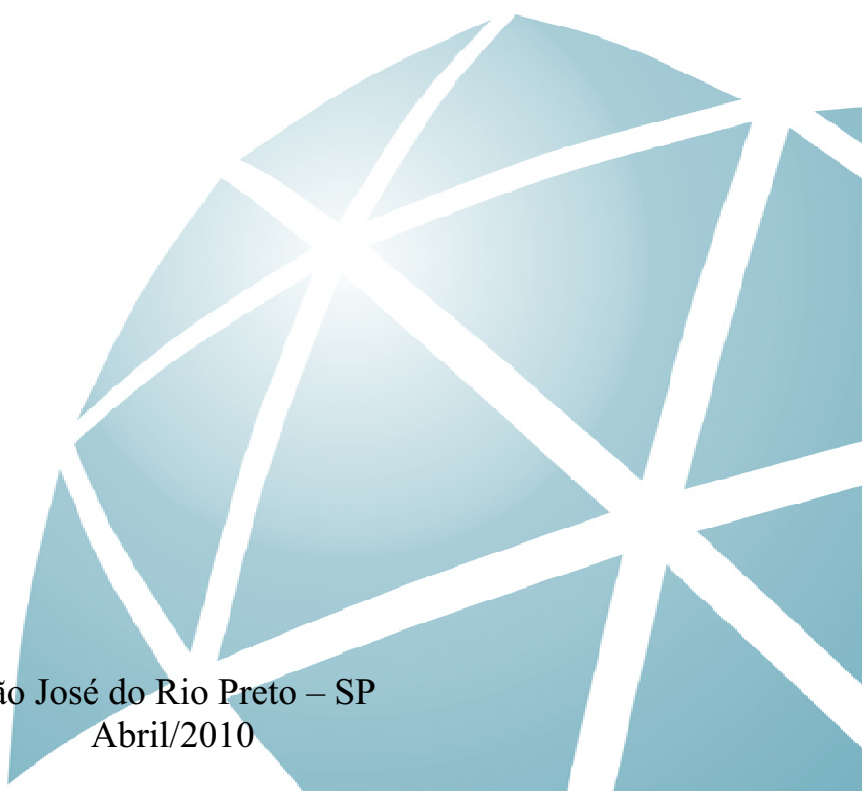




Lílian Hiromi Tomonari Yamasaki

**Análise por ferramentas de bioinformática da proteína
não-estrutural 5A do vírus da hepatite C genótipo 1 e 3
em amostras pré-tratamento**

São José do Rio Preto – SP
Abril/2010



Lílian Hiromi Tomonari Yamasaki

**Análise por ferramentas de bioinformática da proteína
não-estrutural 5A do vírus da hepatite C genótipo 1 e 3
em amostras pré-tratamento**

Trabalho de Dissertação apresentado como parte das atividades para a obtenção do título de Mestre do Programa de Pós-Graduação em Microbiologia do Instituto de Biociências, Letras e Ciências Exatas – Universidade Estadual Paulista "Júlio de Mesquita Filho"

Área de concentração: Virologia

Orientadora: Prof^ª.Dr^ª. Paula Rahal

Co-orientadora: Dr^ª. Helen Andrade Arcuri

São José do Rio Preto – SP
Abril/2010

Lílian Hiromi Tomonari Yamasaki

Análise por ferramentas de bioinformática da proteína não-estrutural 5A do vírus da hepatite C genótipo 1 e 3 em amostras pré-tratamento

Trabalho de Dissertação apresentado como parte das atividades para a obtenção do título de Mestre do Programa de Pós-Graduação em Microbiologia do Instituto de Biociências, Letras e Ciências Exatas – Universidade Estadual Paulista “Júlio de Mesquita Filho”.

Os componentes da banca de avaliação, abaixo listados, consideram este trabalho aprovado.			
	Nome	Instituição	Assinatura
1	Profa. Dra. Paula Rahal	UNESP/IBILCE	
2	Profa. Dra. Fernanda Canduri	USP/IQSC	
3	Prof. Dr. Carlos Alberto Montanari	USP/IQSC	

Yamasaki, Lílían Hiromi Tomonari.

Análise por ferramentas de bioinformática da proteína não-estrutural 5A do vírus da hepatite C genótipo 1 e 3 em amostras pré-tratamento / Lílían Hiromi Tomonari Yamasaki. – São José do Rio Preto : [s.n.], 2010. 83 f. : il. ; 30 cm.

Orientadora: Paula Rahal

Co-orientador: Helen Andrade Arcuri

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

1. Virologia. 2. Hepatite por vírus. 3. Hepatite C. 4. Vírus da hepatite C. 5. Proteína não-estrutural. 6. Bioinformática. estrutural. I. Rahal, Paula. II. Arcuri, Helen Andrade. III. Universidade Estadual Paulista, Instituto de Biociências, Letras e Ciências Exatas. IV. Título.

CDU - 578

Ficha catalográfica elaborada pela Biblioteca do IBILCE

Campus de São José do Rio Preto - UNESP

Autorizo a reprodução xerográfica para fins de pesquisa.

São José do Rio Preto, ____/____/____

Assinatura

Dedico este trabalho aos meus pais que me ensinaram todos os valores para conseguir mais esta conquista e mostraram que sempre vale a pena ir atrás daquilo que deseja. Em especial, dedico aos meus avós Ken e Takae que foram os mais entusiasmados quando eu iniciei o mestrado, mas não puderam ver a conclusão deste.

AGRADECIMENTOS

- À minha orientadora Profa. Dra. Paula Rahal pela coordenação deste trabalho e por toda a confiança depositada em mim para conduzi-lo e por todas as sugestões, elogios e críticas que permitiram grande amadurecimento profissional e pessoal;
- À co-orientação da Dra. Helen Andrade Arcuri por se dedicar a me instruir em uma área nova para mim, e me transformou em mais uma fiel seguidora da Bioinformática;
- Ao Centro de Estudo em Insetos Sociais (UNESP/Rio Claro) e à Faculdade de Medicina da USP por permitirem que eu utilizasse os computadores em minhas análises;
- À minha “ex-co-orientadora” Ana Carolina Gomes Jardim e à Cintia Bittar que sempre me ajudaram muito tanto desde a iniciação científica;
- Aos colegas do Laboratório de Estudos Genômicos Lenira, Paola, Juci, Marília, Érica, Marina, Tasso, Bruno, Carolzinha Bonfim, Natália, Renata, André, Plínio e Ana Cláudia por todo companheirismo e amizade;
- Aos funcionários da seção de Pós-graduação deste Instituto pela dedicação e eficiência no trabalho;
- Ao Programa de Pós-Graduação em Microbiologia por me aceitar como mais uma integrante em sua equipe de discentes;
- Aos professores das disciplinas cursadas durante o curso de mestrado pelos ensinamentos passados;
- Ao IBILCE por fornecer o espaço físico e à CNPq pelo apoio financeiro para a realização deste trabalho;
- Aos meus colegas de graduação, a famosa Xis-elésima ii Turma, que estavam presentes em Rio Preto durante o mestrado e à aqueles que sempre me apoiaram mesmo de longe, em especial ao Paulo, Carol Viegas, Nadia, Ana Z., Bruna, Paulinha, Cris, Gabriel e Élen;

- Aos integrantes do Laboratório BETA, principalmente à Carmelia, Nádia e Mariellen por sempre me receberem em seu laboratório com café fresco, guloseimas e “amizade remunerada”;
- Às moradoras da república que resido, tanto as que foram (Aline + Sofia, Larissa, Thaísa, Jana, Paulinha, Cris e Cíntia) quanto aquelas que continuam (Bruna, Marcela, Mari, Tuany e Livia) por permitirem que minha estadia aqui em Rio Preto fosse tão agradável;
- Aos bichos de estimação da república, Snoopy, Zara, Kate, Raj, Katezinha 1, Katezinha 2, Katezinho 1 e Katezinho 2 por deixarem eu sempre “judiarem” deles;
- Aos meus amigos de Jacareí que mesmo longe estão torcendo por mim e quando reecontro eles me recebem como se nunca tivéssemos nos separado;
- Aos moradores de todas as repúblicas, casas e apartamentos que sempre me abrigaram durante as minhas viagens para reuniões científicas, disciplinas em outros institutos, cursos, simpósios, etc. por permitirem que eu crescesse intelectualmente, mesmo com orçamento curto;
- Aos meus pais, minhas irmãs, primos, tios e avós por todo o apoio e união que permitiu que enfrentássemos uma fase tão triste sem abalar o andamento do mestrado;
- A todos aqueles que eu não citei, mas que participaram e/ou que participam ainda da minha vida, me apoiando sempre de alguma forma mesmo nos gestos mais pequenos como convidar para sair ou conversar às vezes, e me ajudaram a passar pelos momentos difíceis.

RESUMO

A infecção pelo vírus da Hepatite C (HCV) é considerada um grande problema de saúde pública, desde a sua descoberta em 1989. Entretanto a terapia mais utilizada atualmente, baseada no uso de Peginterferon, tem sucesso em aproximadamente 50% dos pacientes com o genótipo 1. Embora os mecanismos envolvidos nesta resistência viral ainda não sejam esclarecidos, sugere-se que fatores virais e do hospedeiro participam deste. A proteína não-estrutural 5A (NS5A) está envolvida em diversos processos celulares e é um componente essencial para o HCV. Entretanto, sua estrutura e função ainda não foram bem elucidadas. A partir destes fatos, os objetivos do presente estudo foram elaborar um modelo teórico da NS5A e investigar as propriedades estruturais e funcionais *in silico*. Foram analisadas 345 sequências da proteína NS5A do HCV de 23 pacientes infectados com o genótipo 1 ou 3. As composições de aminoácidos e de estrutura secundária demonstraram que há diferença entre os genótipos, podendo indicar que há diferenças nas interações proteína-proteína entre os genótipos, o que pode estar relacionado com a diferença da taxa de resistência ao tratamento. A análise funcional foi realizada com o *ProtFun*, que sugeriu que a NS5A estaria envolvida nas funções celulares de metabolismo intermediário central, tradução, crescimento, transporte, ligação e hormônio. Estas funções variaram entre os domínios, suportando a hipótese de que a NS5A é uma proteína multifuncional. A análise pelo PROSITE indicou vários sítios de glicosilação, fosforilação e miristoilação, que são altamente conservados e podem ter função importante na estabilização da estrutura e função, sendo assim possíveis alvos de novos antivirais. Alguns deles estão em regiões relacionadas com a resposta ao tratamento. Outro sítio altamente conservado encontrado foi o sítio de ligação à integrina (RGD), relacionado com a localização nuclear e perinuclear de outras proteínas e pode estar ligada com formas nucleares da NS5A. Por meio da técnica de *threader*, modelagem molecular comparativa e dinâmica molecular foi elaborada a estrutura 3D da NS5A completa, que poderá ser utilizado posteriormente para outros estudos. A análise das alterações de estrutura/função de NS5A mutantes, mostrou várias alterações estruturais e/ou bioquímicas atuando na perda de função da proteína.

Palavras-chave: Bioinformática Estrutural, Hepatite C, HCV, NS5A

ABSTRACT

Hepatitis C virus (HCV) infects almost 3% of people worldwide and it is considered the main cause of liver chronic diseases and transplants. Until today, there is no effective vaccine and the current most used therapy, based on Peginterferon, is successful only in 50% of patients infected by genotype 1. Although the outcomes of this treatment resistance are unclear, it is suggested host and virus factors may participate in this mechanism. Non-structural 5A (NS5A) protein is involved in several cellular and virus processes and it is a critical component of HCV. However, its structure and function are still uncertain. Regarding these facts, the present study attachments were to elaborate a model of the NS5A protein and to investigate NS5A structural and functional features, using *in silico* tools. It was analyzed 345 sequences of HCV NS5A protein from 23 patients infected by genotypes 1 or 3. Residues and secondary structure composition of all sequences demonstrated that there are differences between genotypes. It may indicate that there are differences in interactions between genotypes, which could be related with the distinct average of treatment resistance. In addition, among those that varied between genotypes, there were amino acids in regions that studies suggested as related with virus persistence. Functional analysis was performed with *ProtFun*. It suggested that NS5A is involved with central intermediary metabolism, translation, growth, transport, ligation and hormone functions in the cell. These functions vary between the domains, strengthening the hypothesis that NS5A is a multifunctional protein. *Prosites* motif search indicated that there are many glycosylation, phosphorylation and myristoylation sites, which are highly conserved and may play an important role in structural stabilization and function. Some of them are in regions that were related with therapy response. In addition, it was identified a highly conserved site of integrin ligation, a site related with nuclear and perinuclear localization of proteins and can be linked to described nuclear forms of NS5A. By threading, homology molecular modeling and molecular dynamics methods, it was elaborated the 3D structure of complete NS5A, which will be useful to other protein studies. Evaluation of alterations in NS5A structure/function in mutants showed many structure and/or biochemical alterations involved in function disruption of the protein.

Keywords: Hepatitis C virus, HCV, NS5A, Structural Bioinformatics

LISTA DE FIGURAS

Figura 1. Estrutura em fita da estrutura do NS5A-D1, sem a hélice anfipática. (Adaptado de Tellinghuisen et al., 2005)	04
Figura 2. Esquema das principais regiões conhecidas da proteína NS5A. (HOFMANN, ZEUZEM, SARRAZIN, 2005).....	07
Figura 3. Média da quantidade (%) de cada aminoácido na NS5A completa por genótipo.....	24
Figura 4. Média da quantidade (%) e cada forma de estrutura secundária resumida por tipos de resposta ao tratamento e/ou genótipo.....	24
Figura 5. Diagrama de Ramachandran mostrando a avaliação geométrica do modelo gerado para a sequência AF009606 (1a).....	27
Figura 6. Modelo estrutural da NS5A gerado por sequência-referência AF009606.....	30
Figura 7. Gráfico mostrando a alteração de RMSD durante a dinâmica molecular. Cada linha representa uma sequência diferente da NS5A.....	32
Figura 8. Modelos da proteína NS5A antes (A) e após (B) a análise por dinâmica molecular da sequência referência AF009606 (genótipo 1a).....	32
Figura 9. Modelos da proteína NS5A antes (A) e após (B) a análise por dinâmica molecular da sequência referência D50481 (genótipo 1b).....	32
Figura 10. Modelos da proteína NS5A antes (A) e após (B) a análise por dinâmica molecular da sequência referência D17763 (genótipo 3).....	33
Figura 11. Resultado da predição de região transmembrânica do programa MEMSAT3.....	34
Figura 12. Representação em cartoon da NS5A completa demonstrando os sítios de N-glicosilação	37
Figura 13. Representação em cartoon da NS5A completa demonstrando o sítio de fosforilação dependente de cAMP e c GMP.....	38
Figura 14. Representação em cartoon da NS5A completa demonstrando os Sítios de fosforilação por PKC	38
Figura 15. Representação em cartoon da NS5A completa demonstrando os sítios de fosforilação por CKII	39
Figura 16. Representação em cartoon da NS5A completa demonstrando o sítio de fosforilação por Tirosina Quinase.....	39
Figura 17. Representação em cartoon da NS5A completa, demonstrando os sítios de n-miristoilação.....	40
Figura 18. Representação em cartoon da NS5A completa, demonstrando o sítio de ligação celular	40

Figura 19. Gráfico demonstrando a porcentagem (eixo y) de sequências da NS5A completa em cada uma das categorias funcionais (eixo x) classificadas pelo ProtFun.	42
Figura 20. Gráfico demonstrando a porcentagem (eixo y) de sequências da NS5A-DI em cada uma das categorias de ontologias do gene (eixo x) classificadas pelo ProtFun.	42
Figura 21. Gráfico demonstrando a porcentagem (eixo y) de sequências da NS5A-DII em cada uma das categorias de ontologias do gene (eixo x) classificadas pelo ProtFun.	43
Figura 22. Gráfico demonstrando a porcentagem (eixo y) de sequências da NS5A-DII na classificação de enzima ou não- enzima (eixo x) do ProtFun.	43
Figura 23. Gráfico demonstrando a porcentagem (eixo y) de sequências da NS5A-DII nas categorias de ontologia do gene (eixo x) classificadas no ProtFun.	44
Figura 24. Gráfico demonstrando a porcentagem (eixo y) de sequências da NS5A-DIII nas categorias de ontologia do gene (eixo x) classificadas no ProtFun.....	44
Figura 25. Gráfico demonstrando a porcentagem (eixo y) de sequências da NS5A-DIII nas categorias de ontologia do gene (eixo x) classificadas no ProtFun.	45
Figura 26. Esquemas da alteração na posição 02.....	49
Figura 27. Esquemas da alteração na posição 06.	50
Figura 28. Esquemas da alteração na posição 09.....	51
Figura 29. Esquemas da alteração na posição 09.....	52
Figura 30. Esquemas da alteração na posição 10.	53
Figura 31. Esquemas da alteração na posição 11.....	54
Figura 32. Esquemas da da deleção da CRS, em comparação da mutação em azul com a forma nativa da referência D17763 em rosa.....	55
Figura 33. Esquemas da alteração na posição 326-328	56
Figura 34. Esquemas da alteração na posição 335-337.	57
Figura 35. Esquemas da alteração na posição 342.	58

LISTA DE TABELAS

- Tabela 1.** Moldes utilizados para montagem do modelo tridimensional do genótipo 1.25
- Tabela 2.** Moldes utilizados para montagem do modelo tridimensional do genótipo 3.25
- Tabela 3.** Valores do *G-factor* para os modelos gerados das sequências referências do Genbank da proteína NS5A do HCV genótipos 1 e 3.....28
- Tabela 4.** Valores do desvio médio-quadrático (RMSD) para os modelos gerados das sequências referências do Genbank da proteína NS5A do HCV genótipos 1 e 3.....29
- Tabela 5.** Ficha técnica das condições gerais das simulações realizadas por dinâmica molecular.....30
- Tabela 6.** Resumo das alterações encontradas da NS5A descritas como comprometedoras da função na literatura.....44

LISTA DE SIGLAS E ABREVIACÕES

Aminoácidos

Código de 3 letras	Código de 1 letra	Nome do aminoácido
Ala	A	Alanina
Arg	R	Arginina
Asn	N	Asparagina
Asp	D	Ácido Aspártico
Cys	C	Cisteína
Gln	Q	Glutamina
Glu	E	Ácido Glutâmico
Gly	G	Glicina
His	H	Histidina
Iso	I	Isoleucina
Leu	L	Leucina
Lys	K	Lisina
Met	M	Metionina
Phe	F	Fenilalanina
Pro	P	Prolina
Ser	S	Serina
Thr	T	Treonina
Trp	W	Triptofano
Tyr	Y	Tirosina
Val	V	Valina

3D – tridimensional

AKT – família protéica de proteínas quinases, também chamadas de proteínas quinases B

