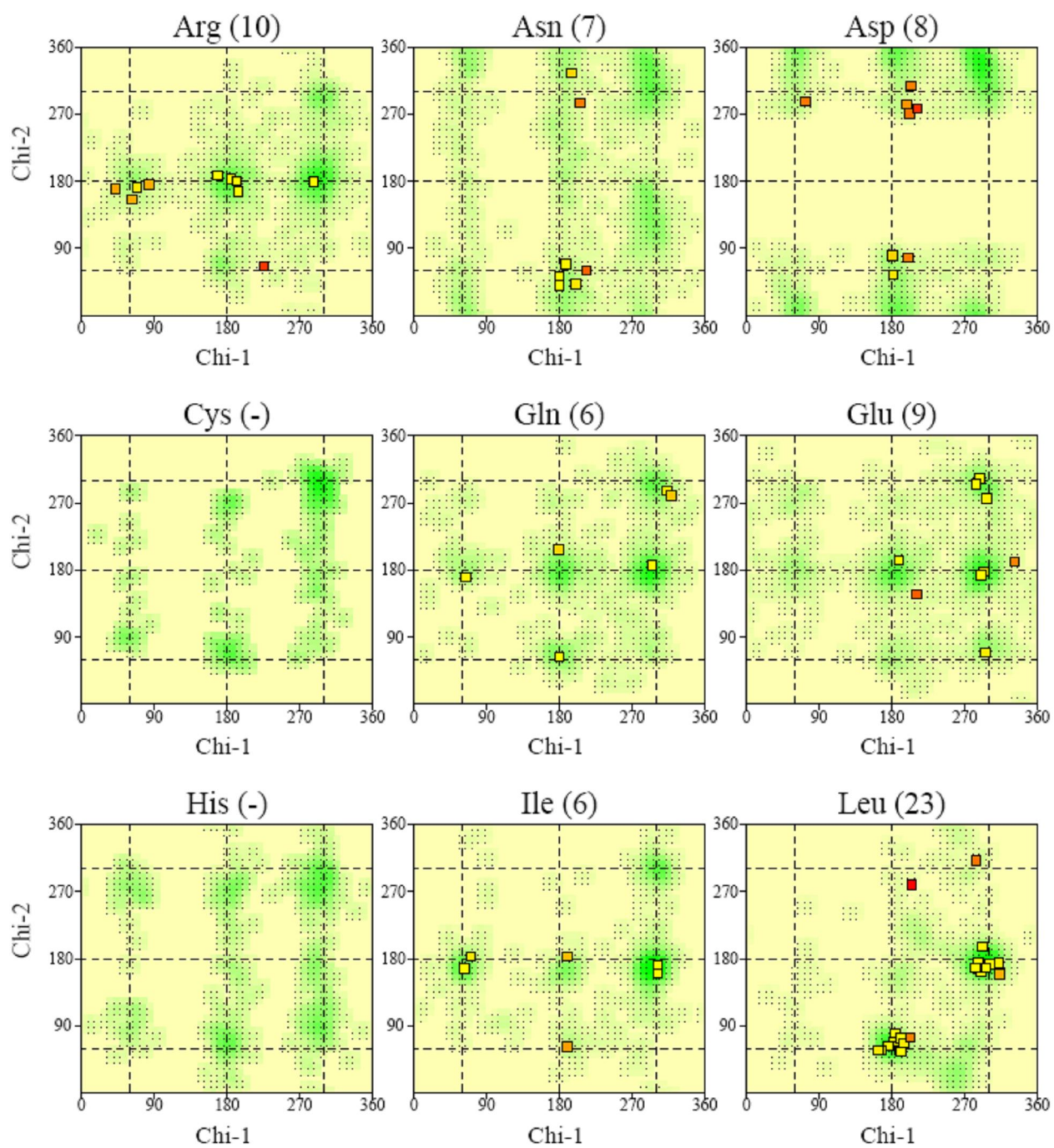


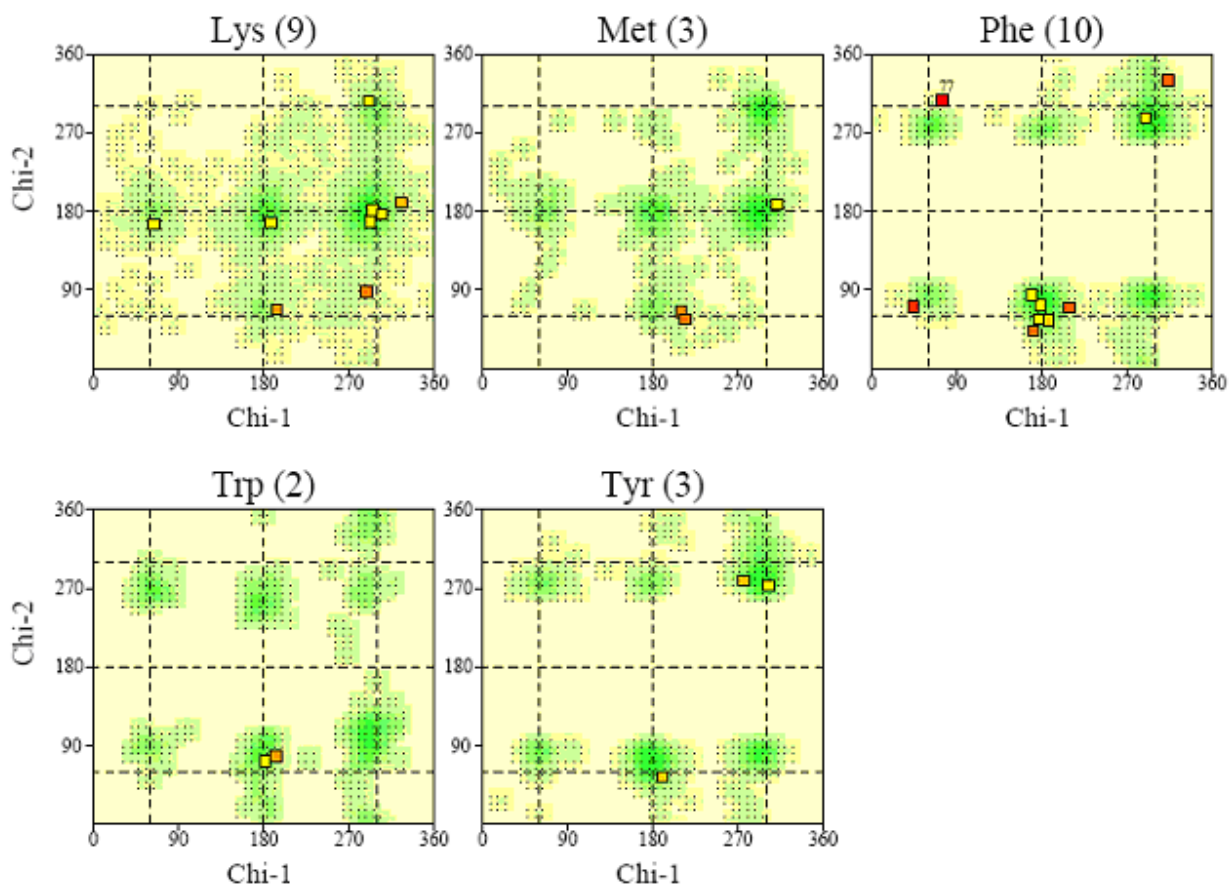


**Figura 32.** (A) Gráfico de Ramachandran do modelo mutante da cisteína protease do *T. cacao*, respectivamente B e C. Em destaque os resíduos Gly e Pro.