C-terminal – carboxiterminal

c-RAF – tirosina quinase de proteína receptora de tirosina

cAMP – adenosina monofosfato cíclico

cGMP – guanosina monofosfato cíclico

CKI – caseína quinase I

CKII – caseína quinase II

CRS – sinal de retenção citoplasmático

DM – dinâmica molecular

DNA – ácido desoxirribonucleico

E – estrutural

EC – classe enzimática

GO – ontologia do gene

GRB2 – proteína adaptadora do receptor do fator de crescimento

HEK – rim de embrião humano

HCV – vírus da hepatite C

HIV – vírus da imunodeficiência adquirida

IFN - Interferon

IRES – sítio interno de entrada do ribossomo

ISDR – região determinante de sensibilidade ao Interferon

L1-CAM – molécula de adesão celular do tipo L1

MEK – proteína quinase ativada por mitógeno

MKK1 – proteína quinase de quinase ativada por mitógeno 1

MM – modelagem molecular

NLS- sinal de localização nuclear

NR – não-respondedor

NS – não-estrutural

N-glicosilação – aminoglicosilação

N-miristoilação - aminomiristoilação

N-terminal – aminoterminal

p70s6K – uma serina/treonina quinase

PDB – Protein Data Bank

PKC – proteína quinase C

PKR – proteína quinase ativada por RNA

R – respondedor

RE – retículo endoplasmático

RMN – ressonância magnética nuclear

RMSD – raiz do desvio médio quadrático

RFT – respondedor ao final do tratamento

RNA – ácido ribonucleico

SH3 - um domínio estrutural da proteína que pode estar envolvido na formação de interações obrigatórias da proteína-proteína

SRC – família de quinases homólogas à proteína oncogênica do vírus de sarcoma de Rous

Syk – proteína-tirosina quinase associada à superfície de imunoglobulinas

TRAF2 – fator associado ao receptor do fator de necrose tumoral

U/UC – uracila/uracila-citosina

UTR – região não-traduzida

V3 – variável 3

VMD – Visual Molecular Dynamics

LISTA DE SÍMBOLOS

5'

3'

5'-cap

kDa – quilo Dalton

α – alfa

β – beta

ps – picossegundos

K – Kelvin

fs – femtosegundo

ns – nanossegundo

ϕ - phi

ψ - psi

C α -N – carbono alfa-nitrogênio

C α -C – carbono alfa-carbono

Å – Ângstrom

° - graus

atm – atmosfera padrão

Sumário

1. Introdução.....	1
1.1. Histórico e Patologia	1
1.2. Epidemiologia e Transmissão.....	1
1.3. O vírus da Hepatite C	2
1.4. A proteína NS5A	3
1.5. Variabilidade viral.....	7
1.6. Tratamento e Tipos de Respostas ao Tratamento	8
1.7. Bioinformática aplicada ao estudo estrutural e funcional de proteínas	9
2. Objetivos	12
2.1. Objetivos Gerais	12
2.2. Objetivos Específicos	12
3. Materiais e métodos.....	13
3.1. Fonte de dados.....	13
3.2. Análise da composição de aminoácidos e estrutura secundária da NS5A completa	14
3.3. Modelagem Molecular Comparativa	14
3.3.1. Modeller	15
3.3.2. Ferramentas de análise dos modelos	16
3.4. Threading aplicado ao domínio II e domínio III	17
3.5. Dinâmica Molecular	18
3.5.1. Pacote de simulação	18
3.5.2. Minimização de energia.....	19
3.5.3. Procedimentos de simulação	19
3.5.4. Análise dos dados de dinâmica molecular.....	20
3.6. Predição de Regiões Transmembrânicas	20
3.7. Predição de Sítios na NS5A.....	20
3.8. Predição de Função da NS5A.....	21
3.9. Análise de mutantes da NS5A.....	22
4. Resultados	24
4.1. Análise da composição de aminoácidos e estrutura secundária da NS5A completa	24
4.2. Modelagem Molecular Comparativa	25
4.3. Dinâmica Molecular	31
4.4. Predição de regiões transmembrânicas	34
4.5. Predição de Sítios na NS5A.....	35
4.6. Predição de Função da NS5A.....	41
4.7. Análise de mutantes da NS5A.....	45
5. Discussão.....	59

6. Conclusão	70
7. Perspectivas	71
8. Referências Bibliográficas.....	72
9. Anexo	83

1.1. Introdução

1.2. Histórico e Patologia

A Hepatite C é considerada a principal hepatite associada às transfusões sanguíneas e adquirida na comunidade. A doença pode evoluir para distúrbios malignos crônicos, incluindo cirrose hepática e carcinoma hepatocelular (ALTER et al., 1989; SAITO et al., 1990). A alta probabilidade de persistência da doença é característica da Hepatite C, uma vez que isto não ocorre em infecções por outros vírus do mesmo grupo. Essa persistência vem sendo relacionada às estratégias virais ativas e passivas de evasão do sistema imune (GALE, FOY, 2005). Esta propriedade e a alta prevalência explicam a grande relevância desta doença para a medicina (APPEL et al., 2008).

1.2. Epidemiologia e Transmissão

O agente etiológico da Hepatite C, chamado de vírus da Hepatite C (HCV) foi isolado de chimpanzés infectados com sangue de pacientes que apresentaram a hepatite não-A e não-B após transfusão (CHOO et al., 1989).

Estima-se que aproximadamente 300 milhões de pessoas estejam infectadas pelo HCV (BROWN, GAGLIO, 2003; CHAMBERS et al., 2005), e a cada ano aproximadamente 3 - 4 milhões de novos casos surgem (SY, JAMAL, 2006; HUANG, HU, 2006). No Brasil, a porcentagem de anticorpos anti-HCV detectados apresenta diferentes taxas de acordo com a região: 0,9 a 2,4% no Norte, 1,7 a 3,4% no Nordeste, 1,0 a 1,4% no Centro-Oeste, 0,8 a 2,8% no Sudeste e 1,1 a 2,1% no Sul (CAMPIOTTO et al., 2005).

A transmissão do HCV é principalmente via parenteral. Com a introdução de métodos para a detecção de anti-HCV em amostras de sangue a partir de 1990, o número novos casos de Hepatite C por transfusão de sangue decresceu em vários países (SOLDAN et al., 2003).

Entretanto, o uso de drogas intravenosas tornou-se a principal via identificada de transmissão (ESTEBAN, SAULEDA, QUER, 2008).

Outras vias de transmissão também são estudadas, como a infecção ocupacional por meio de agulhas contaminadas que resulta em soroconversão em aproximadamente 3% dos casos, representando um risco de transmissão intermediária entre o apresentado pelo HIV e pelo vírus da hepatite B (WASLEY, ALTER, 2000; ESTEBAN, SAULEDA, QUER, 2008). Em um estudo com 481 pacientes, 68% dos usuários de cocaína intranasal eram portadores do vírus, o que torna o uso de drogas não-injetáveis uma possível rota de transmissão (CONRY-CANTILENA et al., 1996).

A infecção do vírus por via sexual não é comprovada e estudos nesta área apresentam resultados contraditórios. As vias de transmissão vertical (de mãe para filho) e a intrafamiliar também são estudadas. No entanto, em mais de 40% dos casos de infecção pelo HCV não são reconhecidos fatores de risco epidemiológico (MORADPOUR et al., 2001).

1.3. O vírus da Hepatite C

O HCV é classificado na família *Flaviviridae*, gênero *Hepacivirus*, embora seja diferente em vários aspectos da organização do genoma em relação aos demais membros dessa família (SIMMONDS, 2004), principalmente quanto à organização das proteínas estruturais e à tradução independente de 5'-cap (LIN et al., 1997).

O RNA genômico deste vírus é uma fita simples de polaridade positiva, com aproximadamente 9.600 pares de bases. Assim como os flavivírus e os pestivírus, membros da mesma família, o genoma viral é composto por uma região não-traduzida 5' (5'UTR), uma região aberta de leitura que codifica uma poliproteína precursora de aproximadamente 3.000 aminoácidos e uma região não-traduzida 3' (3' UTR) (LIANG et al., 2007).

A 5' UTR é altamente conservada e possui um sítio de entrada de ribossomo (IRES), essencial para a tradução de RNA viral. Esta é seguida das regiões codificadoras das seguintes proteínas estruturais: “core” (Proteína C), Envelope 1 (E1) e Envelope 2 (E2); uma proteína

associada à membrana (p7) e proteínas não-estruturais 2 (NS2), 3 (NS3), 4A (NS4A) e 4B (NS4B), 5A (NS5A) e 5B (NS5B). Seguido da NS5B, a 3' UTR é composta por uma pequena região variável, uma cauda poli (U/UC) e uma região altamente conservada. A poliproteína traduzida é posteriormente clivada por proteases virais e celulares, produzindo as proteínas estruturais e não-estruturais (MORADPOUR et al., 2001; MORADPOUR; BLUM, 2004; LYRA et al., 2004; LINDENBACH et al., 2005).

O ciclo de vida proposto para o HCV inclui (1) ligação na superfície celular por meio de um receptor de membrana ainda não identificado e internalização na célula hospedeira, (2) liberação citoplasmática e desempacotamento do RNA viral, (3) tradução mediada pela IRES (4) processamento da poliproteína por proteases celulares e virais, (5) replicação de RNA, (6) montagem e empacotamento da partícula viral, (7) maturação viral e (8) liberação da célula hospedeira (MORADPOUR et al., 2001).

1.4. A proteína NS5A

A proteína não-estrutural 5A (NS5A) é uma fosfoproteína crítica para o HCV de aproximadamente 447 aminoácidos e pode estar relacionada principalmente a replicação de RNA viral, modulação das vias de sinalização celular, resposta ao Interferon, patogenicidade e regulação da apoptose. Entretanto, a função enzimática e estrutura precisas desta proteína são desconhecidas, estudos apontam para uma função através da ligação com outras proteínas do HCV e com proteínas da célula hospedeira. (REED et al., 1997; MACDONALD, HARRIS, 2004; LIANG et al., 2007; LOVE et al., 2009).

A NS5A do HCV é constituída de três domínios: o domínio I (resíduos 1 a 213), o domínio II (resíduos 250 a 342) e o domínio III (resíduos 356 a 447) (TELLINGHUISEN et al., 2004, 2005). No domínio I (NS5A-D1), os primeiros 30 aminoácidos no N-terminal da NS5A são correspondentes a um domínio de ligação à membrana (código de acesso ao PDB: 1R7C), que apresenta uma hélice alfa anfipática na membrana modelo e pode ser responsável pela localização

da proteína no retículo endoplasmático. (LIU et al., 2006, BRASS et al., 2002). Recentemente, foi observado que monômeros do domínio I podem interagir entre si, formando dímeros tanto por meio da cristalização como *in vivo*. Esse estudo também indica que a dimerização da NS5A pode apresentar papel importante na ligação do RNA durante a replicação viral e/ou na maturação e formação das partículas virais (LOVE et al., 2009).

Tellinghuisen e colaboradores (2005), por meio da técnica de difração por raios X, demonstraram que a estrutura do NS5A-DI é altamente conservada. Baseando-se na estrutura tridimensional, o mesmo estudo dividiu o domínio I em dois subdomínios: IA e IB (Figura 1) (TELLINGHUISEN et al., 2005).

O subdomínio IA (Figura 1) consiste em um aminoterminal com uma hélice alfa anfipática de ligação à membrana, uma extensa conformação em “fita” adjacente a uma folha beta anti-paralela com três fitas e uma hélice alfa. Esses elementos constituem a estrutura do sítio de ligação ao zinco (código de acesso ao PDB: 1ZH1), composto por quatro resíduos de cisteína (Cys) altamente conservadas Cys 39, Cys 57, Cys 59 e Cys 80. A propriedade de ligação ao zinco é essencial para a replicação do HCV (GALE et al., 1998). O subdomínio IB (Figura 1) consiste de duas folhas beta anti-paralelas circundadas por uma extensiva espiral randômica (TELLINGHUISEN et al., 2005).

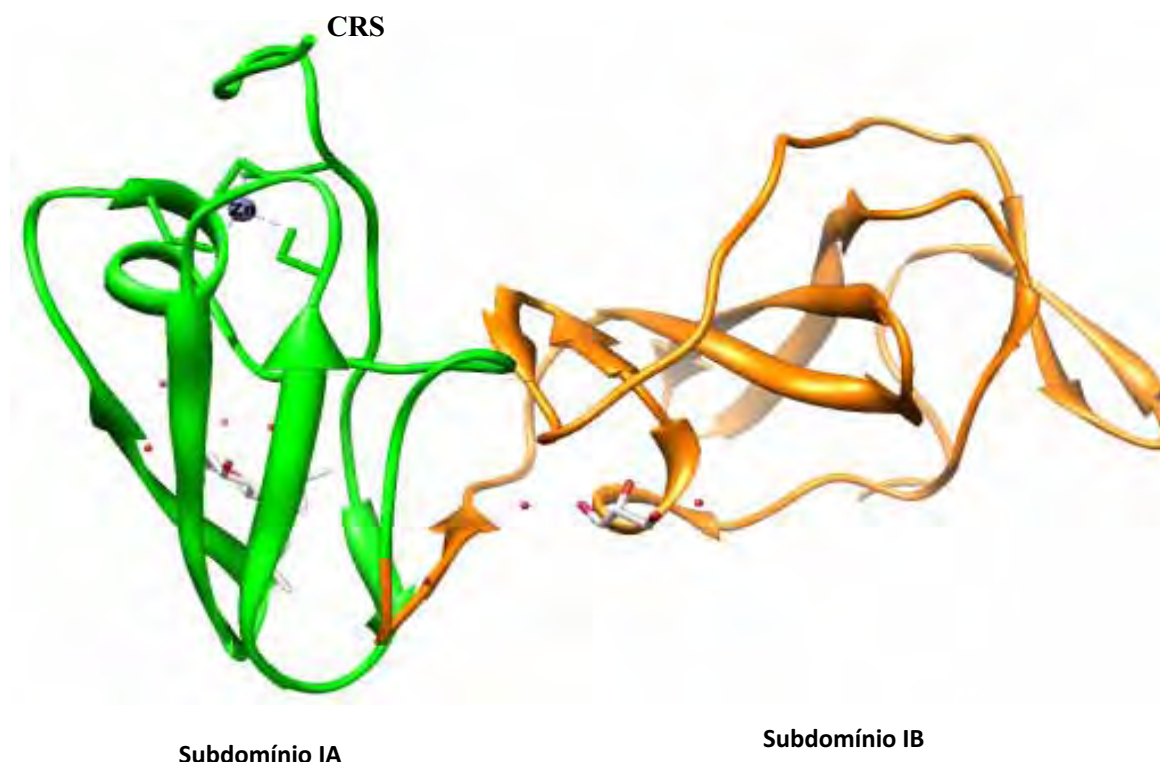


Figura 1. Estrutura do NS5A-DI. A região em verde representa o aminoterminal e em laranja, o carboxiterminal. A esfera cinza representa o átomo de zinco e CRS indica a região onde está a hélice anfipática. (Adaptado de Tellinghuisen et al., 2005).

O domínio II da NS5A (NS5A-D2) do HCV tem sido relacionado a várias vias de regulação biológica. Esse domínio interage com a NS5B (SHIROTA et al., 2002), além de compreender a região determinante de sensibilidade ao *Interferon* (ISDR), o domínio de ligação à proteína quinase ativada por RNA (PKR) (LIANG et al., 2007) e uma região rica em prolina, semelhante ao motivo de ligação ao Homólogo SRC 3 (SH3). Esse motivo é encontrado em diversos grupos de proteínas virais e celulares de sinalização da transdução (KRAJCSI, WOLD, 1998).

Os domínios II e III se caracterizam por apresentarem enovelamento não-definido. Proteínas sem enovelamento demonstram alta flexibilidade, característica que pode permitir a adaptação da estrutura local ou global ao ambiente celular. Essa flexibilidade intrínseca da estrutura permite o reconhecimento de um grande número de ligantes sem comprometer a especificidade protéica (WRIGHT, DYSON, 1999). O domínio II é apontado como uma região

multifuncional, que pode interagir com várias proteínas incluindo a NS5B, a PKR, a TRAF2 (fator associado ao receptor do fator de necrose tumoral) e GRB2 (proteína adaptadora do receptor do fator de crescimento), participar na ativação de transcrição, suprimir a IRES e apresentar função anti-apoptótica (LIANG et al., 2007).

O domínio III indica não ser essencial para o processo de replicação do RNA viral, mas foi demonstrado que participa da montagem do vírion. Além disso, foram identificados diversos sítios de fosforilação neste domínio (HUGHES et al., 2009; APPEL et al., 2008).

Uma evidência da relação entre a sequência viral e a resposta ao tratamento foi demonstrada pela interação da proteína NS5A com a PKR. Esta proteína tem sido relacionada à ativação de apoptose celular e a resposta imune mediada por Interferon, a interação com a NS5A parece inibir esta função. (SIMMONDS, 2004).

Enomoto e colaboradores (1996) encontraram uma associação entre a resposta ao tratamento e uma região no domínio II da NS5A denominada de região determinante da sensibilidade ao Interferon (ISDR). Como a ISDR com mais 26 aminoácidos C-terminal forma a *PKR-binding*, foi sugerido que o mecanismo de ligação à PKR seria um determinante crucial na persistência viral e potencial na determinação de outros aspectos da interação vírus-hospedeiro (SIMMONDS, 2004). Além disso, mutações no domínio V3, localizado no domínio III e na ISDR também vêm sendo relacionadas à resposta ao tratamento (Figura 2) (PUIG-BASAGOITI, 2005).

A NS5A é comumente detectada no citoplasma celular, próxima da membrana nuclear ou do retículo endoplasmático e aparelho de Golgi. Um evento de processamento pós-traducional da NS5A por caspases celulares pode ativar o sinal de localização nuclear (NLS) (domínio III), resultando na forma da NS5A de 31kDa. A forma de 31 kDa foi localizada no núcleo, sugerindo que a NS5A pode agir como ativador transcricional (TANJI et al., 1995). Outros estudos indicam que a presença da região N-terminal de 27 aminoácidos (domínio I), denominada sinal de retenção citoplasmática (CRS) é capaz de manter a NS5A no citoplasma, mostrando uma atividade dominante sobre a NLS (Figura 2) (IDE et al., 1996). Mais tarde, foi demonstrada a CRS era

responsável pela associação da NS5A à membrana do retículo endoplasmático, essencial para a formação do complexo replicativo e consequentemente para a replicação viral (BRASS et al., 2006).

As formas nucleares da NS5A ainda possuem função obscura, mas estudos indicam que estas possam exercer papel importante na regulação da transcrição. Outro importante dado é que esses mutantes podem carregar para o núcleo proteínas associadas, como a quinase c-RAF (tirosina quinase de proteína receptora de tirosina), responsável pela sinalização de diversas vias celulares (BÜRCKSTÜMMER et al., 2006; SAUTER et al., 2009).

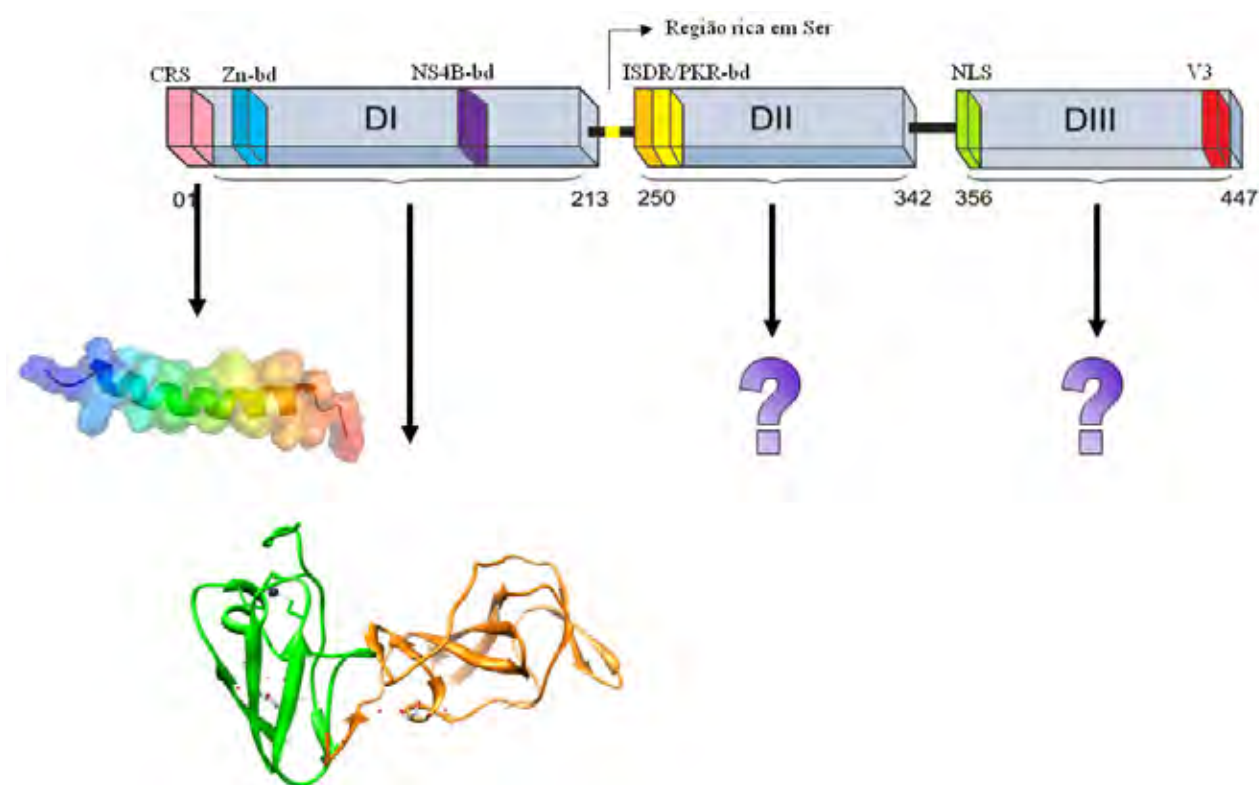


Figura 2. Esquema das estruturas e principais regiões conhecidas da proteína NS5A. CRS: sítio de retenção citoplasmático, Zn-bd: domínio de ligação ao zinco, NS4B-bd: sítio de ligação à NS4B, ISDR: região de sensibilidade ao Interferon, PKR-bd, sítio de ligação à PKR, NLS: sinal de localização nuclear, V3: região variável V3 (Baseado em HOFMANN. ZEUZEM. SARRAZIN. 2005)

1.5. Variabilidade viral

Estudos sugerem que a infecção pelo HCV é um processo dinâmico, ou seja, apresenta alta taxa de produção e destruição de vírions. Neumann e colaboradores (1998) mostraram que a meia-vida viral é de apenas 2,7 horas e a média de produção e liberação do vírion ocorre em taxas de aproximadamente $1,3 \cdot 10^{12}$ /paciente em um dia. Desse modo, juntamente com a ausência de atividade de correção da polimerase viral, a alta taxa replicação determina a base para a variabilidade genética do HCV (MORADPOUR et al., 2001).

Além disso, o HCV coevoluiu com as populações humanas por séculos, ou até por milênios o que resultou numa ampla diversidade genética (PYBUS et al., 2001). Desse modo, atualmente o HCV é classificado em três categorias: genótipos, subtipos e *quasispecies* (SIMMONDS, 1994). Até o momento, reconhecem-se sete grandes genótipos do vírus (1 a 7), que podem ser divididos em mais de 50 subtipos (denominados por letras minúsculas) (KUIKEN, SIMMONDS, 2009).

O termo *quasispecies* refere-se a uma população de genomas relacionados de HCV, coexistindo em um mesmo indivíduo infectado (MORADPOUR et al., 2001). Uma das hipóteses para a persistência viral inclui a alta heterogeneidade do vírus, que permite a emergência de variantes resistentes ao tratamento. Vários estudos focaram a composição de *quasispecies* como fator preditivo da resistência viral ao tratamento. Entretanto, esses dados ainda são controversos, pois diferentes resultados podem ser obtidos dependendo do tipo de estudo realizado, das técnicas utilizadas e da região analisada (OKADA et al., 1992; GONZALEZ-PERALTA et al., 1996; GRAHOVAC et al., 2001; THELU et al., 2001; TANG et al., 2002; SALMERON et al., 2006).

1.6. Tratamento e Tipos de Respostas ao Tratamento

A resposta ao tratamento antiviral do HCV pode ser classificado em três grupos: resposta virológica sustentada ou respondedor (R), resposta ao final do tratamento e recorrência da infecção (RFT), e não-resposta (NR). A R é definida pela ausência de detecção do RNA de HCV seis meses após o término do tratamento. Os pacientes com resposta transitória (RFT) apresentam

ausência do RNA viral durante o tratamento e detecção após final da terapia. Os pacientes NR apresentam RNA de HCV detectado durante e após o término da terapia (FELD, HOOFNAGLE, 2005).

A monoterapia com Interferon alfa forneceu o primeiro tratamento do vírus para os pacientes com hepatite C crônica. Essa substância não inibe uma função enzimática específica do vírus, mas induz modificações nas respostas do sistema imune e estabelece um estado antiviral inespecífico nas células infectadas pela ativação de diferentes genes celulares. Dessa forma, a inibição da replicação do HCV é uma consequência, em parte, da inibição generalizada de tradução na célula infectada (MORENO et al., 2006).

Recentemente, formas peguiladas de Interferon vêm sendo desenvolvidas e a terapia mais utilizada trata-se de uma combinação de Peginterferon e Ribavirina, um análogo de nucleosídeo. Com essa combinação, o tratamento tem sido mais eficaz. Por essa razão, Peginterferon alfa-2a e Peginterferon alfa-2b são as últimas inovações para o tratamento da hepatite C (FRIED, HADZIYANNIS et al., 2004).

No entanto, as terapias atualmente disponíveis para o tratamento do HCV são efetivas em apenas uma fração dos pacientes, apresentando sucesso em aproximadamente 40% dos indivíduos infectados com HCV genótipo 1 e 80% para os genótipos 2 e 3 (KLENERMAN et al., 2009; HUANG, HU, 2006). A baixa susceptibilidade a tratamentos antivirais e a alta taxa de persistência das infecções provavelmente estão relacionados a uma complexa relação entre fatores virais e do hospedeiro (TORRES-PUENTE et al., 2008). Os fatores virais incluem o genótipo do HCV, nível de replicação (carga viral) e complexidade genética da população de *quasispecies* antes do início do tratamento (MORENO et al., 2006).

1.7. Bioinformática aplicada ao estudo estrutural e funcional de proteínas

A bioinformática tem várias definições. De forma geral, mas não ampla, a bioinformática é uma área interdisciplinar envolvendo biologia, informática, matemática e estatística com o objetivo de analisar dados de seqüências biológicas e prever a função e estrutura de macromoléculas biológicas (BAYAT, 2002).

Nos últimos anos, o rápido progresso no campo da biologia molecular, juntamente com o avanço na tecnologia genômica, resultaram no crescimento explosivo de informações biológicas distribuídas em uma variedade de bases de dados biológicos. Dessa forma, atualmente é preciso saber como analisar em conjunto a grande quantidade de dados produzir novos conhecimentos biológicos (YE, 2008).

Assim, um dos grandes desafios da bioinformática começa a aparecer na era pós-genômica, na qual a proteômica se torna um dos principais alvos de estudo juntamente com o entendimento estrutural e funcional de proteínas. Por exemplo, ferramentas de bioinformática são utilizadas na identificação e caracterização funcional e estrutural de uma proteína por meio de buscas em bancos de dados biológicos públicos e no manuseio do vasto e complexo conjunto de dados de cristalografia de raios X e RMN para prever modelos 3D de moléculas de proteínas por modelagem molecular comparativa, quando utilizadas pelos biólogos estruturais (BURLEY et al., 1999).

A bioinformática pode ser útil em diversos campos da biologia molecular por utilizar métodos computacionais robustos (processamento paralelo, redes neurais, etc.) e por aplicar técnicas analíticas quantitativas a modelagem de sistemas e problemas biológicos (GIBAS, JAMBECK, 2001).

O foco inicial da bioinformática fora o desenvolvimento de ferramentas para análise de seqüências de nucleotídeos e aminoácidos, com a finalidade de identificar relações evolucionárias. A partir do final da década de 80, gradualmente foi-se percebendo que a estrutura tridimensional é mais conservada do que a seqüência, e desta forma, informações estruturais podem revelar relações filogenéticas que estariam ocultas ao nível seqüencial. Os dados estruturais auxiliam na

classificação de domínios proteicos em famílias e em como entender melhor a evolução de novas sequências, estruturas e funções. Além destas linhas, atualmente a bioinformática também permite a predição de complexos proteína-proteína e proteína-ligante, por meio de ferramentas essenciais para o desenvolvimento de novos medicamentos baseados em estrutura (ORENGO et al., 2003).

2. Objetivos

2.1. Objetivos Gerais

A partir de ferramentas de bioinformática, este estudo tem como objetivo analisar, comparar e caracterizar as variações nas estruturas e investigar a função da proteína NS5A em *quasispecies* do HCV genótipo 1 e 3, em amostras pré-tratamento de pacientes submetidos posteriormente à terapia baseada em Interferon e Ribavirina.

2.2. Objetivos Específicos

1) Aplicar a metodologia de modelagem molecular comparativa e dinâmica molecular na obtenção de estruturas-modelo da proteína NS5A do genótipo 1 e 3 em amostras pré-tratamento;

2) Analisar e comparar as estruturas-modelo quanto às possíveis alterações nas características físico-químicas entre as amostras de diferentes tipos de resposta e entre os genótipos 1 e 3;

3) Investigar as possíveis funções da proteína NS5A comparando as seqüências com os bancos de dados de padrões e *motivos* como o PROSITE e pelo programa *ProtFun*;

5) Avaliar se há correlação entre as alterações encontradas e os diferentes tipos de resposta ao tratamento.

3. Materiais e métodos

3.1. Fonte de dados

O seqüenciamento da região NS5A completa de amostras do genótipo 1 do vírus foi realizado no projeto de mestrado “*Análise comparativa da variação entre quasiespecies do Vírus da Hepatite C genótipo 1 em amostras pré-tratamento de pacientes tratados com Peginterferon*”. As seqüências do genótipo 3 foram obtidas no projeto de mestrado “*Análise das quasiepecies de HCV genótipo 3 pela região NS5A antes do tratamento*”. Estes estudos foram realizados com pacientes que apresentaram diferentes respostas ao tratamento baseado em Interferon, acompanhados no ambulatório de hepatologia do Hospital de Base da Faculdade de Medicina de São José do Rio Preto (FAMERP). A cópia do parecer do Comitê de Ética encontra-se no Anexo. As seqüências da proteína NS5A foram obtidas a partir da amostra de sangue de 23 pacientes infectados com o HCV. Destes, 11 foram infectados com o genótipo 1 (6 com o genótipo 1a e 5 com o 1b) e 12 com o genótipo 3. Dos pacientes com o genótipo 1 do vírus, 3 responderam ao tratamento (R), 4 não responderam (NR) e 4 apresentaram reincidência viral após o término do tratamento (RFT). Daqueles infectados com o genótipo 3, 5 pacientes foram respondedores, 4 não respondedores e 3 respondedores ao final do tratamento. Para cada paciente foram obtidas 15 seqüências da NS5A completa. Desta forma, há aproximadamente 165 seqüências da proteína NS5A completa do genótipo 1 e 180 do genótipo 3. No total, estão armazenadas no banco aproximadamente 345 seqüências da NS5A completa. Além destas seqüências, foram escolhidas aleatoriamente 9 seqüências referências do *Genbank* (três para cada genótipo) para avaliar se os resultados observados eram característicos da população analisada ou do vírus. Para o genótipo 1a, foram utilizadas as seqüências de códigos AF009606, AF011752 e NC004102. Do genótipo 1b, as seqüências com os códigos EU256081, D50481 e AF356877 foram utilizadas. Do genótipo 3 foram utilizadas as seqüências NC009824, AM493623 e D17763.

3.2. Análise da composição de aminoácidos e estrutura secundária da NS5A

completa

A porcentagem de cada aminoácido na proteína e a porcentagem de cada um dos tipos de estruturas secundárias (hélice α , folha β e *coil*) foram calculadas no programa PROF (ROST & SANDER, 1993), contido no pacote de programas do *Predict Protein Server* (www.predictprotein.org/) (ROST *et al.*, 2004). Os resultados das análises das sequências utilizadas como referência retiradas do *Genbank* e de cada uma das sequências alvo foram inseridas em uma tabela com os valores para cada aminoácido e outra para as porcentagens dos tipos de estrutura secundária em cada sequência. Foram calculadas a média da porcentagem para cada paciente. Estes valores foram analisados juntamente com os valores para as sequências referências do *GenBank*.

3.3. Modelagem Molecular Comparativa

A modelagem molecular comparativa é o principal método computacional utilizado para prever estruturas terciárias de proteínas. Esta é uma técnica de desenvolvimento de modelos 3D que parte de uma sequência primária conhecida de uma proteína cuja estrutura 3D é indeterminada e utiliza como molde uma ou mais sequências de proteínas cujas estruturas 3D já foram resolvidas e, que possuem similaridade com a proteína de interesse. Portanto, a modelagem molecular comparativa consiste em algumas etapas que são (SÁNCHEZ, ŠALI, 1997):

1. **Busca por moldes:** consiste em realizar uma busca por proteínas que possuem suas estruturas tridimensionais resolvidas. Esta busca é realizada em bancos de dados de estruturas terciárias de proteínas, sendo que o principal deles é o *Protein Data Bank* (PDB) (BERMAN, *et al.* 2000) (<http://www.rcsb.org/pdb>). O principal parâmetro usado para realizar a busca dos moldes é a porcentagem de identidade (superior a 30%) entre a sequência primária do molde e sequência da proteína que será modelada. Existem várias ferramentas na Internet que realizam esta busca.

2. **Alinhamento:** para construir o arquivo de alinhamento de entrada que será utilizado pelo programa *Modeller* na construção do modelo, foi utilizado o próprio alinhamento executado por este programa que utiliza o comando ALIGN2D, o qual é baseado também no algoritmo de programação dinâmica (NEEDLEMAN & WUNSCH, 1970).
3. **Construção do modelo tridimensional:** nesta etapa, o software escolhido irá receber como entrada o arquivo contendo as seqüências alinhadas e o arquivo contendo as coordenadas tridimensionais do molde. Dessa forma, o programa irá construir a estrutura tridimensional da proteína alvo baseado nas coordenadas tridimensionais do molde. Com exceção da etapa de análise, todas as outras etapas acima descritas podem ser realizadas pelo programa *Modeller* (SANCHEZ et al., 1997), no presente estudo utilizou-se a versão 9v6 do programa.
4. **Avaliação dos modelos:** A qualidade do modelo predito determina a informação que pode ser extraída dele. Assim, estimar a precisão do modelo 3D da proteína é essencial para interpretá-lo. O modelo pode ser avaliado como um todo bem como em regiões individuais, com base na similaridade entre as seqüências do molde e do alvo, observando resíduos importantes em regiões da proteína como o sítio ativo e sua conservação (SÁNCHEZ & ŠALI, 1997).

3.3.1. Modeller

O ponto inicial para construção do modelo, utilizando o programa *Modeller* é quando o alinhamento é feito entre a seqüência de aminoácidos alvo e a(s) seqüência(s) de aminoácidos do molde (GIBAS, JAMBECK, 2001), um outro arquivo que também é essencial para construir o modelo 3D da seqüência de aminoácidos alvo é o arquivo de coordenadas atômicas 3D do molde. O programa *Modeller* gera, a partir da estrutura do molde, um conjunto de restrições que são aplicadas à seqüência de aminoácidos alvo a ser modelada. Estas restrições limitam, por exemplo, a distância entre dois resíduos no modelo que está sendo construído, sendo esta distância baseada naquela entre os dois resíduos equivalentes na estrutura molde. Esse cálculo de restrições também

é aplicado nas ligações angulares (entre três átomos) e nos ângulos diedrais (entre quatro átomos) (SALI & OVERINGTON, 1994; SALI *et al.*, 1995).

O cálculo das restrições é baseado em análises estatísticas entre estruturas comparativas. Além dessas restrições um campo de forças também controla as propriedades estereoquímicas entre os átomos, aplicando restrições químicas, adicionais na conformação molecular 3D (SALI & BLUNDELL, 1993). Finalmente, o modelo é obtido pela otimização da função objetiva no espaço cartesiano. Vários modelos ligeiramente diferentes podem ser calculados variando a estrutura inicial (SALI & OVERINGTON, 1994; SALI & BLUNDELL, 1993).

3.3.2. Ferramentas de análise dos modelos

Um requerimento básico para um modelo é ter uma boa qualidade estereoquímica. Os programas utilizados para este tipo de análise foram o programa PROCHECK (LASKOWSKI *et al.*, 1993) que pertence ao pacote CCP4 (Collaborative Computational Project N° 4) e o programa WHATCHECK (HOOFT *et al.*, 1996). As características bioquímicas de um modelo que são checadas por estes programas incluem comprimento de ligação, ângulo de ligação, ligação peptídica e planaridade de anéis da cadeia lateral, quiralidade, ângulos de torção da cadeia principal e da cadeia lateral e choques entre pares de átomos não ligados.