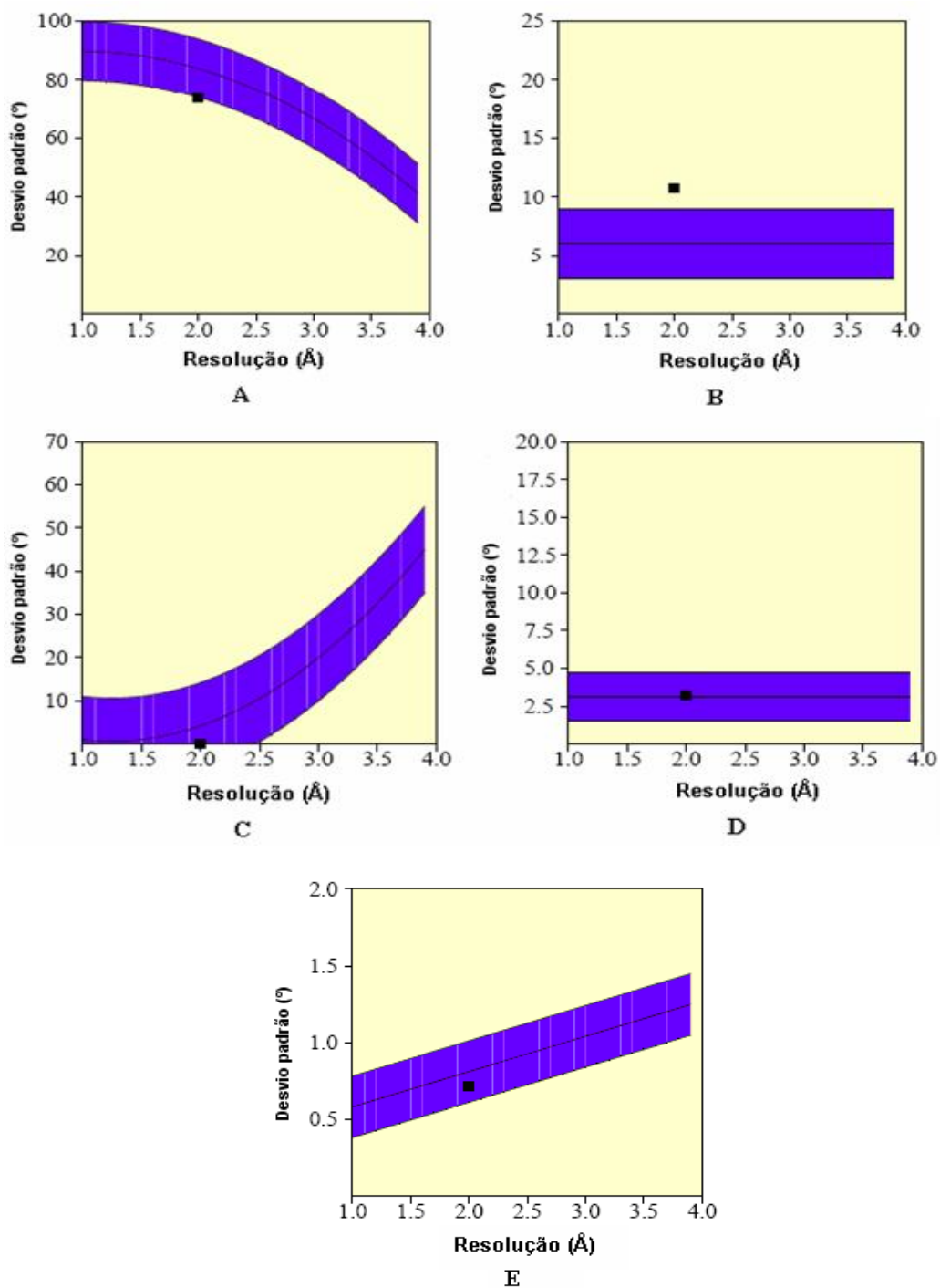


**Figura 33.** Representação gráfica de cada átomo na cadeia de proteína mutante da cisteína protease do *T. cacao* durante a DM. As projeções em verde referem-se a um valor de energia favorável e as projeções em vermelho a valores desfavoráveis de energia.