Para analisar quanto ao desvio da posição de cada átomo da estrutura modelada em relação aos parâmetros observados por ENGH e HUBER, (1991) utilizamos o desvio médio quadrático (RMSD) da geometria ideal (KEARSLEY, 1989), que é calculado como uma função de cada um dos átomos de uma estrutura protéica (ENGH & HUBER, 1991). O RMSD da geometria ideal utiliza como parâmetros o comprimento e ângulo de ligação ideal, derivados de uma análise detalhada (ENGH & HUBER, 1991) de estruturas de pequenas moléculas em *Cambridge Structural Database* (CSD) (ALLEN *et al.*, 2002). O programa X-PLOR (BRÜNGER, 1992; SCHWIETERS *et al.*, 2003) foi utilizado para executar os cálculos de comprimento e ângulos de ligação.

3.4. *Threading* aplicado ao domínio II e domínio III

As técnicas experimentais, apesar de apresentarem uma precisão superior aos demais métodos, são lentas e caras, pois necessitam de equipamentos, laboratórios e pessoas especializados. Por outro lado, as aplicações teóricas são mais acessíveis, sendo realizadas em qualquer computador. Dentre elas, os métodos no campo empírico mais conhecidos são a modelagem por homologia (MARTÍ-RENOM et al., 2002) e o *threading* (HIGGINS, TAYLOR, 2000). O programa mGenTHREADER reconhece o tipo de enovelamento protéico através de uma busca baseado em métodos de perfil-perfil. Atualmente, os melhores métodos de identificação de dobramento da proteína são aqueles que aplicam este método. Vários estudos demonstraram que o método perfil-perfil apresentam desempenho superior aos métodos baseados em sequência-perfil. Além disso, métodos baseados em perfil-perfil permitem a detecção de maior número de homólogos remotos com maior confiabilidade (OHLSON et al., 2004; MCGUFFIN et al., 2006). O programa mGenTHREADER é disponível no endereço eletrônico: <http://bioinf.cs.ucl.ac.uk/psipred/>. Um exemplo de sucesso na aplicação desta técnica, utilizando-se múltiplos *templates* foi a predição da proteína Ycdx (código do PDB: 1M65), uma proteína de *E. coli*. Por meio da técnica de *threading*, foram encontrados quatro *templates* com componentes estruturais comuns. A predição se mostrou correta para as seis primeiras folhas beta de oito que compõem a proteína (BUJNICKI, 2008). Como não estão disponíveis estruturas dos domínios II e III da NS5A obtidas experimentalmente, as sequências dessas regiões foram submetidas ao programa mGenTHREADER. Os possíveis *templates* identificados foram utilizados na montagem de um “*template* quimérico” contendo partes de diversas proteínas. Este *template* foi utilizado na modelagem comparativa das sequências completas da NS5A.

3.5. Dinâmica Molecular

Existem duas abordagens na modelagem computacional de estruturas protéicas. A primeira é baseada em métodos que aplicam parâmetros extraídos de uma base de dados de estruturas existentes para avaliar e aperfeiçoar estruturas ou prever a estrutura de uma sequência. A segunda abordagem é baseada na simulação de forças físicas e dinâmica molecular. As simulações físico-químicas são utilizadas para testar como modelo protéico dobra-se em sua compacta e funcional, forma nativa de uma não-funcional, não-compacta, forma desnaturada (GIBAS, JAMBECK, 2001).

Na dinâmica molecular, a energia é adicionada ao sistema da proteína pela simulação de um “aquecimento”. A energia adicionada ao sistema promove uma força oposta que movimenta os átomos na molécula para fora das conformações tradicionais. As ações e reações de centenas de átomos em um dado sistema molecular podem ser simuladas utilizando esta abstração (GIBAS, JAMBECK, 2001).

3.5.1. Pacote de simulação

Para a realização das simulações optamos por utilizar o programa NAMD (PHILLIPS et al. 2005) na versão 2.7b 1, que consiste de um conjunto de programas capazes de executar todos os passos necessários para a simulação da dinâmica molecular, desde a preparação dos sistemas até a análise dos resultados obtidos. Esse programa se trata de um dos mais conhecidos programas de simulação, o que lhe confere inúmeras vantagens. Dentre elas destacamos o fato de existir uma grande comunidade de usuários, o que facilita a busca por informações novas, troca de conhecimento e solução de possíveis problemas.

3.5.2. Minimização de energia

Cada sistema a ser simulado passou por um processo de minimização de energia, processo este que tem por finalidade reduzir a energia potencial do sistema pela busca de mínimos de energia, o que favorece a correta termalização do sistema, pois em uma simulação de dinâmica molecular, ocorre constantemente a conversão de energia potencial em cinética, e picos de energia potencial podem levar ao aparecimento de regiões com temperatura acima do esperado, o que pode comprometer o resultado final obtido. Todas as simulações foram equilibradas por 300 ps.

3.5.3. Procedimentos de simulação

Todas as simulações de dinâmica molecular foram realizadas conforme amostra isotérmica e isobárica em temperatura de 310 K. O passo de integração utilizado foi de 2 fs e o tempo real de 5 ns, o campo de força e o arquivo de topologia escolhidos foram o do programa CHARMM22 (BRÜNGER, 1992).

O tratamento eletrostático foi feito com raio de corte (*Cut-off*) tanto para as interações de Coulomb quanto para as de Van der Waals, com valor para estes raios de corte de 1,4 nm, o que é sempre menos que a metade do comprimento do menor lado da caixa de simulação. Esse cuidado com o valor do raio de corte garante que nenhum átomo ou molécula presente na caixa de simulação seja considerado mais de uma vez nos cálculos das interações eletrostáticas.

3.5.4. Análise dos dados de dinâmica molecular

Em simulações de Dinâmica Molecular (DM), mesmo após a etapa de equilíbrio e termalização, quando os átomos do sistema atingem uma distribuição de velocidades compatível com a temperatura, o relaxamento das estruturas ainda continua por tempos que vão além de centenas de picosegundos (ps). Isto ocorre geralmente devido as diferenças entre o ambiente aquoso da simulação de macromoléculas biológicas e a estrutura cristalina inicial. Para avaliar o comportamento global das estruturas foi calculado a Raiz do Desvio Médio Quadrático (RMSD) para os sistemas, no tempo, tomando como referência a estrutura após as minimizações. Estas

análises permitem acompanhar dinamicamente as variáveis da estrutura em relação à estrutura antes da dinâmica molecular.

O programa VMD foi utilizado para visualizar as estruturas geradas ao longo do processo de dinâmica molecular.

3.6. Predição de Regiões Transmembrânicas

A predição de regiões transmembrânicas foi desenvolvida no programa MEMSAT3, disponível no endereço eletrônico <http://bioinf.cs.ucl.ac.uk/psipred/psiform.html>. Este programa utiliza a metodologia de redes neurais para a predição de regiões transmembrânicas em proteínas, demonstrando uma precisão de até 80% e uma baixa taxa de falsos positivos (0,5%) (JONES, 2007). Todas as sequências alvo e referência foram submetidas a este programa.

3.7. Predição de Sítios na NS5A

A predição dos padrões e *motivos* da proteína NS5A foi realizada em todas as sequências alvo e nas sequências referência do *Genbank*, na base de dados do PROSITE (BAIROCH *et al.*, 1997), contido no *Predict Protein Server* (ROST *et al.*, 2004).

O PROSITE é um banco de dados de padrões, os quais não estão desenvolvidos fundamentadas em uma análise automatizada. Em vez disso, são cuidadosamente selecionados com base em dados publicados na literatura básica ou em pesquisas descrevendo a funcionalidade de determinados grupos de proteínas (BAIROCH *et al.*, 1997).

Um padrão de sequência é uma região localmente conservada de uma sequência ou um pequeno padrão de sequências compartilhado por um conjunto de sequências. O termo padrão geralmente se refere a qualquer padrão de sequência que pressuponha uma função de uma proteína, uma característica estrutural ou membro de uma família. O uso mais comum das análises

baseadas em padrões é a detecção de padrões de sequência que correspondem às características estruturais ou funcionais das proteínas (GIBAS, JAMBECK, 2001).

Os padrões são gerados a partir de alinhamentos de várias sequências e podem ser exibidos como padrões de aminoácidos como os do banco de dados PROSITE. Podem ser criados para famílias de proteínas ou conjuntos de proteínas cujos membros são evolutivamente relacionados (GIBAS, JAMBECK, 2001).

Os *motivos* encontrados foram comparados de acordo com os grupos de resposta ao tratamento e de acordo com os genótipos. Para a montagem das ilustrações foi utilizado o programa *Pymol* e o modelo gerado por modelagem molecular comparativa da sequência referência AF009606 (1a).

3.8. Predição de Função da NS5A

Para predição de função nas sequências alvo e referência da proteína NS5A dos genótipos 1 e 3 foi utilizado o programa *ProtFun* (JENSEN *et al.*, 2002). Este programa utiliza outros inúmeros servidores de predição de propriedades protéicas para obter informações de vários aspectos pós-traducionais e de localização intracelular, informações que são integradas para a predição final da sequência alvo em uma das 12 funções celulares segundo o esquema de Riley; uma classe enzimática (EC) se esta for uma enzima ; umas das 14 categorias de ontologia genética (JENSEN *et al.*, 2002). A bio-ontologia é um recurso de bioinformática que captura informações biológicas de forma estrutural. Essa ciência integra dados compartilhados por diversas bases de dados e são utilizadas como base para extrair propriedades decorrentes de grandes *datasets* produzidos por análises de transcriptomas e proteomas. Atualmente, a ferramenta mais utilizada da bio-ontologia é a Ontologia do Gene (GO), que é utilizado para representar as funções dos produtos genéticos e se tornou uma ferramenta chave para a genômica funcional. Embora o uso do GO tenha sido primariamente focado no modelo de espécies eucariotas, ele também pode ser utilizado para descrever e modelar funções genéticas em patógenos e em organismos de vida livre

(NANDURI et al., 2005; SEVERINO et al., 2007; RHEE et al., 2008). Através da GO não é possível descrever todos os tipos de dados biológicos, esta é limitada para três funções de produtos gênicos que são centrais para a fisiologia (assim como para a patofisiologia): funções moleculares individuais, processos biológicos gerais e localização celular. Além disso, a GO não descreve diretamente os processos genéticos, funções ou componentes que não unicamente atribuídos a mutantes ou doenças. Por exemplo, tumorigênese não é uma função válida nos termos da GO, pois a formação de tumor não é uma função normal de células eucariotas (RHEE et al., 2008).

3.9. Análise de mutantes da NS5A

No banco de sequências da NS5A completa utilizado, foram encontradas alterações descritas na literatura como comprometedoras da função da NS5A por Penin e colaboradores (2004) e em Tellinghuisen e colaboradores (2008) em estudos *in vitro*.

As formas mutantes encontradas em nosso banco de dados passaram pela predição de estrutura por modelagem molecular por homologia, *threading* e dinâmica molecular. As estruturas resultantes foram analisadas em comparação com estruturas das sequências referência do *Genbank* do genótipo correspondente.

Para visualização, confecção de figuras de cada mutante (hidrofobicidade e carga) foram utilizados os programas Pymol e Molegro Molecular Viewer.

4. Resultados

4.1. Análise da composição de aminoácidos e estrutura secundária da NS5A completa

Todas as 345 sequências-alvo da NS5A foram analisadas quanto a porcentagem de cada um dos vinte aminoácidos em sua sequência. A partir destes dados, foi gerado um gráfico com a média por genótipo da porcentagem de todos os aminoácidos. A Figura 3 resume esses dados, mostrando apenas os valores dos aminoácidos que variaram de acordo com o genótipo e/ou subtipo. Podemos observar que a composição de alanina (apolar) na NS5A apresenta uma pequena variação entre os subtipos do genótipo 1 e entre o genótipo 1 e 3. A composição de ácido glutâmico (polar ácido), glutamina (polar neutro) e tirosina (apolar) apresentou diferença do genótipo 1a para 1b e 3. Além disso, todas as *quasispecies* do genótipo 1a apresentaram exatamente 2% de tirosina, enquanto que para os genótipos 1b e 3 este valor apresentou pequena variação em algumas sequências. A composição de cisteína (apolar), valina (apolar) e treonina (polar neutro) apresentaram valor diferente entre os genótipos 1 e 3. Não houve variação da composição de aminoácidos quanto ao tipo de resposta ao tratamento do paciente.

Da mesma forma, para a análise da composição de estrutura secundária na NS5A, todas as sequências-alvo foram submetidas ao programa. Os resultados da análise de estrutura secundária são resumidos na Figura 4.

Não foi possível observar relação entre a composição de estrutura secundária e o tipo de resposta. Entretanto, pode-se observar que há uma diferença entre os genótipos. O genótipo 1 tende a apresentar maior quantidade de folhas beta e menor de *coil* do que o genótipo 3.

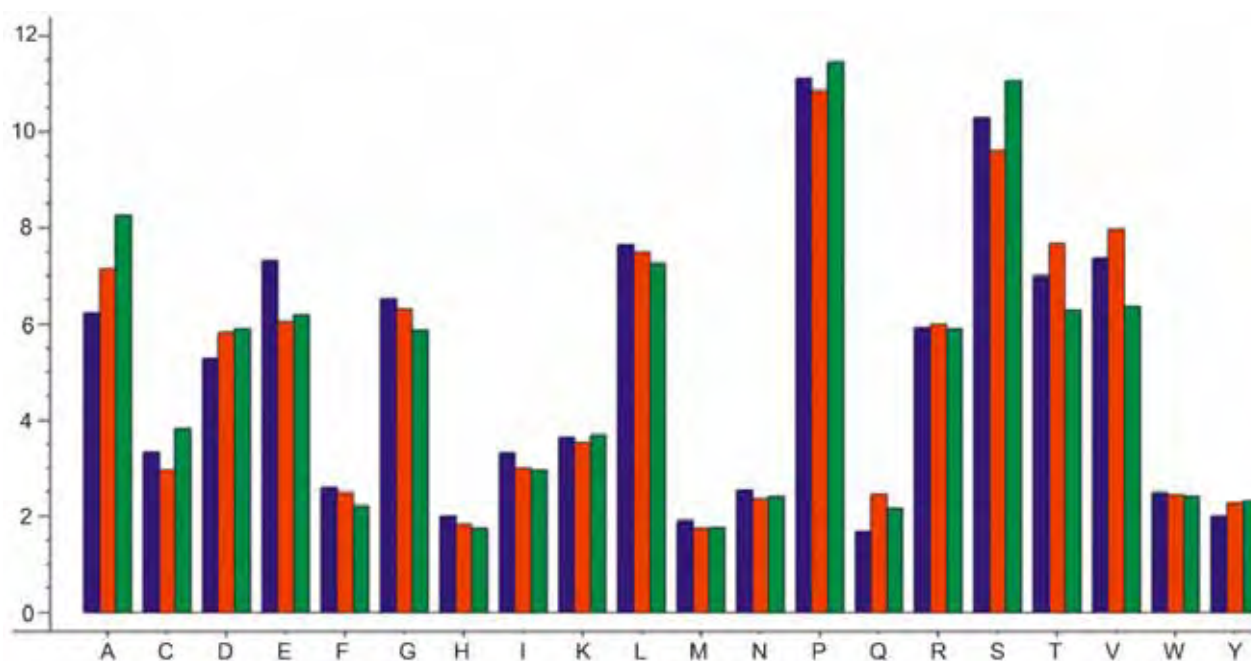


Figura 3. Média da quantidade (%) de cada aminoácido na NS5A completa por genótipo. Azul – 1a, Vermelho – 1b e Verde – 3.

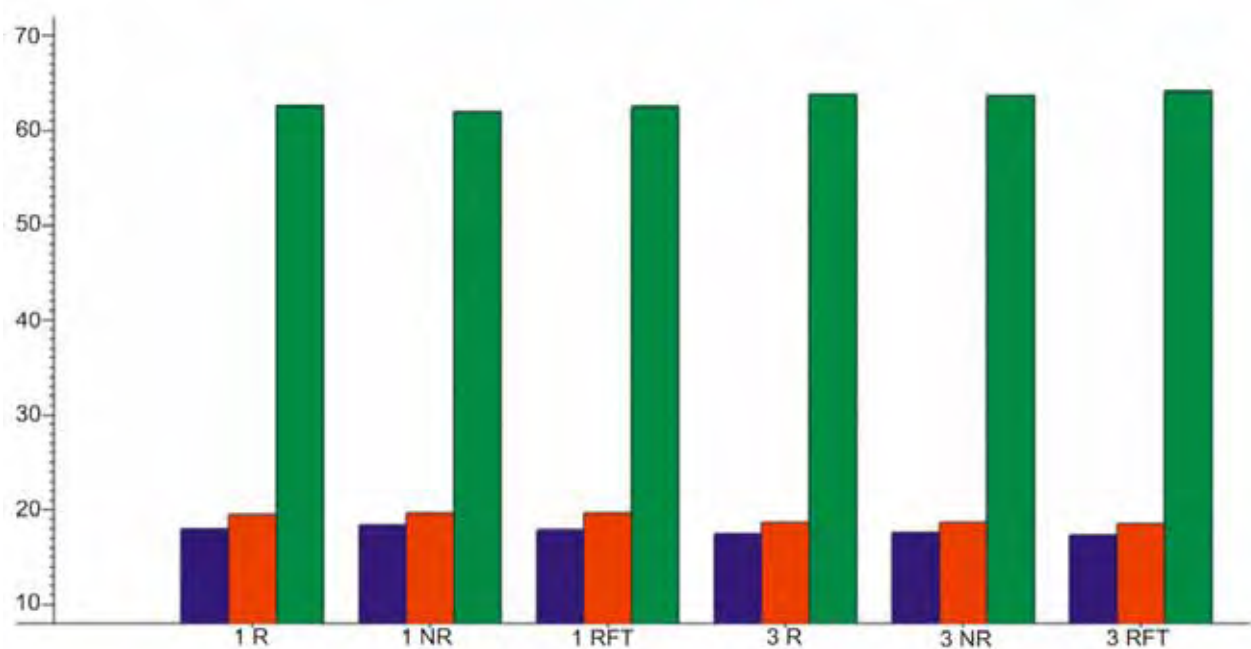


Figura 4. Média da quantidade (%) e cada forma de estrutura secundária resumida por tipos de resposta ao tratamento e/ou genótipo. Azul – Hélice, Vermelho – Folha e Verde – Coil.

4.2. Modelagem Molecular Comparativa

A procura por estruturas já determinadas da NS5A no PDB com identidade de sequência primária superior a 30% resultou em duas estruturas parciais da proteína. A estrutura com código 1R7C correspondia aos resíduos 1 a 31 (região de âncora de membrana), obtida a partir da sequência AF009606 (genótipo 1a) do *Genbank*. A estrutura 1ZH1 correspondia aos resíduos 36 a 198 (domínio de ligação ao zinco), obtida a partir da sequência BAH08638 (genótipo 1b). Desta forma, foram encontrados moldes para os resíduos 1 a 198, praticamente todo o domínio I (resíduos 1 a 213).

Como não foram encontrados moldes a partir do resíduo 199 utilizando o método de alinhamento sequencial, foi utilizado o método de alinhamento estrutural no programa mGenThreader. Os moldes encontrados com similaridade estrutural dos resíduos 199 ao 448/452 são descritos na Tabela 1. A partir destes moldes, foram gerados *templates* quiméricos diferentes para o genótipo 1 e 3, contendo estruturas de sequências parciais de outras proteínas com perfil estrutural semelhante ao da NS5A (Tabela 1 e 2).

Tabela 1. Moldes utilizados para montagem do modelo tridimensional do genótipo 1.

Região da NS5A	Região do molde	Cód. PDB	Nome da proteína
01 - 31	01-31	1R7C	NS5A (parcial)
36 - 198	36 - 198	1ZH1	NS5A (parcial)
199 - 325	09 - 142	1V9N	Malato Dehidrogenase
329 - 360	62 - 165	1OQY	Proteína de Reparo de DNA
382 - 448			
363 - 381	92-118	2C3F	Hialuronidase

Tabela 2. Moldes utilizados para montagem do modelo tridimensional do genótipo 3.

Região da NS5A	Região do molde	Cód. PDB	Nome da proteína
01 - 31	01 - 31	1R7C	NS5A (parcial)
36 - 198	36 - 198	1ZH1	NS5A (parcial)
199 - 325	09 - 142	1VPY	Hialuronidase
329 - 360	62 - 165	1OQY	Proteína de Reparo de DNA
382 - 452			
363 - 381	92-118	2C3F	Hialuronidase

A partir destes moldes, foram modeladas 9 sequências referência (três de cada genótipo).

Para cada sequência foram gerados 1000 modelos. Destes, foi selecionado aquele que apresentou

melhor qualidade estereoquímica utilizando o programa *Procheck* e a função objetiva do *Modeller*.

O programa *Procheck* foi utilizado para analisar a geometria global dos modelos através do Diagrama de Ramachandran (avalia os ângulos de ligação phi (ϕ) e psi (ψ), referentes às ligações C α -N e C α -C, respectivamente) e o *G-factor* (verifica os ângulos de ligação e a geometria covalente da proteína). Uma estrutura de boa qualidade deve ter 90% ou mais de seus resíduos nas regiões favoráveis. As áreas em branco correspondem a regiões onde existem choques estereoquímicos na proteína, válido para todos os aminoácidos exceto a glicina que é o único aminoácido que não tem cadeia lateral e pode adotar os ângulos de torção em todos os quadrantes do diagrama. As áreas em amarelo são regiões limitantes, porém permitidos e as áreas vermelhas correspondem à conformação onde não há nenhum choque estereoquímico, ou seja, regiões permitidas aonde se encontram os aminoácidos em hélices e folhas. Todos os modelos escolhidos apresentaram boa qualidade estereoquímica, com mais de 90% dos resíduos em regiões favoráveis como mostra a Figura 5.

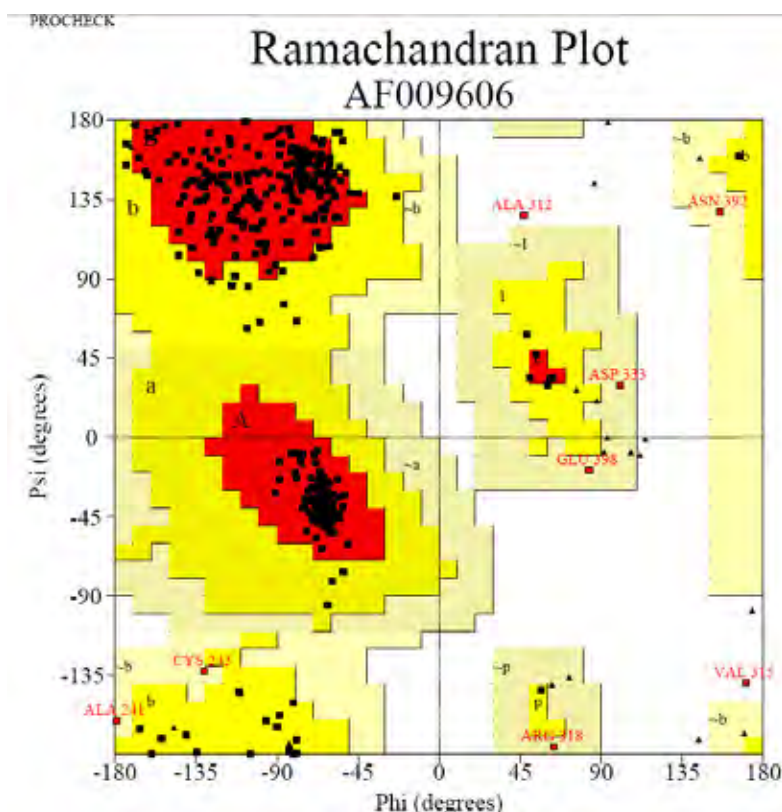


Figura 5. Diagrama de Ramachandran mostrando a avaliação geométrica do modelo gerado para a sequência AF009606 (1a). As áreas em vermelho representam regiões altamente favoráveis (85,2% dos resíduos) para hélice-alfa (a) e folha-beta(b), em amarelo escuro regiões permitidas (12%), em amarelo claro regiões generosamente permitidas (1,6%) e em branco regiões não permitidas (0,5%).

O *G-factor* global é uma medida da qualidade estereoquímica da estrutura como um todo. O *G-factor* médio dos ângulos de torção e geometria covalente representa uma medida do desvio de uma dada propriedade estereoquímica para cada resíduo de aminoácido da estrutura. O valor ideal deste fator deve ser acima de -0,5. Valores no intervalo de -0,5 a -1,0 significam que existem resíduos na proteína que devem ser investigados e, valores abaixo de -1,0 indicam que os modelos apresentam qualidade estereoquímica ruim e devem ser investigados. O *G-factor* global é a melhor medida da qualidade estereoquímica total de uma molécula, pois representa a média de todos os *G-factor* obtidos para cada aminoácido da estrutura. Os valores de *G-factor* médio e global calculados para os modelos gerados indicam que a estrutura elaborada possui boa qualidade estereoquímica e compatível com o molde utilizado, uma vez que todos os valores estão acima de -0,50 (Tabela 3).

Tabela 3. Valores do *G-factor* para os modelos gerados das sequências referências do Genbank da proteína NS5A do HCV genótipos 1 e 3.

Nome da sequência referência	Genótipo	<i>G-factor</i> médio dos Ângulos de torção	<i>G-factor</i> médio da Geometria covalente	<i>G-factor</i> Global
AF009606	1a	-0,12	-0,31	-0,19
AF011752	1a	-0,17	-0,44	-0,26
NC004102	1a	-0,12	-0,31	-0,19
EU256081	1b	-0,14	-0,35	-0,22
D50481	1b	-0,11	-0,33	-0,19
AF356877	1b	-0,15	-0,37	-0,22
NC009824	3	-0,11	-0,36	-0,20
AM493623	3	-0,12	-0,35	-0,21
D17763	3	-0,11	-0,36	-0,20

Através do programa *Whatcheck* foi constatado que os ambientes químicos de resíduos individuais se enquadram nos padrões freqüentemente encontrados em estruturas protéicas. Segundo os valores de referência do programa para modelagem molecular, os modelos finais podem ser considerados de boa qualidade; os relatórios não serão apresentados aqui por serem extensos.

Para analisar o desvio médio da posição dos átomos de uma estrutura, calculou-se a raiz do desvio-médio quadrático (RMSD) da geometria ideal (Tabela 4). O valor do RMSD da geometria ideal para o comprimento e ângulo de ligação foi calculado através do programa X-PLOR, que gera um arquivo com os valores do RMSD para as ligações e ângulos diedros e impróprios bem como contatos não ligados. Valores ideais para o comprimento e ângulo de ligação devem estar abaixo de 0,06Å e 10,0°, respectivamente, para que a geometria da proteína como um todo esteja em equilíbrio. Os valores de RMSD calculados para o comprimento de ligação (0,021 Å) e para o ângulo de ligação (de 2,377° a 2,898°) indicam que a geometria do modelo teórico elaborado para a NS5A completa dos genótipos 1 e 3 está em equilíbrio.

A partir destas avaliações foi constatado que por meio do *template* quimérico elaborado é possível realizar a modelagem molecular comparativa com sucesso para a proteína NS5A dos genótipos 1 e 3.

Tabela 4. Valores do desvio médio-quadrático (RMSD) para os modelos gerados das sequências referências do Genbank da proteína NS5A do HCV genótipos 1 e 3.

Nome da sequência referência	Genótipo	RMSD para o comprimento de ligação	RMSD para o ângulo de ligação
AF009606	1a	0,021 Å	2,618°
AF011752	1a	0,021 Å	2,796°
NC004102	1a	0,021 Å	2,618°
EU256081	1b	0,021 Å	2,898°
D50481	1b	0,021 Å	2,621°
AF356877	1b	0,021 Å	2,672°
NC009824	3	0,021 Å	2,377°
AM493623	3	0,021 Å	2,648°
D17763	3	0,021 Å	2,377°

O modelo gerado da sequência referência AF009606 é mostrado na Figura 6. Outros modelos da proteína NS5A do HCV resultantes da modelagem molecular por homologia e *threading* são mostrados nas Figuras 8A, 9A e 10A.

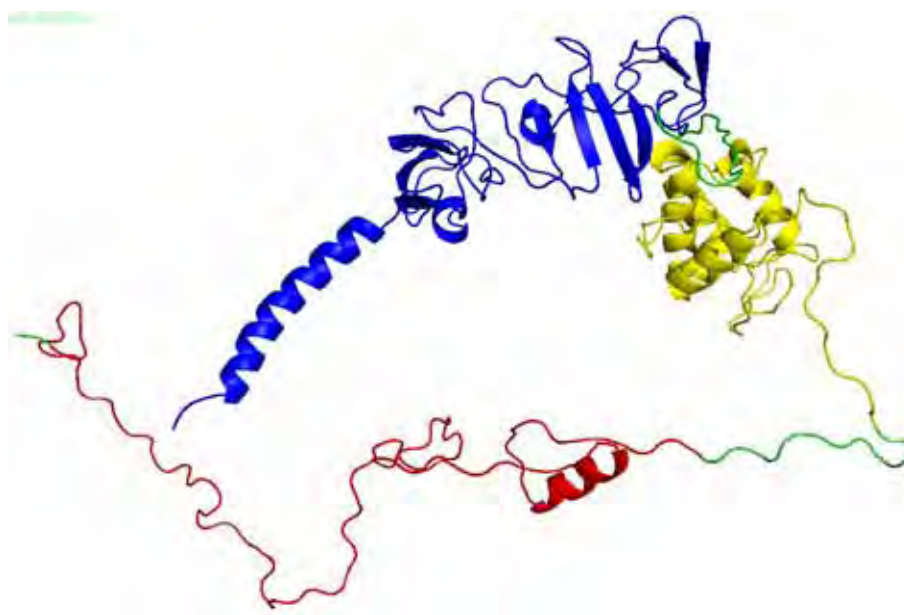


Figura 6. Modelo estrutural da NS5A gerado por sequência-referência AF009606 (genótipo 1a), mostrando em azul o domínio I, em amarelo o domínio II e em vermelho o domínio III.

4.3. Dinâmica Molecular

As simulações de dinâmica molecular foram realizadas com o objetivo de avaliar a estabilidade do modelo gerado e investigar o possível dobramento da proteína NS5A. Os

resultados das análises de RMSD para os modelos são mostrados na Figura 7 . Podemos observar que os modelos conseguem se estabilizar numa mesma faixa de valores. O RMSD dos modelos também mostra que a estrutura vai sofrendo mudanças em sua conformação ao longo do tempo por isso o aumento gradativo.

Passaram por esta análise todas as sequências referência do Genbank e sequências que apresentaram alterações descritas na literatura como comprometedoras da função *in vitro* (descritas na Tabela 6, pág.44).

Na Tabela 5, são mostrados um resumo dos parâmetros utilizados na dinâmica molecular da NS5A. Através dessa análise foi possível constatar que houve compactação de todas formas analisadas da proteína ao final da dinâmica, como também podemos conferir comparando os modelos antes e depois da aplicação da dinâmica (Figuras 8 a 10).

A dinâmica molecular de anelamento não foi realizada, pois a grande quantidade de região não-estruturada da proteína não permitiria o retorno à sua forma após a diminuição de temperatura.

Tabela 5. Ficha técnica das condições gerais das simulações realizadas por dinâmica molecular.

Tempo de simulação	5ns (5000000 passos de 2fs)
Programa	NAMD 2.7b 1
Campo de Força	CHARM22
Interação Eletrostática	Raio de Corte
Raio de Corte (Coulomb)	1,4° angstroms
Interação de van der Waals	Raio de Corte
Raio de Corte (van der Waals)	1,4 angstroms
Outros	Temperatura constante de 310K (Berendsen) Pressão Constante de 1 atm (Berendsen) Condições Periódicas de Contorno (PBC)

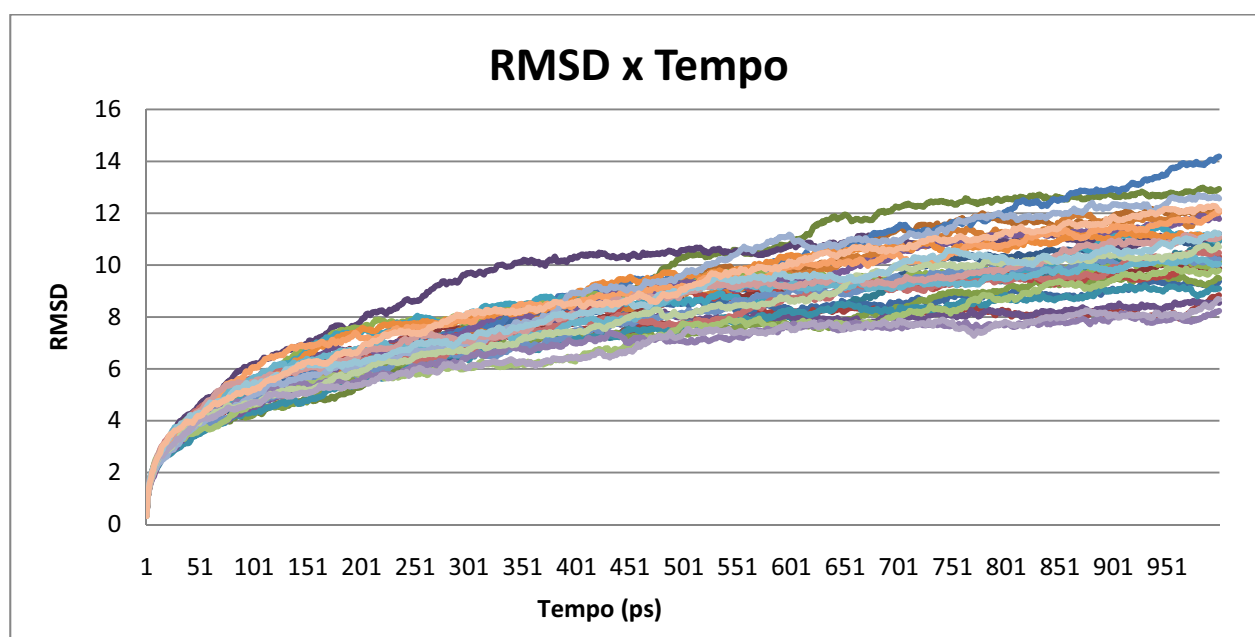


Figura 7. Gráfico mostrando a alteração de RMSD durante a dinâmica molecular. Cada linha representa uma sequência diferente da NS5A.

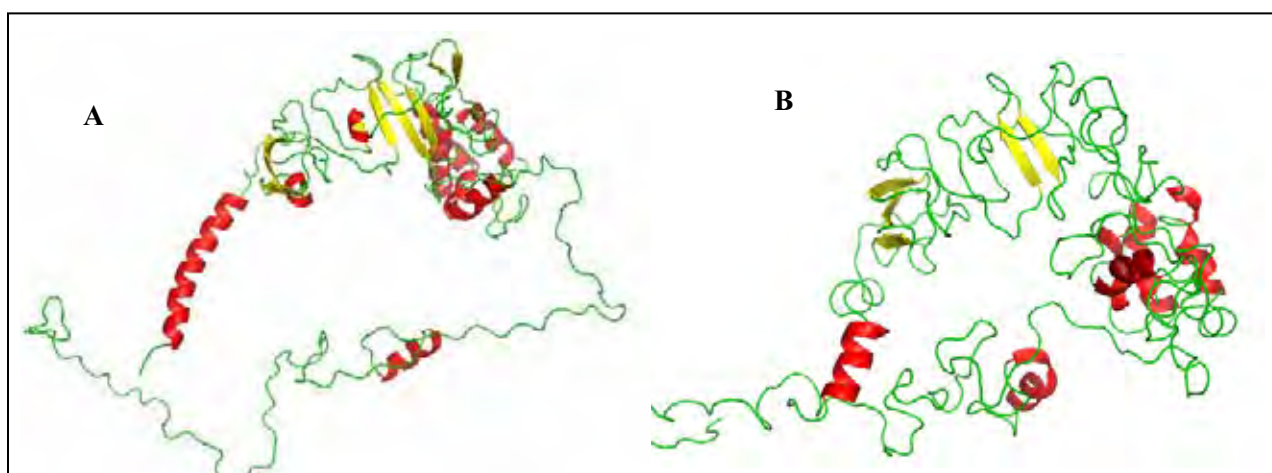


Figura 8. Modelos estruturais da proteína NS5A antes (A) e após (B) a análise por dinâmica molecular da sequência referência AF009606 (genótipo 1a).