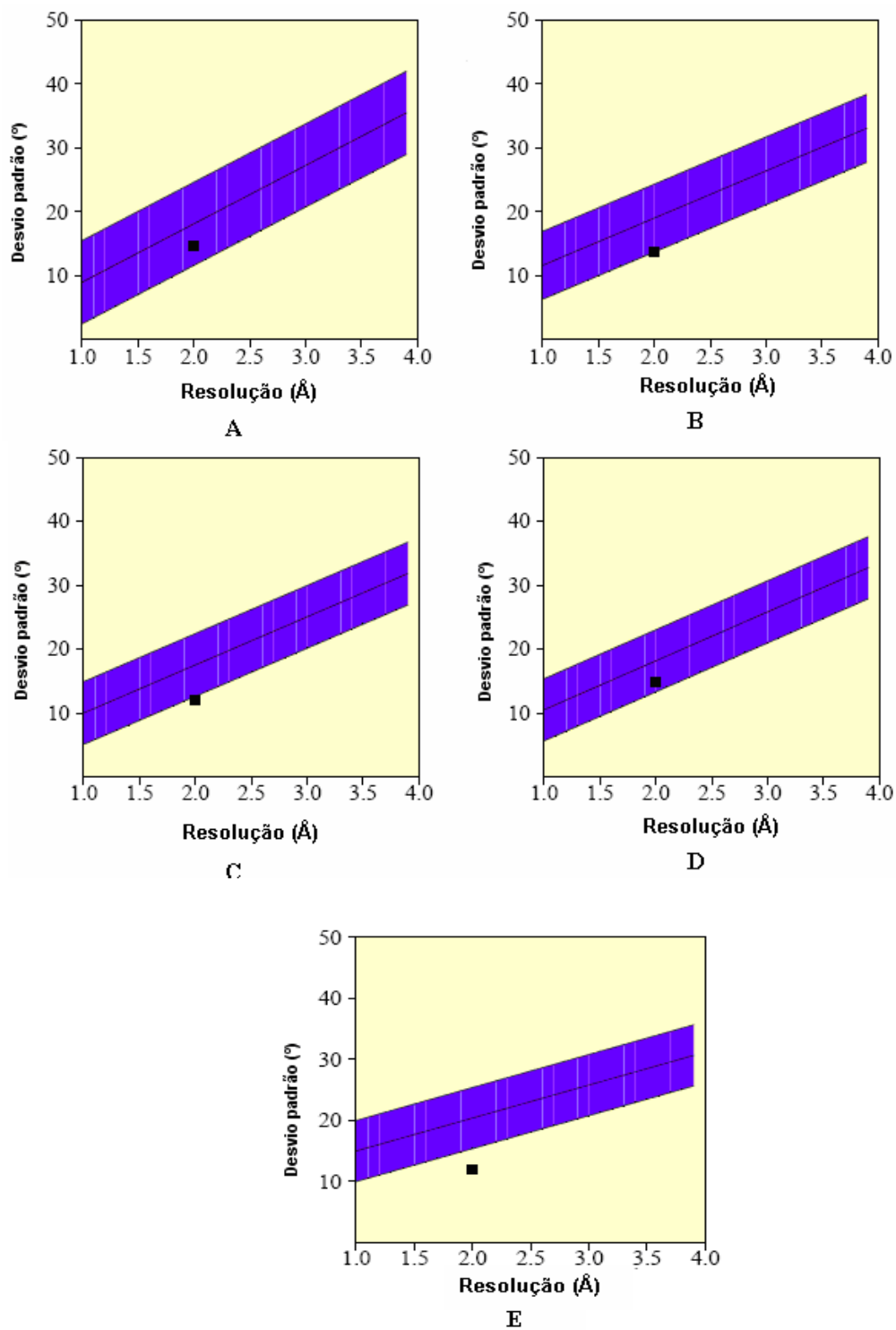




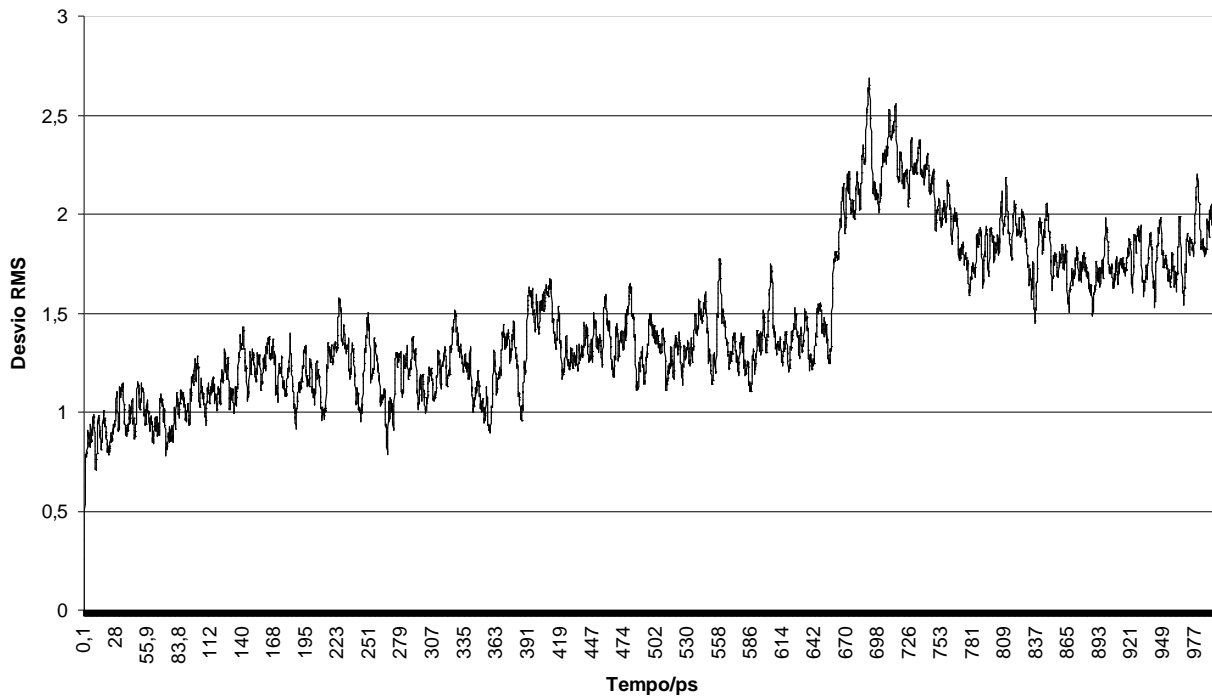
**Figura 34.** Gráficos Chi-1 e Chi-2 dos resíduos da estrutura mutante. Os números de resíduos são mostrados entre parêntese. As regiões sombreadas representam áreas favoráveis, com exceção às áreas com coloração vermelha, estas demonstram os resíduos com ângulo desfavorável.



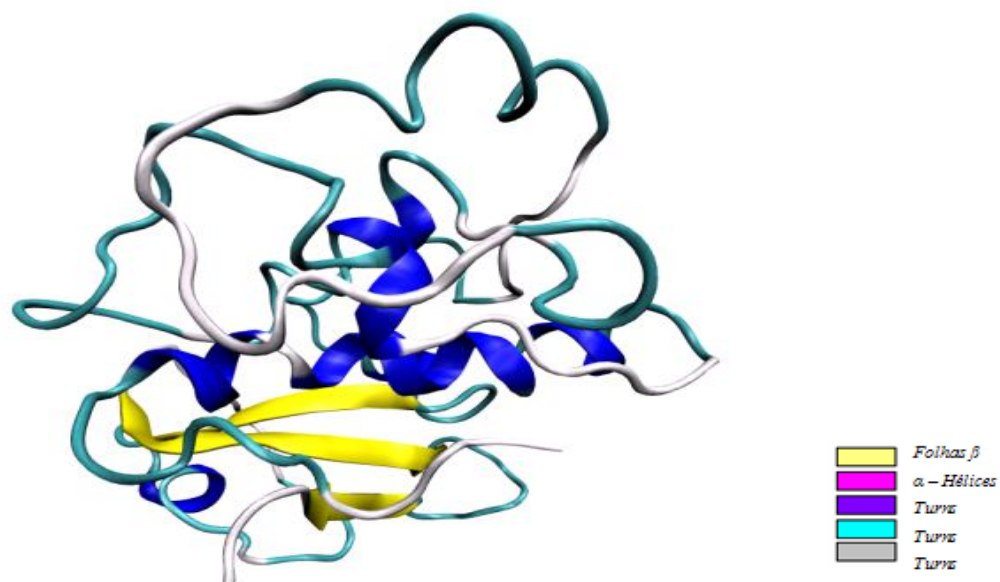
**Figura 35.** Qualidade da cadeia principal do modelo mutante da cisteína protease do *T. cacao*: (A) Avaliação do gráfico de Ramachandran; (B) Planaridade da ligação peptídica; (C) Interações ruins; (D) Distorção dos carbonos alfa; (E) Energia das ligações de hidrogênio.



**Figura 36.** Desvio padrão para os ângulos diedros Chi1 do modelo mutante da cisteína protease do *T. cacao*. (A) Conformação gauche menos; (B) Conformação trans; (C) Conformação gauche mais; (D) Somatório de todos os desvios padrões para Chi1; (E) Conformação trans para o ângulo torcional Chi2. O eixo das abscissas está descrito em graus, e o das coordenadas está descrito em Å.



**Figura 37.** Gráfico do RMS do modelo mutante da cisteína protease do *T. cacao*. Estabilidade estrutural alcançada a partir de 837ps.



**Figura 38.** Modelo mutante da cisteína protease do *T. cacao*.



## 6.7. A INTERAÇÃO DO MODELO MUTANTE COM METAIS

Qualquer alteração nas condições ambientais que possa influenciar ou reduzir a atividade enzimática de uma proteína deve ser analisado. O interesse sobre acumulação e toxicidade de metais tem crescido nos últimos anos como consequência das exposições ocupacionais e ambientais. Embora os mecanismos com que isso acontece estejam sendo elucidados a nível atômico somente nas últimas décadas (D'HAESE, 1995; MOREIRA, 2004). Com intuito de validar o modelo mutante como potencial agente quelante de íons metálicos, foram realizados experimentos de QM/MM no programa Gaussian 3,0 empregando os íons metálicos:  $Zn^{+2}$ ,  $Cu^{+1}$ ,  $Cu^{+2}$  e  $Cd^{+2}$  ao resíduo Cys742 da cisteína rotease do *T.cacao*. Para o cálculo da energia de ligação entre a estrutura e os íons foram realizados os seguintes procedimentos para análise dos sistemas: proteína sem ligante e somente o ligante, complexo ligante-proteína. A energia de ligação foi obtida através da equação descrita por (CASE, 2005) Figura 39 e Tabela 3.

$$E_{final} = E_{cpx} - (E_{lig} + E_{ptn}) \text{ onde:}$$

$E_{final}$ : Energia potencial final de ligação do complexo ligante-proteína;

$E_{cpx}$ : Média da energia potencial do complexo ligante-proteína;

$E_{lig}$ : Média da energia potencial somente do ligante;

$E_{ptn}$ : Média da energia potencial somente da proteína.