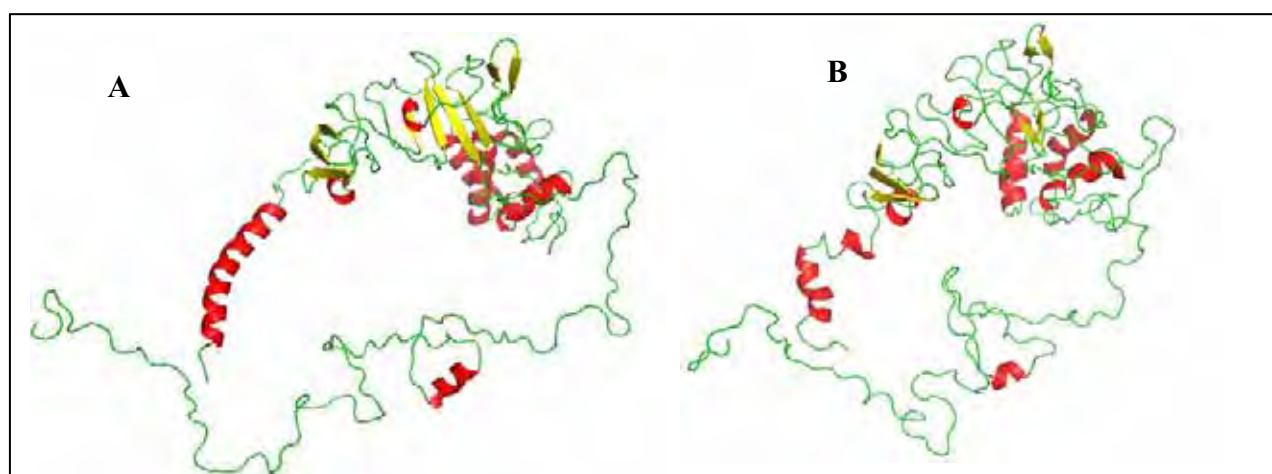


Figura 8. Modelos estruturais da proteína NS5A antes (A) e após (B) a análise por dinâmica molecular da sequência referência D50481 (genótipo 1b).

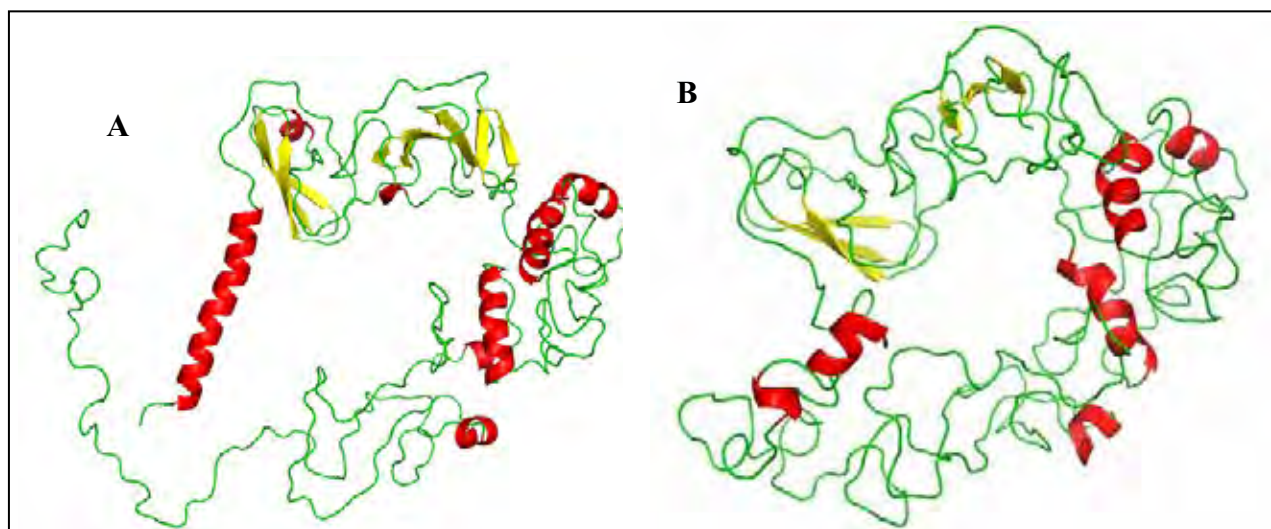


Figura 9. Modelos estruturais da proteína NS5A antes (A) e após (B) a análise por dinâmica molecular da sequência referência D17763 (genótipo 3).

4.4. Predição de regiões transmembrânicas

Todas as sequências-alvo e as sequências referência, analisadas pelo MEMSAT3 da NS5A completa dos genótipos 1 e 3 apresentaram a predição de uma região transmembrânica próximo ao N-terminal, entre os resíduos 32-51 (Figura 11).

1: (in) 32-51 (-3.18)

```

      10      20      30      40      50      60
+++++IIIIIXXXXXXXXXXXOOOOO-----
SGSWLRDIDWICEVLSDFRTWLKAKLMPQLPGIPLVSCQRGYRGVWRGDGIMHTRCHCGA
      70      80      90     100     110     120
-----
EITGHVKNGTMRIVGPKTCRNMWSGTFPINAYTTGPCTPLPAPNYKFALWRVSAEEYVEV
      130     140     150     160     170     180
-----
RQVGDFHYVTGMTTDNLKCPQVPSPEFFTELDGVRLLHRFAPPCKPLLREEVSFRVGLHE
      190     200     210     220     230     240
-----
YPVGSQPLPCEPEPDVAVLTSMLTDPSSHITAEAAKRRRLARGSPPSMASSASQLSAPSLKG
      250     260     270     280     290     300
-----
TCTANHDSPPDAELIEANLLWRQEMGGNITRVESENKVVVLDSFDPLVAEEDEREISVPAE
      310     320     330     340     350     360
-----
ILRKSRRFTSALPIWARPDYNPPLVETWKKPDYEPVHVHGCPLPPPKSPFVPPPRKKRTV
      370     380     390     400     410     420
-----
VLTESSVSTALAEATKSFSGSSSTSGITSDNTATSPEFAPSGCSPDSDAESYSSMPPLEG
      430     440
-----
EPGDPDLSDGSWSTVSSEAGTEDVVC

```

Legenda:

+ : Dentro do loop

- : Fora do loop

X : Segmento da hélice
transmembrânica central

I : Dentro da capa da
hélice

Figura 11. Resultado da predição de região transmembrânica do programa MEMSAT3.

4.5. Predição de Sítios na NS5A

As 345 sequências e as 9 sequências referências da NS5A do genótipo 1 e 3 foram analisadas quanto à presença de padrões, realizado pelo PROSITE. Os resultados mostraram que

a NS5A apresenta diversos *motivos*, principalmente de modificações pós-traducionais. Os sítios mais conservados são listados a seguir, com modelos demonstrando a localização dos sítios na sequência referência AF009609 (1a). Todos os sítios identificados foram localizados em regiões de superfície da proteína. Não foram identificadas correlações entre o número ou a localização dos *motivos* e a resposta ao tratamento ou o genótipo/subtipo.

- **Sítios de N-glicosilação:** Possíveis sítios de glicosilação (sequência consenso N-{P}-[ST]-{P}) foram encontrados em todas as sequências do genótipo 1 e 3. O número de sítios encontrados variou de 2 a 5 para o genótipo 1 e de 3 a 6 para o genótipo 3. Em duas posições, esta sequência apresentou-se conservada. Todas as sequências, exceto na referência AF011752 (1a) apresentaram sítio de glicosilação na posição 69 e todas as sequências, exceto as de um paciente do genótipo 3 (NR) apresentaram na posição 268. Entretanto, todas as sequências deste mesmo paciente apresentaram um sítio em uma região próxima, na posição 258 que não foi encontrada em sequências de outros pacientes. Outras posições de glicosilação foram variáveis. Não foi observada relação entre quantidade e/ou localização dos sítios de glicosilação e a resposta ao tratamento (Figura 12).

- **Sítio de Fosforilação por Proteína Quinase Dependente de cAMP (Adenosina monofosfato cíclico) e cGMP (Guanosina monofosfato cíclico):** A análise pelo PROSITE também encontrou possíveis sítios de fosforilação por proteína quinase dependente cAMP e cGMP (sequência consenso [RK](2)-x-[ST]). O número de sítios encontrados variou de nenhum, em sequências de um paciente RFT do genótipo 1 e dois NR do genótipo 3. As sequências dos demais pacientes apresentaram pelo menos um sítio na posição 357 (Figura 13).

- **Sítio de Fosforilação por Proteína Quinase C (PKC):** Nas 345 sequências-alvo e referência da NS5A, foram identificados vários possíveis sítios de fosforilação por PKC (sequência consenso [ST].[RK]). Estes sítios se mostraram altamente conservados nas posições 71, 238 e 328, sendo identificados em todas as sequências. Entretanto, as sequências de todos os pacientes apresentaram

pelo menos 5 sítios de fosforilação (em sequências de dois pacientes com o genótipo 1, um NR e outro RFT), podendo variar até o número de oito (nas sequências de um paciente do genótipo 3 NR) (Figura 14).

- **Sítio de Fosforilação por Caseína Quinase II (CKII):** A análise pelo PROSITE, demonstrou que a NS5A dos genótipos 1 e 3 possuem vários possíveis sítios de fosforilação por CKII (sequência consenso [ST]-x(2)-[DE]). Não foram identificadas posições conservadas para esse *motivo*, que foi encontrado em regiões dos três domínios. Entretanto, todas as sequências apresentaram este sítio. O número de sítios variou de quatro (em sequências de um paciente com o HCV 1a RFT) até oito (em sequências de dois pacientes com o HCV 1a, um RFT e outro NR e um do genótipo 3 R) (Figura 15).

- **Sítio de Fosforilação por Tirosina Quinase:** A busca por padrões referentes ao sítio de fosforilação por tirosina quinase (sequência consenso [RK]-x(3)-[DE]-x(2)-Y ou [RK]-x(3)-[DE]-x(2)-Y) em todas as sequências da NS5A completa, resultou na identificação de um possível sítio altamente conservado nas posições 122 ou 123. Apenas três *quasispecies* do HCV de três pacientes (um do genótipo 1 RFT e dois do genótipo 3 NR) não apresentaram nenhum sítio de fosforilação por tirosina quinase. As sequências de um paciente com o HCV do genótipo 1a NR apresentou um possível sítio na posição 406 também (Figura 16).

- **Sítio de N-miristoilação:** Possíveis sítios de N-miristoilação (sequência consenso G-{EDRKHPFYW}-x(2)-[STAGCN]-{P}) foram encontrados em todas as sequências, com variação de número e localização, mesmo entre sequências da NS5A obtidas da amostra de um mesmo paciente. O número de sítios identificados variou de quatro (em um paciente NR infectado com o HCV 3) até oito (em cinco pacientes com o HCV 1 que apresentaram diferentes respostas ao tratamento) (Figura 17).

- **Sequência de Ligação Celular:** A busca por padrões da sequência de ligação celular (sequência consenso RGD) encontrou um *motivo* em todas as sequências-alvo e referência na posição 48. Este

sítio apresentou-se altamente conservado quanto à localização e número em todas as sequências da NS5A do genótipo 1 e 3 (Figura 18).



Figura 12. Representação em *cartoon* da NS5A completa, de acordo com a sequência referência AF009606 (genótipo 1a), demonstrando os sítios de N-glicosilação representados em *stick* e superfície.



Figura 13. Representação em *cartoon* da NS5A completa, de acordo com a sequência referência AF009606 (genótipo 1a), demonstrando o sítio de fosforilação dependente de cAMP e c GMP representados em *stick* e superfície.



Figura 14. Representação em *cartoon* da NS5A completa, de acordo com a sequência referência AF009606 (genótipo 1a), demonstrando os Sítios de fosforilação por PKC representados em *stick* e superfície.

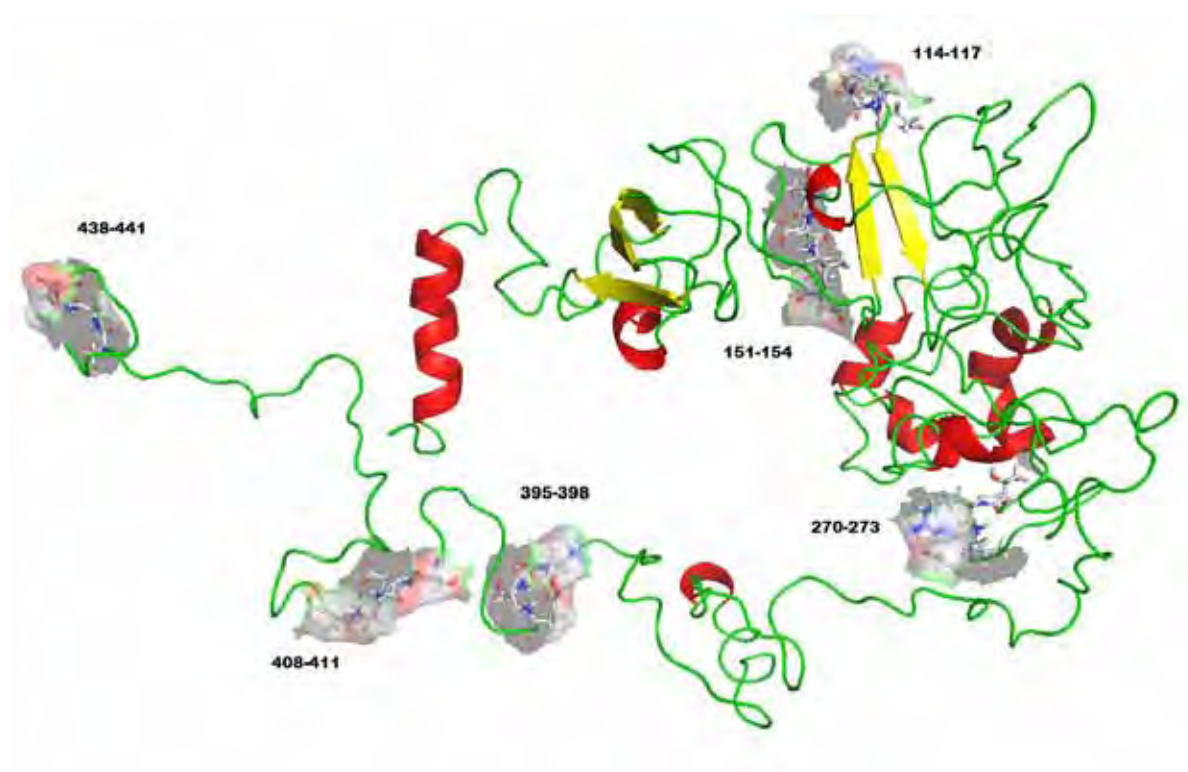


Figura 15. Representação em *cartoon* da NS5A completa, de acordo com a sequência referência AF009606 (genótipo 1a), demonstrando os sítios de fosforilação por CKII representados em *stick* e superfície.

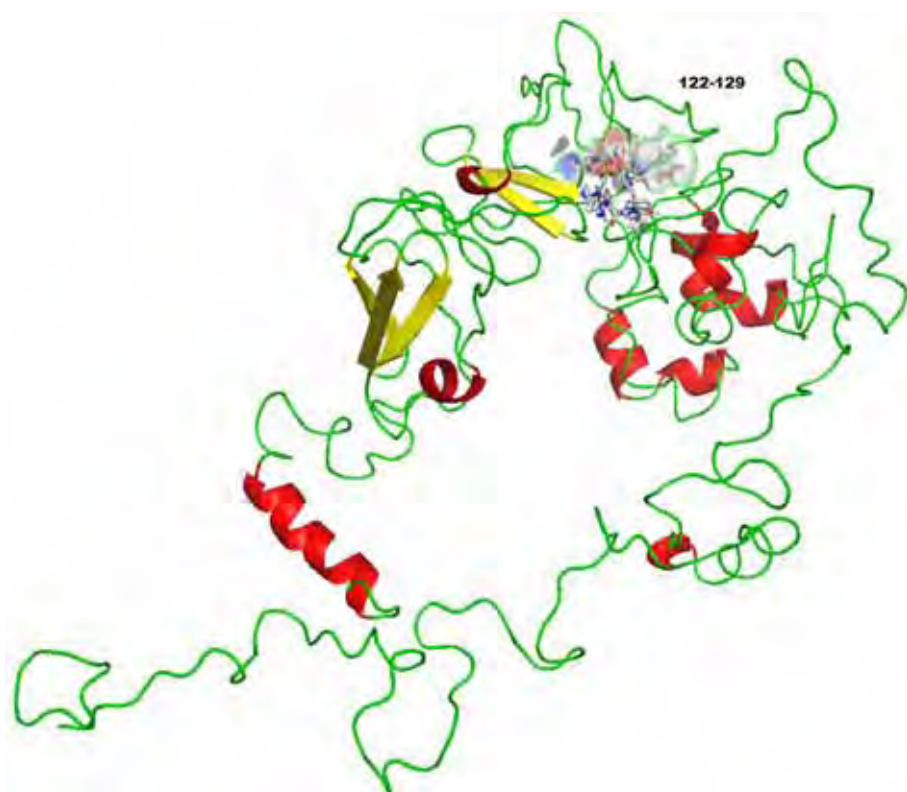


Figura 16. Representação em *cartoon* da NS5A completa, de acordo com a sequência referência AF009606 (genótipo 1a), demonstrando o sítio de fosforilação por Tirosina Quinase da representado em *stick* e superfície.

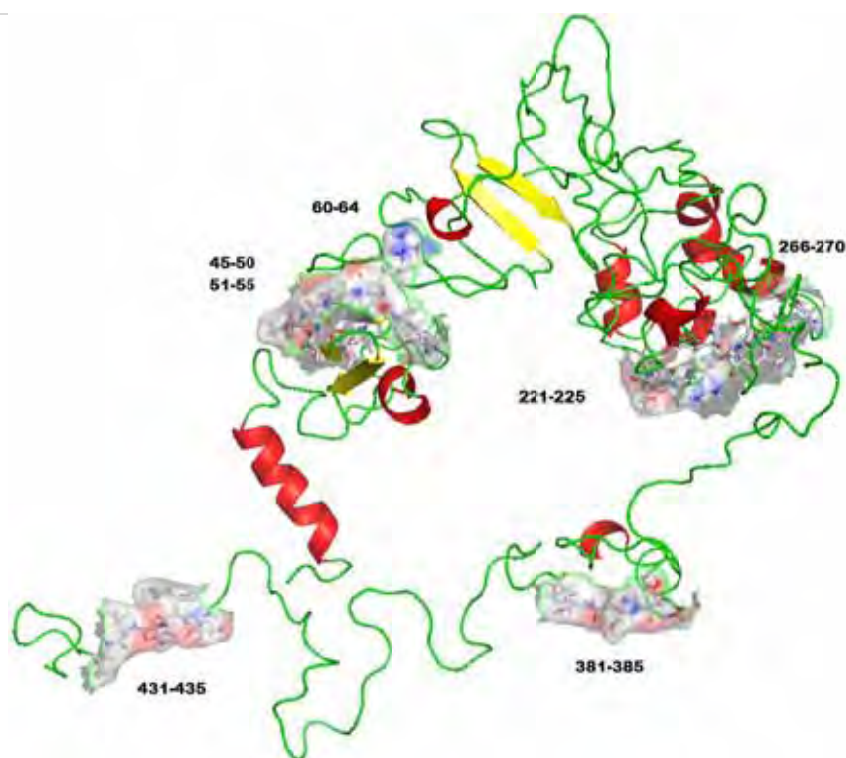


Figura 17. Representação em *cartoon* da NS5A completa, de acordo com a sequência referência AF009606 (genótipo 1a), demonstrando os sítios de n-miristoilação representados em *stick* e superfície.

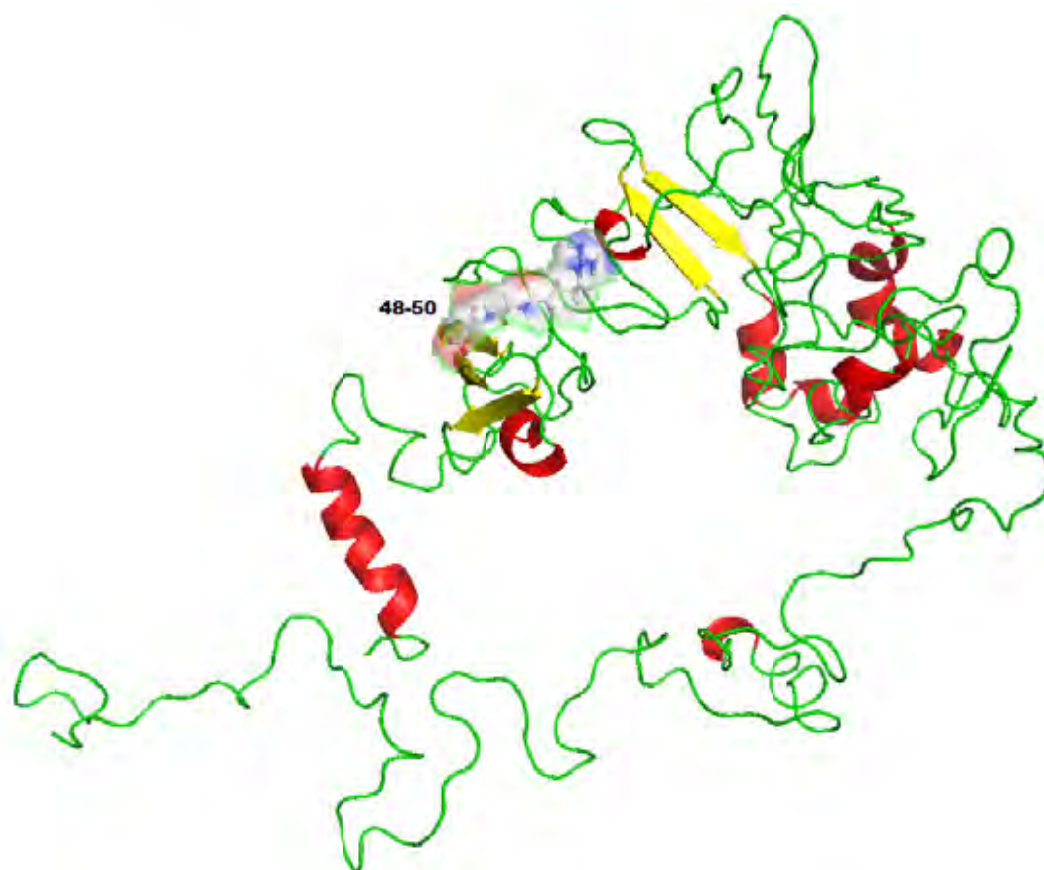


Figura 18. Representação em *cartoon* da NS5A completa, de acordo com a sequência referência AF009606 (genótipo 1a), demonstrando o sítio de ligação celular da referência AF009606 (genótipo 1a) representados em *stick* e superfície.

4.6. Predição de Função da NS5A

Foram analisadas no *ProtFun* as sequências completas da proteína NS5A e os domínios separados. Para a NS5A completa, 60% das sequências foram classificadas na categoria funcional de metabolismo central intermediário e outras em funções regulatórias, purinas e pirimidinas e replicação e transcrição (Figura 19). Todas as sequências foram classificadas como enzimas de classe desconhecida e categoria de ontologia do gene de regulação e transcrição.

As sequências do domínio I foram todas classificadas na categoria funcional de metabolismo energético. Todas foram classificadas como enzimas, entretanto 99% das sequências não tiveram classe enzimática definida, 1% fora classificada como hidrolase. De todas as sequências, 82% foram classificadas como fator de crescimento e 17% como regulação da transcrição na categoria de ontologia do gene (Figura 20). A análise do domínio II demonstrou que 64% das sequências foram classificadas na categoria funcional de tradução, 30% na categoria de transporte e ligação e 5% na categoria de metabolismo energético (Figura 21). Sessenta e um por cento (61%) foram classificadas em enzimas e 39% em não-enzimas (Figura 22). Todas as sequências classificadas como enzimas não tiveram classe enzimática definida pelo programa. Dentre as categorias de ontologia do gene, 52% das sequências do domínio II foram classificadas como fator de crescimento, 26% como transcrição e outros como hormônio, canal de íon dependente de voltagem, resposta imune, regulação da transcrição ou desconhecida (Figura 23). As sequências do domínio III foram classificadas na categoria funcional de transporte e ligação (99%) ou funções regulatórias (1%) (Figura 24). Todas as sequências foram classificadas na categoria de não-enzima. O *ProtFun* classificou 44% das sequências na categoria ontológica de fator de crescimento, 35% em hormônio, 17% em categoria desconhecida e as outras em resposta ao estresse, transdutor de sinal ou canal de íon dependente de voltagem (Figura 25).

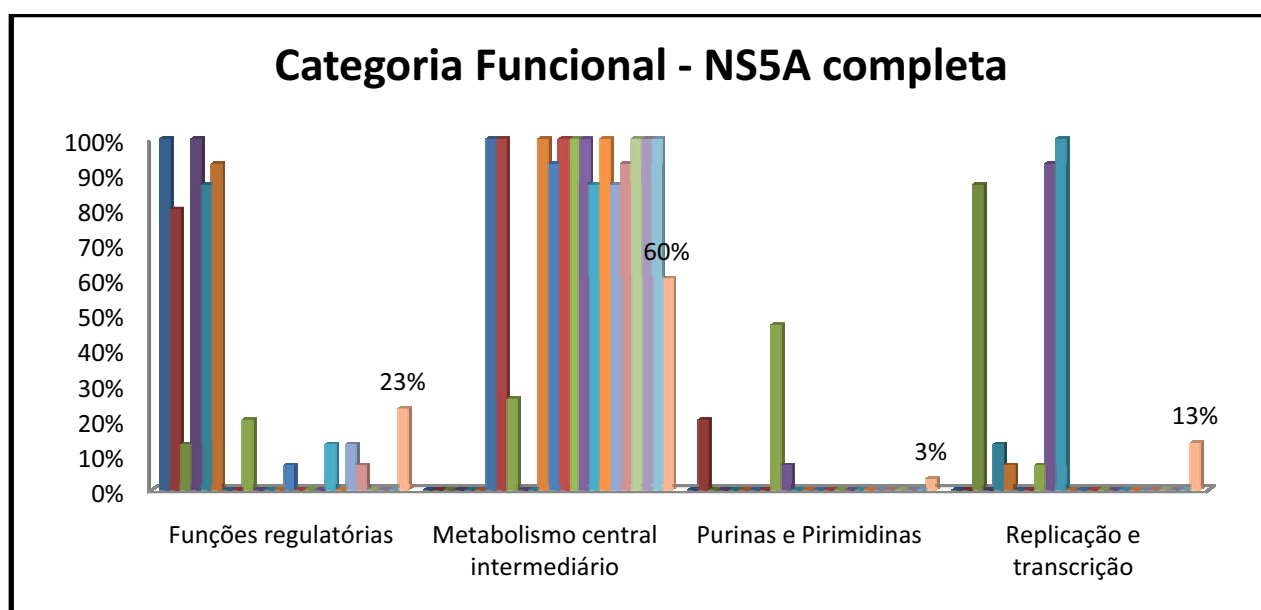


Figura 19. Gráfico demonstrando a porcentagem (eixo y) de sequências da NS5A completa em cada uma das categorias funcionais (eixo x) classificadas pelo *ProtFun*. Cada barra representa as sequências de um paciente diferente, exceto a barra com os valores indicados que representam as porcentagens levando em consideração todas as sequências.

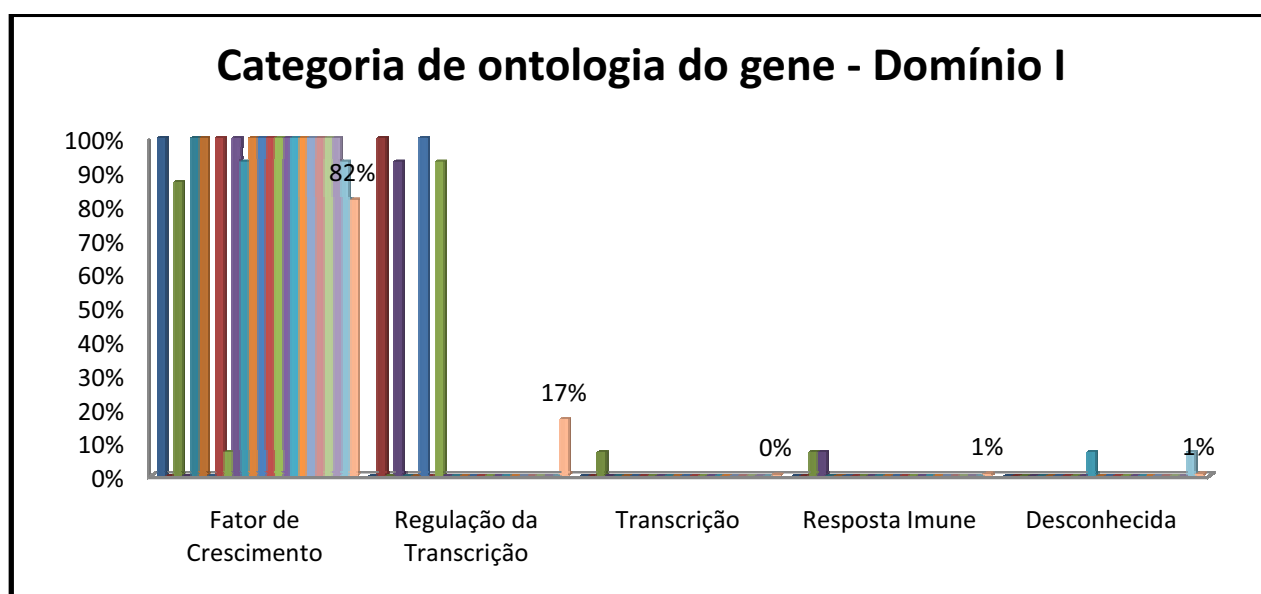


Figura 20. Gráfico demonstrando a porcentagem (eixo y) de sequências da NS5A-DI em cada uma das categorias de ontologias do gene (eixo x) classificadas pelo *ProtFun*. Cada barra representa as sequências de um paciente diferente, exceto a barra com os valores indicados que representam as porcentagens levando em consideração todas as sequências.

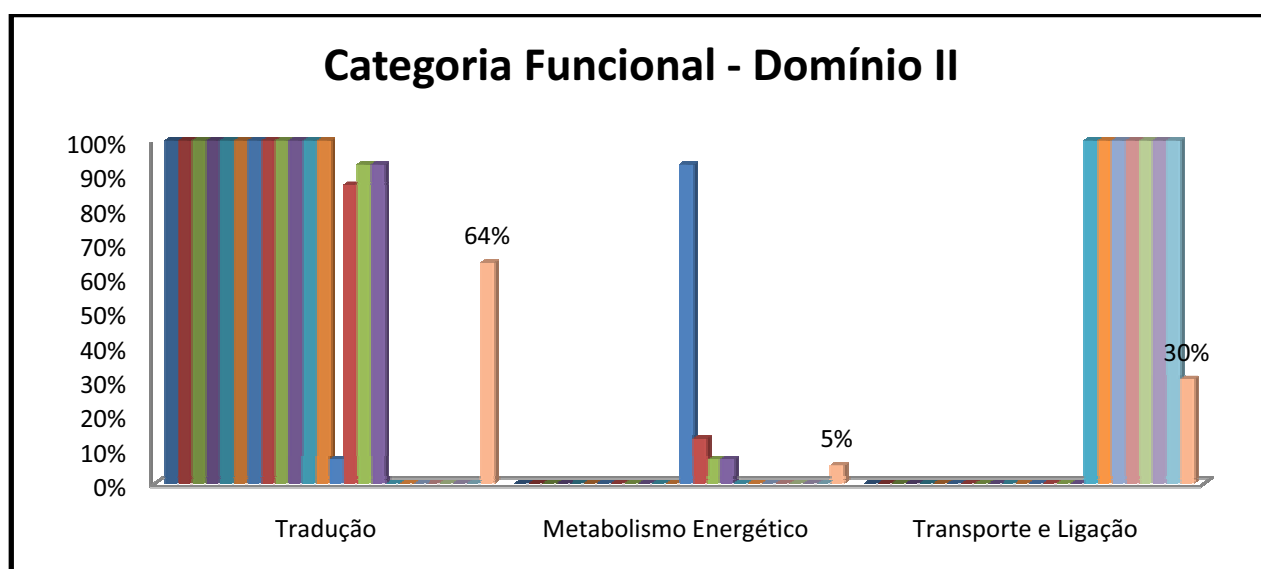


Figura 21. Gráfico demonstrando a porcentagem (eixo y) de sequências da NS5A-DII em cada uma das categorias de ontologias do gene (eixo x) classificadas pelo *ProtFun*. Cada barra representa as sequências de um paciente diferente, exceto a barra com os valores indicados que representam as porcentagens levando em consideração todas as sequências.

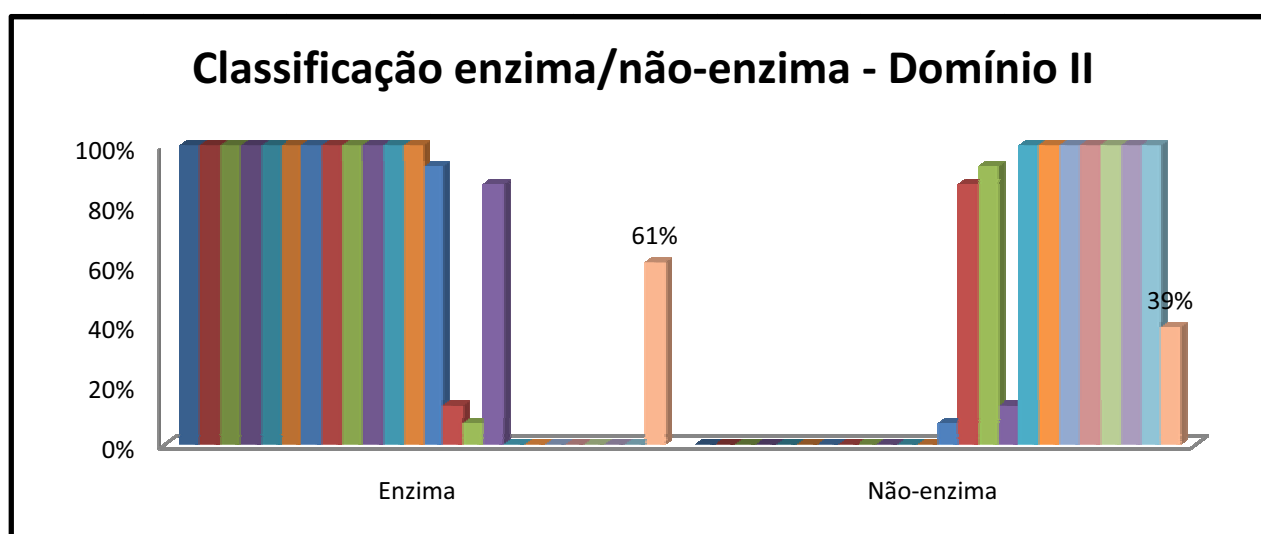


Figura 22. Gráfico demonstrando a porcentagem (eixo y) de sequências da NS5A-DII na classificação de enzima ou não-enzima (eixo x) do *ProtFun*. Cada barra representa as sequências de um paciente diferente, exceto a barra com os valores indicados que representam as porcentagens levando em consideração todas as sequências.

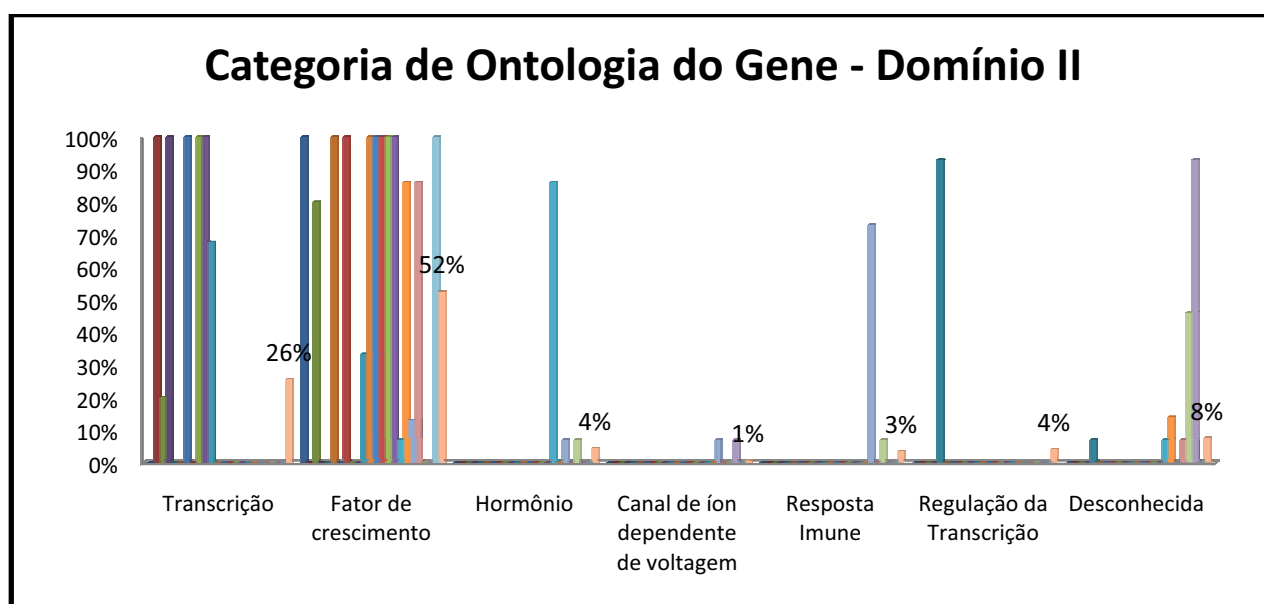


Figura 23. Gráfico demonstrando a porcentagem (eixo y) de sequências da NS5A-DII nas categorias de ontologia do gene (eixo x) classificadas no *ProtFun*. Cada barra representa as sequências de um paciente diferente, exceto a barra com os valores indicados que representam as porcentagens levando em consideração todas as sequências.