**Figura 39.** Equação da obtenção da energia de ligação ligante-proteína.

**Tabela 3.** Comparação entre os valores de energia do modelo mutante e sua interação com metais (kcal/mol).

| <b>Metais</b>          | <b>C. Protease</b>           | <b>C. Protease -<br/>Metal</b> | <b>Ligação</b>                 |
|------------------------|------------------------------|--------------------------------|--------------------------------|
| <b>Zn<sup>+2</sup></b> | <b>-274 x 10<sup>3</sup></b> | <b>-1389 x 10<sup>3</sup></b>  | <b>-857 x 10<sup>3</sup></b>   |
| <b>Cu<sup>+1</sup></b> | <b>-274 x 10<sup>3</sup></b> | <b>-30 x 10<sup>3</sup></b>    | <b>-1063 x 10<sup>3</sup></b>  |
| <b>Cu<sup>+2</sup></b> | <b>-274 x 10<sup>3</sup></b> | <b>-122 x 10<sup>3</sup></b>   | <b>-58916 x 10<sup>3</sup></b> |
| <b>Cd<sup>+2</sup></b> | <b>-274 x 10<sup>3</sup></b> | <b>-29 x 10<sup>3</sup></b>    | <b>-59801 x 10<sup>3</sup></b> |

Ao analisar as simulações de QM/MM para o complexo enzima-metal de ligação expresso na Tabela 3, percebeu-se que o sistema apresenta uma acentuada variação de energia de ligação entre os sistemas compostos pelos íons Zn<sup>+2</sup>, Cu<sup>+1</sup> e Cu<sup>+2</sup>. Esse comportamento, sugere uma instabilidade energética do complexo sistema constituído pela estrutura mutante da cisteína protease do *T. cacao* e os íons. Entretanto o resultado da energia de ligação para o experimento realizado com o Cd<sup>+2</sup>, apresentou valor energético favorável à estabilização do sistema testado. A energia de ligação apresentou-se negativa em -59801 x 10<sup>3</sup> kcal/mol que significa a quantidade de energia a ser liberada para que o sistema permaneça em equilíbrio (CASE, 2005). A energia de ligação negativa nesse caso indica que o sistema possui a estabilidade necessária para o estabelecimento de uma análise físico-química da interação do modelo mutante da cisteína protease do *T. cacao* com metais e embasar estudos mais complexos da interação dos metais com o modelo proposto, em particular com o Cd<sup>+2</sup>. Um elemento potencialmente tóxico podendo estar presente no ambiente, natural ou antropicamente através de indústrias de fundições, de processos de galvanoplastia, através de eletrodos de baterias, da fabricação de plásticos coloridos e pigmentos de tintas. Este metal causa impactos a saúde e meio ambiente, sendo comprovadamente um agente cancerígeno e teratogênico (BAIR, 2002).



## 7. CONSIDERAÇÕES FINAIS

Uma dos grandes avanços na Biologia Molecular do século XXI consistiu na elucidação de genomas de seres vivos. A predição estrutural de sequências de aminoácidos é função da Biologia Computacional que através de suas ferramentas algorítmicas tem sido fundamental para resolução de sequências de aminoácidos e/ou nucleotídeos em modelos estruturais 3D que favorece o desenvolvimento racional de estratégias para a produção de novos produtos. Uma vez determinado a conformação espacial de um receptor é possível prever qual a estrutura do novo ligante e/ou propor estruturas de proteínas mutantes que possam favorecer a sua aplicação industrial. Desta forma há um ganho de tempo significativo e uma redução nos custos das pesquisas. O presente trabalho teve como objetivo central a determinação estrutural da cisteína protease do *T. cacao* por Modelagem Comparativa.

O desenvolvimento do modelo 3D da cisteína protease do *T. cacao* por modelagem comparativa foi possível devido a identificação da proteína molde 1PCI|A que apresentou um satisfatório grau de homologia, permitindo a construção da proteína homóloga. Três modelos foram contruídos experimentalmente a partir de alterações em nível das *gaps* presentes entre o alinhamento das sequências. Entretanto, a partir do emprego dos métodos teóricos para construção e validação de estruturas construímos três moldes, e conseqüentemente um modelo protéico próximo da realidade.

As cisteínas proteases encontram-se amplamente distribuídas na natureza e representam um dos seis grupos de enzimas mais estudadas. A cisteína protease do *T. cacao* é uma enzima composta por estruturas alfa/beta/alfa, predominantemente composta por alfas hélices, assim como a maioria das papaínas descritas na literatura. O sítio catalítico da papaína esta representado pelos aminoácidos Cys e His com carga de +1 e -1 respectivamente. O fator que mais chamou a atenção nessa estrutura ocorreu durante a etapa da otimização do modelo, uma vez que, otimizações com números de ciclos superiores ao necessário conduziram as estruturas a “desnaturação” da conformação espacial.

A determinação estrutural do sítio catalítico da enzima por meio da proteína molde viabilizou a compreensão do mecanismo de catálise e hidrólises resultante da atividade cinética da enzima investigada. A determinação da estrutura 3D da cisteína protease do *Theobroma cacao* através de modelagem comparativa poderá contribuir para uma provável

aplicação tecnológica desta enzima favorecendo o fortalecimento da caucultura no Brasil, uma vez que foi performada análises sobre o sítio ativo da cisteína protease do *T. cacao*. Adicionalmente a descrição do modelo 3D desta enzima contribuiu para a construção de um modelo mutante que apresentou afinidade química com o cádmio, um elemento químico altamente tóxico gerado principalmente pelas atividades de mineração e industriais, tais como manufatura de baterias, produção de fertilizantes e estabilização de plásticos.

A contaminação do ambiente por metais pesados vem se tornando um problema global, sendo uma ameaça cada vez mais presente em muitos ecossistemas, especialmente próximo a áreas industriais e urbanas. Para conter a poluição causada por metais pesados são necessárias estratégias biotecnológicas que empreguem a teoria no desenvolvimento de produtos capazes de favorecer a resolução de problemas sociais. Neste contexto, se insere o modelo mutante da cisteína protease do *T. cacao*. Este poderia ser empregado para a fitorremediação/quelação dejetos industriais, de plásticos em geral, em indústrias que utilizem no seu processo de fabricação o cloreto de polivenio e pintura que contem pigmentos de cádmio, nas indústrias produtoras de baterias para transporte automotivo e aéreo, indústrias produtoras de baterias para computadores, aparelhos celulares, solda, freios de automóveis, gasolina, indústria do aço e ligas odontológicas. Na indústria alimentícia a biotecnologia poderia ser utilizada para reduzir os impactos dos resíduos gerados através da produção de embalagens de alimentos, pois o cádmio é usado como revestimento de utensílios a serem utilizados no preparo de alimentos, principalmente nos enlatados; a biotecnologia poderia ser usada também na indústria com atividade relacionada a produção de alimentos refinados: farinha de trigo branca e arroz branco e industrialização do chá e do café.

## 8. REFERÊNCIAS

ALLEN, M.P.; TILDESLEY, D.J. Computer simulations of liquids. **Oxford University Press**: London 1989.

ALTSCHUL, S.F.; GISH, W.; MILLER, W.; MYERS, E.W.; LIPAMAN, D.J. Basic local alignment search tool. **Journal Molecular Biology**. v. 215, p. 403, 1990.

ALVARENGA, P; AMAROLI, P.; CÁCERES, J.; EGUIZÁBAL, C.; FERNÁNDEZ, J. A.; FOWLER, W.; LAURIA, A.; FUENTES, H. L.; MELHADO, O. E.; PANAMENO E WALTER, K. História de El Salvador. **El Salvador: Ministério de Educación**; Centro América, n. 249, p. 49-52, 1941.

ANDEBRHAN, T.; FIGUEIRA, A.; YAMADA, M. M.; CASCARDO, J.; FURTEK, D. B. Molecular fingerprinting suggests two primary outbreaks of witches' broom disease (*Crinipellis Perniciosa*) of *Theobroma Cacao* in Bahia, Brasil. **European Journal of Plant Pathology**, n.105, p.167-175, 1999.

ANKERI, L.S.; JURIS, P.C. Quantitative structure-retention relationship studies of odor-active aliphatic compounds with oxygen-containing functional groups. **Analytical Chemistry**. v. 62, p. 2676-2684, 1990.

ARBAISAH, S.M.; ASBI, B.A.; JUNAINAH, A.H.; JAMILAH, B. Determination of optimum conditions for pectinesterase extraction from soursop fruit (*Anona Muricata*) using response surface methodology. **Food Chemistry**, v. 55, n.3, p. 289-292, 1996.

ATAUALPA A.C.; BRAGA H.M.; MORGON, N.H. Cálculos teóricos de afinidades por próton de n-alquilaminas usando o método ONIOM. **Química Nova**, v. 29, n. 2, p. 187-193, 2006.

BAIR, C. Environmental Chemistry. 2. ed. New York: Bookman, 2002.

BARRETT, A.J. Introduction: The classification of proteinases. **Ciba Found**, v. 75, p. 1-13, 1980.

BARRETT, A.J.; RAWLINGS, N.D.; WOESSNER, J.F.; JR. E.D.S. Handbook of proteolytic enzymes. **Elsevier**, London: Academic Press, 2004.

BASTOS, C.N. Epifitologia, hospedeiros e controle de vassoura-de-bruxa (*Crinipellis pernicios*a (Stahel) Singer), Ilhéus: Ceplac/Cepec, **Boletim Técnico**, n. 168, p. 21, 1990.

BASTOS, E. Cacau: a riqueza agrícola da América. São Paulo: **Ícone**, p.130, 1987.

BERMAN, H. M.; WESTBROOK, J.; FENG, Z.; GILLILAND, G.; BHAT, T. N.; WEISSIG, H.; SHINDYALOV, I. N.; BOURNE, P. E. The protein data bank. **Nucleic Acids Research**, v. 28, p. 235-242, 2000.

BERNSTEIN, H.J. 2000. Disponível em <http://www.OpenRasMol.org/Copyright.html.RASMOL>. Acesso em: 25 mar. 2007.

BLUNDELL, T.L. Comparative protein modelling by satisfaction of sapatinal restraints. **Journal Molecular Biology**. v. 234, p. 779-815, 1993.

BOBTELSKY, M.; JORDAN, J. The metallic complexes of tartrates and citrates, their structure and behavior in dilute solutions. **Journal of the American Chemical Society**. v. 67, n.1824, 1945.

BRANDEAU, J. El cacao. Barcelona: Blume, 1992.

BRANDEN, C.; TOOZE, J. Introduction to protein structure, Gerland: New York, 1991.

CAPRILES D.R.L.; REYS H.E. Contenido de polifenoles en dos variedades de *Theobroma Cacao L.* y su relacion con la resistencia a *Ceratocystis fimbriata*. **Agronomia Tropical**, v. 18, p. 339-355, 1968.

CAPRILES D.R.L.; SCHULZ, E.S.; MUNOZ, A. El contenido de acido clorogenico con diferentes variedades de cacao y su relacion con la resistencia contra el hongo *Ceratocystis fimbriata*. **Agronomia Tropical**, n. 16, p. 273-284, 1964.

CARVALHO, A.C.L.F. Computação Bioinspirada. Instituto de Ciências Matemáticas e Computacionais da Universidade de São Paulo, 2003.

CASCARDO, J.C.M.; PEREIRA, G.A.G. Biochemical changes during the development of witches' broom: The most important disease of cocoa in Brazil caused by *Crinipellis perniciosa*. **Journal of Experimental Botany**. V. 56, n. 413, p. 865-877, 2005.

CASE, D.A.; DARDEN, T. A.; CHEATHAM, T.E.; SIMMERLING, C.L.; WANG, J.; DUKE, R.E. LUO, R.; MERZ, K.M.; WANG, B.; PEARLMAN, D.A.; CROWLEY, M.; BROZELL, S.; TSUI, V.; GOHLKE, H.; MONGAN, J.; HORNAK, V.; CUI, G.; BEROZA, P.; SCHAFMEISTER, C.; CALDWELL, J.W.; ROSS, W.S.; KOLLMAN, P.A. **AMBER 9**: User manual. University of California, San Francisco, 2006, 320 p.

CASTRO, R.M. Perspectivas da indução de resistência em cacauzeiro contra *Crinipellis perniciosa* através do benzotiadiazole (Bth). **Fitopatologia Brasileira**. v. 25, p. 149-156, 2000.

CEPLAC. Comissão Executiva do Plano da Lavoura Cacaueira. Disponível Em: [Http://Www.Ceplac.Gov.Br/Radar/Radar\\_Cacau.Htm](http://Www.Ceplac.Gov.Br/Radar/Radar_Cacau.Htm). Acesso Em: 28 de Julho de 2008.

CLARK, T.A. A Handbook of Computational Chemistry; Wiley, New York, 1985.

COLE, G.T.; HUNG, C.Y. The parasitic cell wall of *C. immitis*. **Medical Mycology**. v. 39 (Suppl 1), p. 31-40, 2001.

COMPANHIA DAS DOCAS DO ESTADO DA BAHIA, 2002. **Apresenta em estatística, os principais produtos movimentados**. Disponível em: <http://www.cobeda.com.br/portoalheus>. Acesso em: 25 mar. 2007.

COSTA, O.V.; FONTES, L.E.F.; COSTA, L.M.; KER, J.C.; ARAÚJO, Q.R.; NACIF, P.G.S. Histórico de uso e ocupação das regiões pastoris dos municípios de Itapé e Itajú do Colônia - Bahia. In: Reunião Brasileira de Manejo e Conservação do Solo e da Água, 18, 1999, Ilhéus. Anais... Ilhéus: CEPLAC, 1999. CD Rom. den Landboun in Suriname, Paramaribo. p. 33-26.

CONRATH, U.; PIETERSE, M.J.; MAUCHI-MANI, B. Priming in plant-pathogen interaction. **Trends in Plant Science**. v. 7, n. 5, p. 210-216, 2002.

CRONSHAW, D.K. Fungicide application together with cultural practices to control cocoa disease in Ecuador. **Tropical Agriculture**, v. 56, n. 2, p. 165-171, 1979.

DAPPRICH, S.; KOMÁROMI, I.; BYUN, K.S.; MOROKUMA, K.; FRISCH, M.J.; **Journal Molecular Structure**. (THEOCHEM), v. 1, n. 461, 1999.

DDBJ. Disponível em: <<http://www.ddbj.nig.ac.jp/>>. Acesso em: 09. jan. 2009.

DIAS, L.A.S. **Melhoramento genético do cacauzeiro**. Editora Folha de Viçosa Ltda. 578 p., 2001.

DIXON, M.; WEBB, E.L. Enzyme kinetics. In: *Enzymes*. New York: Academic Press, 1979. p. 47.

ELLIOT, M.A.; KAROONUTHAISIRI, N.; HUANG, J.; BIBB, M.J.; COHEN, S.N.; KAO, C.M.; BUTTNER, M.J. The chaplins: a family of hydrophobic cell-surface proteins involved in aerial mycelium formation in *Streptomyces coelicolor*. **Genes & development**. v. 17, p. 1727-1740, 2003.

EMBL. Disponível em: <<http://www.ebi.ac.uk/embl/>>. Acesso em: 09 jan.2009.

EVANS, H.C.; SOLORZANO, G.R. Witches' broom disease: wrong experiment, right results. In International Cocoa Research Conference, p. 8, 1981. **Proceedings**. Cartagena, Colômbia, p. 415-418, 1982.

FORESMAN, J. B.; FRISCH, Æ. **Exploring Chemistry With Electronic Structure Methods 2**. Ed. Pittsburgh, Gaussian, Inc. 1996.

FRANCESCHI, V.R.; KAROKENE, P.; KREKLING, T.; CHRISTIANSEN, E. Phloem parenchyma cells are involved in local and distant defense responses to fungal inoculation or bark-beetle attack in norway spruce (Pinaceae). **American Journal Botany**. v. 87, n. 3, p. 314-326, 2000.

FRANCESCHI, V.R.; KREKLING, T.; BERRYMAN, A.A.; CHRISTIANSEN, E. Specialized phloem parenchyma cells in norway spruce (pinaceae) are a primary site of defense reactions. **American Journal Botany**. v. 85, p. 601-615, 1998.

FRIAS, G.A. Infection biology of *Crinipellis perniciosa* on vegetative flushes of cacao. **Plant Disease**. v. 75, p. 552-556, 1991.

FRISCH, Æ.; FRISCH, M.J.; TRUCKS, G.W. **Gaussian 03**: User's Reference. Gaussian Inc: Carnegie, 2003, 327p.

FUJITSU, 2002. **BioMedCaChe protein-drug discovery modeling software on windows**. Disponível em: <http://www.cachesoftware.com/biomedcache/index.shtml>. Acesso em : 23 mai 2007.

FUJITSU. **BioMedCache 6.1**: User Guide. Oxford: Fujitsu Limited, 2003, 200 p.

GENBANK. Disponível em: <<http://www.ncbi.nlm.nih.gov/Genbank/>> Acesso em: 09. jan.2009.

GENOME COMPARISON PROJECT. Disponível em: <[http://www.worldcommunitygrid.org/projects\\_showcase/archives/fgc/viewFgcOverview.do](http://www.worldcommunitygrid.org/projects_showcase/archives/fgc/viewFgcOverview.do)>. Acesso em: 09. jan. 2009.

GETHER, U.; KOBILKA, B.K. G. Protein-coupled receptors. II. Mechanism of agonist activation. **Journal of Biological Chemistry**. n. 273, p. 17979-17982, 1998.

GINALSKI, K.; GRISHIN, N.V., et al. Practical lessons from protein structure prediction. **Nucleic Acids Research**. v.33, n.6, p.1874-1891. 2005.

GOLDSMITH-FSCHMAN, S.; HONIG, B. Structural genomics: computacional methods for structure analyses. **Protein Science**. v. 12, p. 1813-1821, 2003.

GOODAY, G. W. Physiology of microbial degradation of chitin and chitosan. **Biodegradation**. v. 1, p. 177-190, 1990.

GRIFFITH, G. W. et al. Autecology and evolution of the with' broom patogen (*Crinipellis pernicioso*) of cocoa. In: Blakeman, J. P.; Williamison, B. **Ecology of Plant Pathogens**. Oxon: CAB International, 1994, p. 245-267.

GRIFFITH, R.; TRACY, T. Antifungal drugs. In: WILLIAMS, D. A.; LEMKE, L. Foye's principles of medicinal chemistry. 5. ed. Philadelphia. Lippincott Williams & Wilkins, 2002, p. 891-903.

GROVES, M.R.; TAYLOR, M.A.J., SCOTT, M.; CUMMINGS, N.J.; PICKERSGILL, R.W.; JENKINS, J.A. The prosequence of procaricain forms an  $\alpha$ -helical domain that prevents access to the substrate-binding cleft. **Structure**. v. 4, n. 10, p. 1193-1202, 1996.

GUEX, N.; PEITSCH, M. C. Swiss-Model And The Swiss-Pdbviewer: An Environment For Comparative Protein Modeling. **Electrophoresis**, n.18, p. 2714-2723, 1997.

HAGEN, J.B. The origins of bioinformatics. **Nature Reviews Genetics**. v.1, n.3, p.231-236, 2000.

HAMMOND-KOSAK, K.; JONES, J.D.G. **Responses to plant pathogens**. In: Buchanan, B.; Gruissem, W.; Jones, R., eds. *Biochemistry and Molecular Biology of Plants*, American Society of plant Physiologists, p. 1102-1156, 2000.

HENSEN, C. A combined QM/MM approach to protein-ligand interactions: polarization effects of the HIV-1 protease on selected high affinity inhibitors. *Journal of Medicinal Chemistry*, n. 47, p. 6673-6680, 2004.

HERSCOVICS, A.; ORLEAN, P. Glycoprotein biosynthesis in yeast. **FASEB Journal**, n. 7, p. 540-550, 1993.

HIGGINS, D.; TAYLOR, W.R. *Bioinformatics sequence, structure, and databanks: a practical approach*. Oxford: Oxford University Press, 2000.

HIRST, D.M.; *A Computational Approach to Chemistry*; Blackwell Scientific Publications, Oxford, 1990; p 85.

HOLMES, K.A., SCHROER, H.J., THOMAS, S.E., EVANS, H.C., KRUPASAGAR, V.; SEQUEIRA, L. Auxin destruction by *Marasmius perniciosus*. **American Journal of Botany**. v. 56, p. 390-397, 1969.

HÖLTJE, H. D. et al. Introduction to comparative protein modeling. In: **Molecular Modeling**: basics principles and applications. 2. ed. Germany: Wiley-VCH GmbH, 2003, p. 87-144.

HÖLTJE, H.D.; FOLKERS, G. Em: *Molecular Modeling: Basic Principles and Applications*; Mannhold, R.; Kubinyi, H.; Timmerman, H., eds.; VCH: Weinheim, 1997.



HUMPHREY, W.; DALKE, A.; SCHULTEN, K. VMD: Visual Molecular Dynamics, **J. Molecular Graphics**. v. 14.1, p. 33-38, 1996.

INOUE, A. G. O.; KIYOSHI, M. Laboratório de Bioinformática do Projeto Genoma Funcional e Diferencial do Fungo *Paracoccidioides brasiliensis* - Projeto Genoma Pb - Universidade de Brasília, 2002.

JORGENSEN, W.L.; TIRADO-RIVES, J.J. **Journal of the American Chemical Society**. n. 110, p. 1657, 1998.

KABSCH, W.; SANDER, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. **Biopolimers**, v. 22, p. 2577-2637, 1983.

KABSCH, W.; SANDER, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. **Biopolimers**, v. 22, p. 2577-2637, 1983.

KARADAKOV, P. B.; MOROKUMA, K.; **Chemical Physics Letters**. n. 317, p. 589, 2000.

KLEBE G. Recent developments in structure-based drug design. **Journal of Molecular Medicine**. v. 78, p. 269-281, 2000.

KOHN, W.; SHAM, L.J. Self-consistent equations including exchange and correlation effects. **Physical Review**. v. 140, p. 1133, 1965.

KRUPASAGAR, V. Auxin destruction by *Marasmius perniciosus*. (M.S. Thesis). Madison. University of Wisconsin. 1968.

KÜES, U. Life History and developmental processes. In: The Basidiomycete *Coprinus cinereus*. **Microbiology and Molecular Biology Reviews**. v. 64, p. 316-353, 2000.

KUNAKBAEVA, Z.; CARRASCO, R.; ROZAS, I. An approximation to the mechanism of inhibition of cysteine proteases: Nucleophilic Sulfur addition to michael acceptors type compounds. **Journal of Molecular Structure**. v. 626, p. 209-216, 2003.

KUNDROTAS, P.J.; ALEXOV, E. Predicting 3D structures of transient protein-protein complexes by homology. **Biochimica et Biophysica Acta**. 2006.

LAEMMLI, U. K. Nature cleavage of structural proteins during assembly of head of bacteriophage-t4. **Nature**. v. 227, n. 5259, p. 680, 1970.

LANA, T. G. Caracterização genética e fisiológica de *Crinipellis pernicioso*. 2000. 106p. Tese (doutorado) – Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba.

LASKOWSKI, R. A.; MACARTHUR, M. W.; MOSS, D. S.; THORNTON, J. M. PROCHECK: A Program to Check the Stereochemical Quality of Protein Structures. **Journal Applied Crystal Structure**. v. 26, p. 283-291, 1993.

LAWRENCE, J. S., CAMPÊLO, A. M. F. L.; FIGUEIREDO, J. M. Enfermidades do cacauero: doenças fúngicas que ocorrem nas folhas, ramos e tronco. **Agrotrópica** (Brasil). v. 3, n. 1, p. 1-14.

LEHNINGER, A.L.; NELSON, D.L.; COX, M.M. Princípios de Bioquímica. 2. ed. São Paulo: Reimpressão, 2000.

LINDAHL, E.; HESS, B.; SPOEL, D.V. (2001) Gromacs 3.0: a package for molecular.

LTSCHUL, S.F.; GISH, W.; MILLER, W.; MYERS, E.W.; LIPAMAN, D.J. Basic Local Alinmente Search Tool. **Journal Molecular Biology**, v. 215, p. 403, 1990.

LYBRAND, T.P. Computer Simulation of Biomolecular Systems Using Molecular Dynamics and Free Energy Perturbation Methods. In: LIOKWITZ, K. B.; BOYD, D. B. Reviews in computacional chemistry. United States of America: VCH Publishers, 1990, p. 295-320.

MACDONALD, N.A.; JORGENSEN, W.L.J. **Chemical Physics**. v. 102, p. 8049, 1998.

MAJERLE, A., JERALA, R. Protein inhibitors form complexes with procathepsin .1 and augment cleavage of the propeptide. **Archives of Biochemistry and Biophysics**. v. 417, p. 5358, 2003.

MARINHO, P.L. O estado e a economia cacauero da Bahia. **Anais Eletrônico...** Congresso brasileiro de história econômica, 3. São Paulo, 2001.

MARTÍNEZ, M.; LÓPEZ-SOLANILLA, E.; RODRÍGUEZ-PALENZUELA, P.; CARBONERO, P.; DÍAZ, I. Inhibition of plant-pathogenic fungi by the barley cystatin Hv-CPI (Gene Icy) is not associated with its cysteine-proteinase inhibitory properties. **MPMI**. v. 16, n. 10, p. 876–883, 2003.

MASERAS, F.; MOROKUMA, K.; On the application of the IMOMO (integrated molecular orbital + molecular orbital) method. **Journal of Computational Chemistry**. v. 16, p. 1170, 1995.

MATSUBBARA, T.; SIEBER, S.; MOROKUMA, K. Quantum Chemistry. **International Journal Quantum Chemistry**. v. 60, p.1101, 1996.

MAZAL, S.; BELENGHI, B.; DELLEDONNE, M.; MENACHEM, E.; LEVINE, A. The involvement of cysteine proteases and protease inhibitor genes in the regulation of programmed cell death in plants. **Plant Cell**. v. 11, p. 431-444. 1999.

MCKERROW, J.H.; ENGEL, J.C.; CAFFREY, C.R. Cysteine protease inhibitors as chemotherapy for parasitic infections. **Bioorganic and Medicinal Chemistry**. v. 7, p. 639-644, 1999.

MELO, F.; FEYTMANS, E. "Novel knowledge-based mean force potential at atomic level". **Journal of Molecular Biology**. n. 267, p. 207-222, 1997.

MOTTRAM, J.C., HELMS, M.J., COOMBS, G.H., SAJID, M. CLAN. Cysteine peptidases of parasitic protozoa. **Trends In Parasitology**. v. 19, p. 182-187, 2003.

NIELLA, G.; RESENDE, M.L.; CASTRO, H.A. aperfeiçoamento da metodologia de produção artificial de basidiocarpos de *Crinipellis pernicioso*. **Fitopatologia Brasileira**. v. 24, p. 523-527, 1999.

OLIVA, G. Bioinformática na Medicina. **Gazeta Médica Bahia**. v. 78 (Suplemento 1), p. 52-58, 2008.

OLIVEIRA, D. C. Alinhamento de Sequências: Monografia: Departamento de Ciência da Computação da Universidade Federal de Lavras, 2002.

OUZOUNIS, C.A. Bioinformatics and the theoretical foundations of molecular biology. **Bioinformatics**, v.18, n.3, p.377-378, 2002.

OUZOUNIS, C.A.; VALENCIA, A. Early bioinformatics: the birth of a discipline – a personal view. **Bioinformatics**, v.19, n.17, p. 2176-2190, 2003.

PATRICK, G.L. An introduction to medicinal chemistry. Oxford University Press Inc., New York, 2001.

PAYNE, M.C.; TERER, M.P.; ALLAN, D.C.; ARAIS, T.A.; NOPOULOS, J.D. **Reviews of Modern Physics**. n. 64, p. 1045, 1992.

PEARSON, W.R. (1998). Empirical statistical estimates for sequence similarity searches. . v. 276, p. 71-84, 1998.

PEREIRA, J.L. et al. First occurrence of witches' broom disease in the principal cocoa-growing region of Brazil. **Tropical Agriculture**. v. 67, n. 2, p. 188-189, 1990.

PERLMAN, D.S; CASE, D.A.; CALDWELL, J.W; ROSS, W.S; CHEATHAM III, T.E; DEBOLT, S; FERGUSON, D; SIEBEL, G. KOLLMAN, P. **Computer Physics Communications**. n. 1, p. 911, 1995.

PHARDATARE, S.U.; DESHPANDANDE, V.V.; SRINIVASAN, M.C. High activity alkaline protease from *Conidiobolus coronatus* (NCL 86.8.20): Enzyme production and compatibility with commercial detergents. **Enzyme and Microbial Technology**. n. 15, p. 72-76, 1993.

PICKERSGILL, R.W., RIZKALLAH, P., HARRIS, G.W. & GOODENOUGH, P.W. Determination of the structure of papaya protease omega. **Acta Cryst. B**. v. 47, p. 766–771, 1991.

POMPE-NOVAK, M.; POLSAK-PRIJATELJ, M.; POPOVIC, T.; STRUKELJ, B., RAVNIKAR, M. The impact of potato cysteine proteinases in plant growth and development. **Physiological and Molecular Plant Pathology**. v. 60, p. 7178, 2002.

POMPEO, F. et al. Dissection of the bifunctional *Escherichia coli* N-Acetylglucosamine –1-phosphate uridylyltransferase enzyme into autonomously functional domains and evidence that

trimerization is absolutely required for glucosamine-1-phosphate acetyltransferase activity and cell growth. **The Journal of Biological Chemistry**, v.276, p.3833-3839, 2001.

POPLE, J.A.; BEVERIDGE, D.L. Approximate Molecular Orbital Theory, McGraw-Hill, New York, 1970.

PURDY, L.H.; SCHMITDT, R.A. Status of cacao witches' broom: biology, epidemiology and management. **Annual Review of Phytopathology**, v. 34, p. 573-6794, 1996.

RANGEL, J. F. Ceplacicaca V Ano 25. Brasilia, Iica. Desenvolvimento Institucional, n. 16. p. 1-L 1, 1982.

RAST, D. M. et al. Cell wall-associated in fungi. **Phytochemistry**, v. 64, p. 339-366, 2003.

RAO, M.B.; TANKSALE, A.M.; DESHPANDE, V.V. Molecular and biotechnological aspects of microbial proteases. **Microbiology and Molecular Biology Reviews**. p. 597-635, 1998.

RAWLINGS, N.D.; BARRETT, A.J. Merops: The Peptidase Database. **Nucleic Acids Research**. n. 28, p. 323-325, 2000.

REEVES, G.A.; THORNTON, J.M. Integrating biological data through the genome. **Human Molecular Genetics**. v.15, n. 1, p. R81-87. 2006.

RESENDE, M.L.V.; COOPER, R.M.; FLOOD, J.; ROWAN, M.G.; BEALE, M.H.; RAMSDEN, J.D. involvement of condensed tannins and phytoalexins in the resistance of cocoa (*Theobroma cacao* L.) to verticillium wilt. **Physiological And Molecular Plant Pathology**. v, 48 p.347-359. 1996.

RESENDE, M.L.V.; NOJOSA, G.B.A.; AGUILAR, M.A.G.; SILVA, L.H.C.P.; NIELLA, G.R.; CARVALHO, G.A.; GIOVANINI, R. Responsable de la resistance induite. **Cafe Cacao The**. v. 25, p. 181-190, 1991.

RINO, JP.; STUDART, N. An interaction potential for materials and molecular dynamics simulations. **Química Nova**. v. 24, n. 6, 2001.

ROCHA, G.B.; Freire, R.O.; Simas, A.M.; STEWART, J.J.P. RM1: A reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I. **Journal of Computational Chemistry**. v. 27, n. 10, p. 1101–1111, 2006.

ROCHA, H.M.; WHEELER, B.E.J. Factors influencing the production of basidiocarps and the deposition and germination of basidiospore of *Crinipellis pernicioso*, the causal fungus of witches' broom on cacao (*Theobroma Cacao L.*). **Plant Pathology**. v. 34, p. 319-328, 1985.

RODRIGUES, A.G. Descripción y medidas de control de la escoba de Bruja. **El Cacaotero Colombiano**, n. 24, p. 38-41, 1983.

RONCHETTI, M.; JACUCCI, G. Applicability of Nosé isothermal reversible dynamics. **Physical Review A**. v. 48, p. 1726–1737, 1993.

ROSA, I.S. Enxertia do cacauero. Ilhéus: CEPALC/SUBES/CEPEC, 1998, 42p

RSCB, 2007. RSCB Protein Data Bank. Disponível em: <http://www.rcsb.org/pdb/>. Acesso em: 23 mar, 2008.

RUBINI, M.R.; RIBEIRO, S.R.T; POMELLA, A.W.V; MAKI, C.S; ARAÚJO, W.L; SANTOS, D.R.; AZEVEDO, J.L. Diversity of endophytic fungal community of cacao (*Theobroma cacao L.*) and biological control of *Crinipellis pernicioso*, causal agent of witches' broom disease. **International Journal of Biological Sciences**. v.1 p. 24-33, 2005.

RYALS, J.A.; NEUENSCHWANDER, U.H.; WILLITS, M.G.; MOLINA, A.; STEINER, H.I.; HUNT, M.D. Systemic acquired resistance. *The Plant Cell*. n.8, p. 1809-1819, 1996.

SAID, S.; PIETRO, R. Enzimas de interesse industrial e biotecnológico. Ed. Eventos. p. 121, 2002.

ŠALI, A.; BLUNDELL, T.L. Comparative protein modelling by satisfaction of sapatinal restraints. **Journal Molecular Biology**. v. 234, p. 779-815, 1993.

SÁNCHEZ, E., SOTO, J.M., GARCIA, P.C., LÓPEZ-LEFEBRE, L.R., RIVERO, R.M., RUIZ, J.M. & ROMERO, L. Phenolic compounds and oxidative metabolism in green bean plants under nitrogen toxicity. *Australian Journal of Plant Physiology* 27:973-978. 2000.

SÁNCHEZ, R.; SALI, A. Comparative protein modelling in genomics. **Journal of Computacional Physics**. n. 151, p. 388-401, 1999.

SANTOS FILHO, O.A.; ALENCASTRO, R.B. modelagem de proteínas por homologia. **Química Nova**, v. 26, n. 2, p. 253-259, 2003.

SANTOS JUNIOR, M.C. et al. Modelling 3d-Structure of Proteins involved in chitin metabolic pathway in *Crinipellis pernicioso* In: Brazilian Symposium On Medicinal Chemistry. Current trends in drug discovery and development, 2., Rio De Janeiro. Abstract Book, P. 72, Rio de Janeiro, 2004.

SANTOS-FILHO, O.A.; ALENCASTRO, R.B. Modelagem de proteínas por homologia. **Química Nova**. v.26, n. 2, 2003.

SCARPARI, L.M.; MEINHARDT, L.W.; MAZZAFERA, P; POMELLA, A.W.V.; SCHIAVINATO, M.A.; SIEGEL, B.Z. Plant Peroxidases - An Organismic Perspective. **Plant Growth Regulation**. n. 12, p. 303-312, 1993.

SEGAL, J.A. Semi-empirical Methods of Electronic Structure Calculation. Plenum Press, New York, 1977.

SHERAZ, G.; SYEED, H.; THOMAS M.P.; RESMINI, M.; VERMA C.S; THOMAS E.W BROCKLEHURST, K. Generation of nucleophilic character in the Cys25/His159 ion pair of papain involves Trp177 but not Asp158. **Biochemistry**. v. 47, p. 2025-2035, 2008.

SILVA, S.DM.V.; LUZ, E.D.M.N.; GRAMACHO, K.P.; ALBURQUERQUE, P.S.B.; BEZERRA, J.L.; Strandardization of the belt system method to assess cacao genotypes to witches' broom resistance. In: **Proceedings 13<sup>th</sup> international Cocoa Research Conference**, Kota Kinabalu, Malaysia, 2000.

SOBERANIS, W. Increased frequency of phytosanitary pod removal in cocoa (*Teobroma cacao* L.) increases yied economically in eastern Peru. **Crop Protetion**. v. 18, p. 667-685, 2000.

SOLORZANO, G. R. 1977. Factores ambientales involucrados em la producción de basidiocarpos y la diseminación de basidiosporos de *Marasmius perniciosus* *Stahel*. Tesis Ing. Agr. Ecuador. Universidad De Manabí.

STAHEL, G. *Marasmius Perniciosus* Nov. Spec., the cause of the krulloten disease of cacao in Suriname. Transl. AMW ter lag, **Departament Van den Landbouw**, Suriname. Bull. n. 25, p. 25, 1915. (Translated from Geerman).

SVENSSON, M.; HUMBEL, S.; FROESE, R.D.J.; MATSUBARA, T.; SIEBER, S.; MOROKUMA, K.; J. "ONIOM: a multilayered integrated MO+MM method for geometry optimizations and single point energy predictions. A test for Diels-Alder reactions and Pt(P(t-Bu)<sub>3</sub>)<sub>2</sub> + H<sub>2</sub> oxidative addition". **Chemical Physics**. v. 100, n. 50, p.19357-63, 1996.

TERVIZAN, S.D.P. Mudanças no sul da Bahia associadas a vassoura-de-bruxa do cacao. In: Internetal Cocoa Research Conference, 12, Salvador. 1996. Proceedings, Lagos: Cocoa Producers Alliance, 1996. p. 1109-1116.

TOMPSON, J.D.; HIGGINS, D.; GIBSON, T.J. Clustal W: improving the sensitivity of progressive multiple alignment through sequence weighting, position-specific gap penalties and Weight Matrix Choice. **Nucleic Acids Research**. v. 22, n. 22, p. 4673-4680, 1994.

TRABULSI, L. R.; CAMPOS, L. C. Generalidades sobre enterobactérias. In: TRABULSI, L. R. et al. **Microbiologia**. São Paulo: Atheneu, 1999. cap. 27, p. 207-213.

VREVEN, D. 3D shape discrimination using relative disparity derivatives. **Vision Research**, v. 46, n. 25, p. 4181-92, 2006.

WENT, F.A.F.C. Krulloten en versteende vruchten van de cacao in Suriname, Amsterdam. Verhandelingen Der K. **Academic Van Wetenschappen**. v. 2, n. 10, p. 40, 1904.

WHEELER, B.E.J.; MEPSTED, R. Pathogenic variability amongst isolates of *Crinipellis perniciosus* from cocoa (*Theobroma cacao*). **Plant Pathology**. v. 37, p. 475-488, 1998.

WOO, T.K.; CAVALLO, L.; ZIEGLER, T.; Tool of the trade in modeling inorganic reactions from balls and sticks to HOMO's and LUMO's. **Theoretical Chemistry**. n. 100, p. 307, 1998.

WORKMAN, R.L.; FEARING, H.W. Potential model calculation of proton-proton bremsstrahlung using the Paris potential. **Physics Review C**. v. 34, p. 1158-64, 1986.

YAMADA-OKABE, T.; et al. Identification and characterization of the genes for N-Acetylglucosamine kinase and N-Acetylglucosamine-phosphate deacetylase in the pathogenic fungus *Candida albicans*. **European Journal of Biochemistry**, v. 268, p. 2498-2505, 2001.