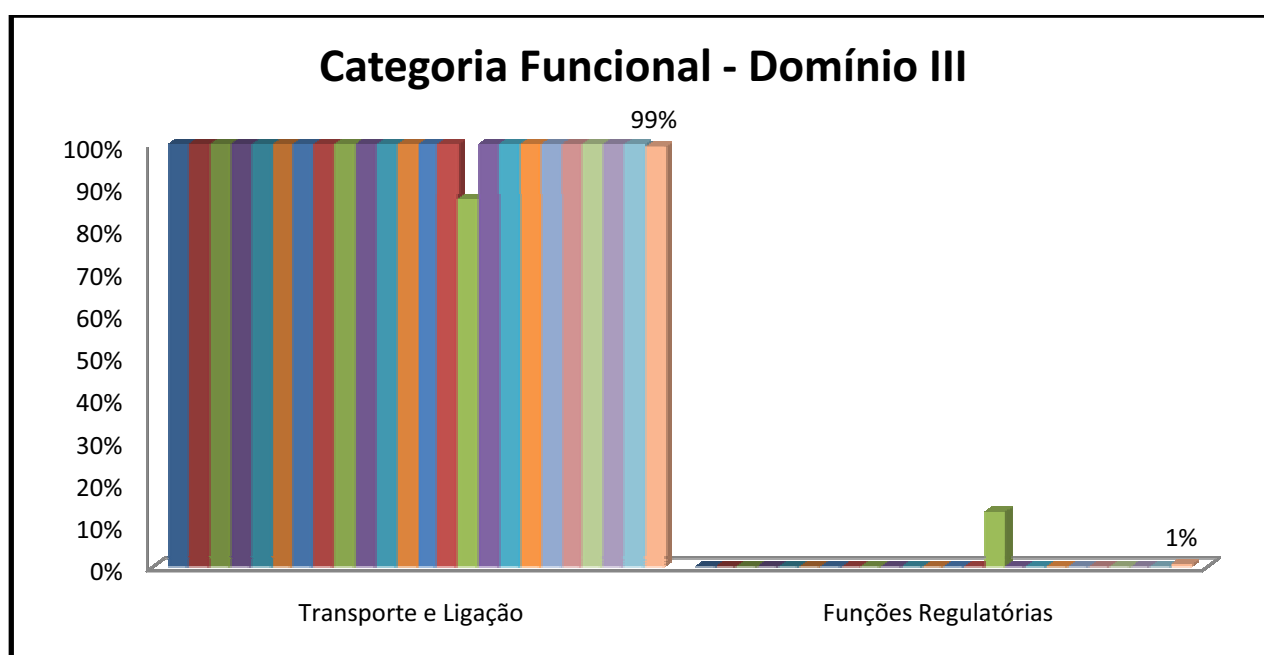


Figura 24. Gráfico demonstrando a porcentagem (eixo y) de sequências da NS5A-DIII nas categorias de ontologia do gene (eixo x) classificadas no *ProtFun*. Cada barra representa as sequências de um paciente diferente, exceto a barra com os valores indicados que representam as porcentagens levando em consideração todas as sequências.

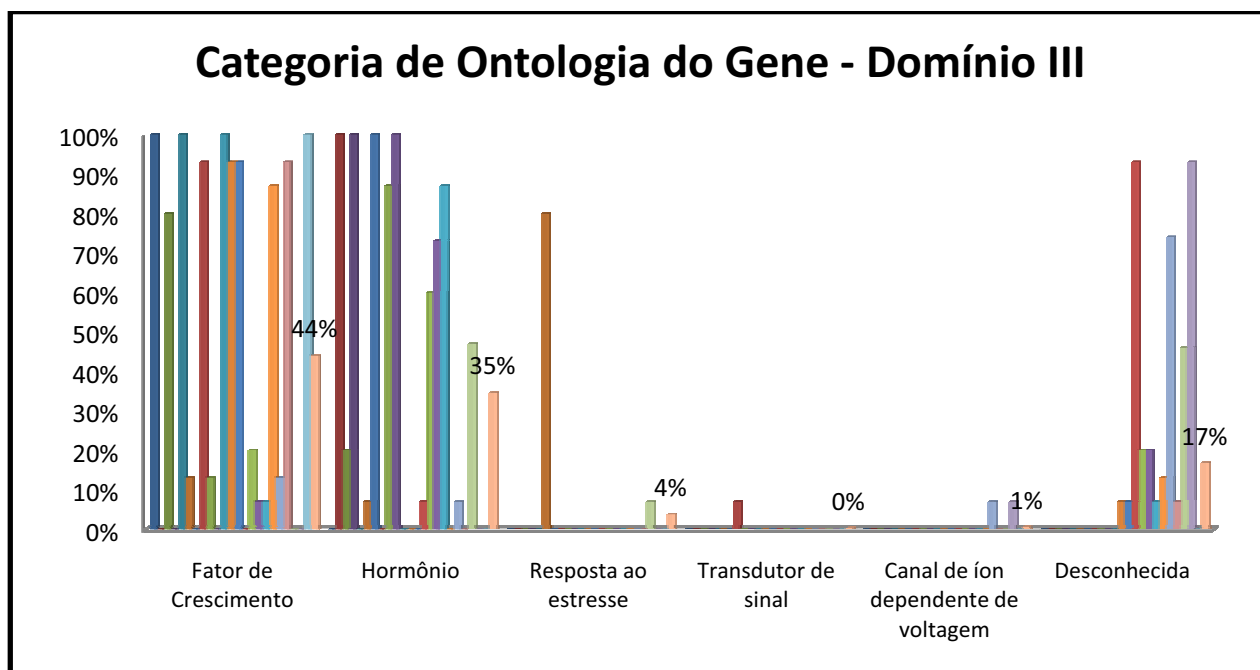


Figura 25. Gráfico demonstrando a porcentagem (eixo y) de sequências da NS5A-DIII nas categorias de ontologia do gene (eixo x) classificadas no *ProtFun*. Cada barra representa as sequências de um paciente diferente, exceto a barra com os valores indicados que representam as porcentagens levando em consideração todas as sequências.

4.7. Análise de mutantes da NS5A

De acordo com as mutações descritas como comprometedoras da função da NS5A por Penin et al. (2004), que analisou a âncora de membrana (resíduos 01 a 33), foram encontradas 2 posições com substituição de aminoácidos no genótipo 1 e 6 para o genótipo 3, incluindo uma deleção dos resíduos 03 a 27 encontrados em na amostra de dois pacientes respondedores ao final do tratamento. A análise baseada nas mutações no domínio II e III descritas por Tellinghuisen et al. (2008) como comprometedoras da função, mostraram que das sequências analisadas da NS5A, 2 posições de alteração foram encontradas no genótipo 1 e 2 no genótipo 3. Alterações para aminoácidos de propriedades semelhantes à da sequência nativa não foram analisadas. Não foi encontrada nenhuma alteração que pudesse ser relacionada ao tipo de resposta ao tratamento. A descrição de todas as alterações encontradas por posição é detalhada na Tabela 6.

Tabela 6. Resumo das alterações encontradas da NS5A descritas como comprometedoras da função na literatura.

	Forma nativa*	Forma mutante*	Número de sequências com alteração
02	Glicina (A)	Ácido Aspártico(P+)	2 (Genótipo 1)
06	Arginina(P-)	Cisteína(P0)	1 (Genótipo 3)
09	Triptofano(A)	(deleção)	2 (Genótipo 1)
		Arginina(P-)	2 (Genótipo 3)
10	Ácido Aspártico(P+)	Treonina(P0)	1 (Genótipo 3)
11	Triptofano(A)	Arginina(P-)	1 (Genótipo 3)
16	Leucina (A)	Serina(P0)	1 (Genótipo 3)
3-27	CRS	(deleção)	2 (Genótipo 3)
326 a 328	Leucina(A), Ácido Glutâmico(P+), Serina(P0)	Triptofano(A), Asparagina(P0), Arginina(P-)	1 (Genótipo 3)
335 a 337	Valina(A), Prolina(A), Prolina(A)	Ácido glutâmico(P+), Prolina(A), Isoleucina (A)	1 (Genótipo 1)
342	Prolina(A)	Serina(P0)	1 (Genótipo 1)

* característica do aminoácido (+) – ácido, (-) – básico, (0) – neutro, (P) – polar, (A) - apolar

Através da análise estrutural e bioquímica das alterações, pode-se observar que a substituição na posição 02 de glicina para ácido aspártico causa alteração da hidrofobicidade do início da hélice que forma a âncora de membrana (Figura 26). Essa alteração foi encontrada em duas sequências na amostra de um paciente NR ao tratamento do genótipo 1.

A alteração na posição 06 de arginina para cisteína na proteína NS5A causa uma alteração na carga desta posição, uma vez que a arginina é positivamente carregada, enquanto que a cisteína é neutra (Figura 27). Essa substituição foi encontrada em uma sequência extraída de um paciente respondedor ao final do tratamento, infectado com o HCV genótipo 3.

A deleção do aminoácido triptofano no resíduo 09 causa a perda de uma volta da hélice da CRS. Em adição, a ausência desse resíduo causa diminuição de hidrofobicidade desta região, uma vez que o triptofano é um resíduo de alta hidrofobicidade (Figura 28). Esta deleção foi presente em duas sequências do genótipo 1, de um paciente não-respondedor e de um respondedor. Nesta mesma posição, foi encontrada uma substituição de triptofano para arginina, que causa uma mudança de polaridade e carga nesta região de hélice, uma vez que diferente do triptofano, o

aminoácido arginina é polar e básico. Além desta alteração, é possível que a hélice seja “quebrada” nesta região, pois a arginina é um resíduo com menor probabilidade de formar hélices (Figura 29). Esta substituição foi encontrada em duas sequências do genótipo 3, presentes em amostras de um mesmo paciente respondedor ao final do tratamento.

A alteração no resíduo 10 de ácido aspártico para treonina, causa uma alteração de carga no local, uma vez que a treonina é neutra e não ácida (Figura 30). Essa substituição foi encontrada na sequência da NS5A do genótipo 3, da amostra de um paciente respondedor ao final do tratamento.

A substituição de triptofano para arginina na posição 11 alterou a hidrofobicidade e a carga da região da CRS. O triptofano é um resíduo apolar e a arginina é polar com característica básica (Figura 31). Essa alteração foi detectada em uma sequência do genótipo 3, em um paciente que respondeu ao tratamento.

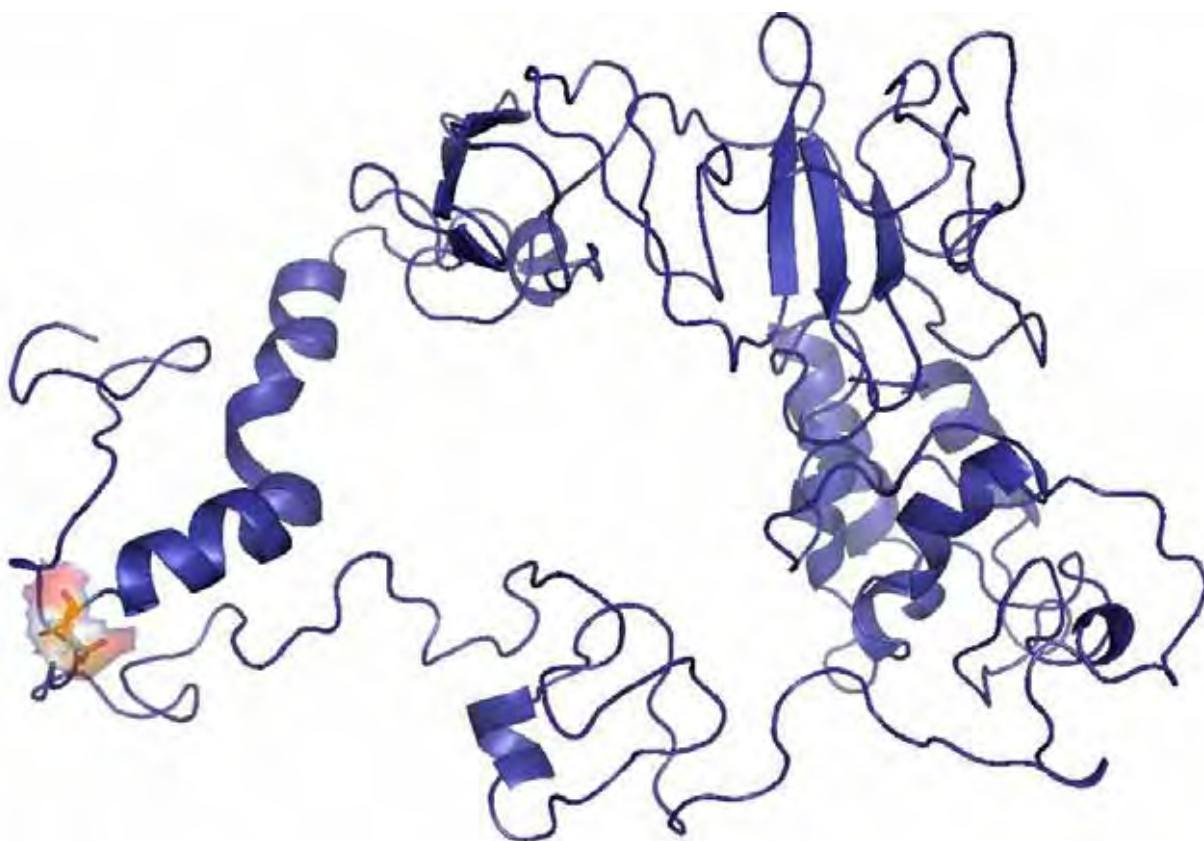
No resíduo 16 foi encontrada uma substituição de leucina para serina. Esta alteração causa uma alteração na hidrofobicidade desta região, uma vez que o primeiro é apolar e o segundo é polar. Essa alteração foi encontrada em uma sequência do genótipo 3 de um paciente respondedor.

Em dois pacientes respondedores ao final do tratamento foram encontrados quasispécies com deleção total da hélice que compõe a âncora de membrana ou CRS. A Figura 32 compara a alteração destes dois mutantes com a sequência referência do genótipo 3.

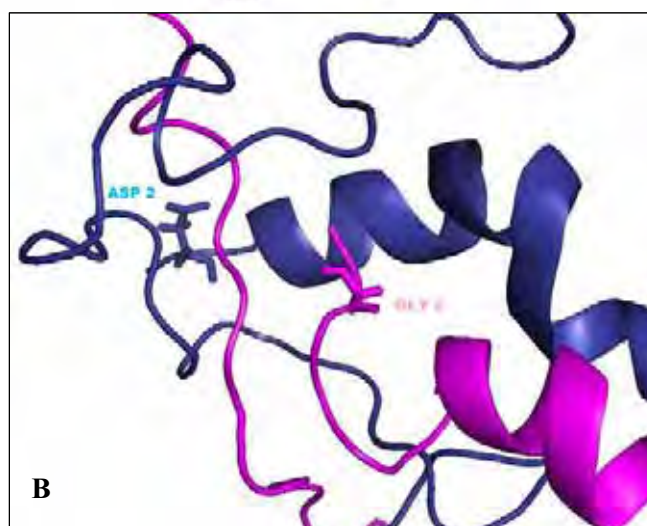
Em uma sequência do genótipo 3 extraída de um paciente respondedor, foi encontrada substituição na tríade leucina, ácido glutâmico e serina (resíduos 326 a 328). Os aminoácidos encontrados na sequência mutante foram triptofano, asparagina e arginina. Como podemos observar na Figura 33, essas substituições causam uma alteração de hidrofobicidade e carga neste local. A estrutura provavelmente não é afetada, uma vez que esta região possui estrutura secundária em *coil*.

A tríade valina, prolina e prolina nas posições 335 a 337 também é descrita como essencial na função da NS5A. Em uma sequência do genótipo 1 de um paciente respondedor ao final do tratamento, houve alteração nesta tríade para ácido glutâmico, prolina e isoleucina. Embora a isoleucina também seja apolar como a prolina, a substituição de valina (apolar) por ácido glutâmico (polar ácido) causa alteração na distribuição de cargas e hidrofobicidade da região (Figura 34).

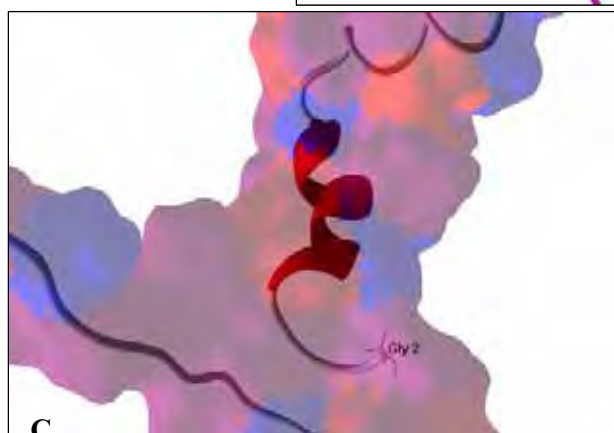
Por último, na posição 342 foi encontrada uma substituição de prolina por serina, o que causa uma alteração de hidrofobicidade (apolar para polar) desta região (Figura 35). Essa substituição foi encontrada em uma sequência do genótipo 1 de um paciente respondedor ao final do tratamento.



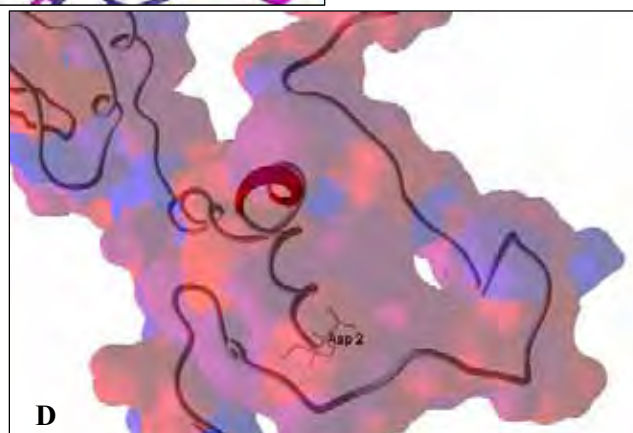
A



B



C



D

Figura 25. Esquemas da alteração na posição 02. A) Localização, B) Comparação da mutação em azul com a forma nativa da referência AF009606 em rosa, C) Padrão de hidrofobicidade na forma nativa e D) forma mutante (azul – hidrofóbico, vermelho – hidrofílico, roxo – pouco hidrofóbico).

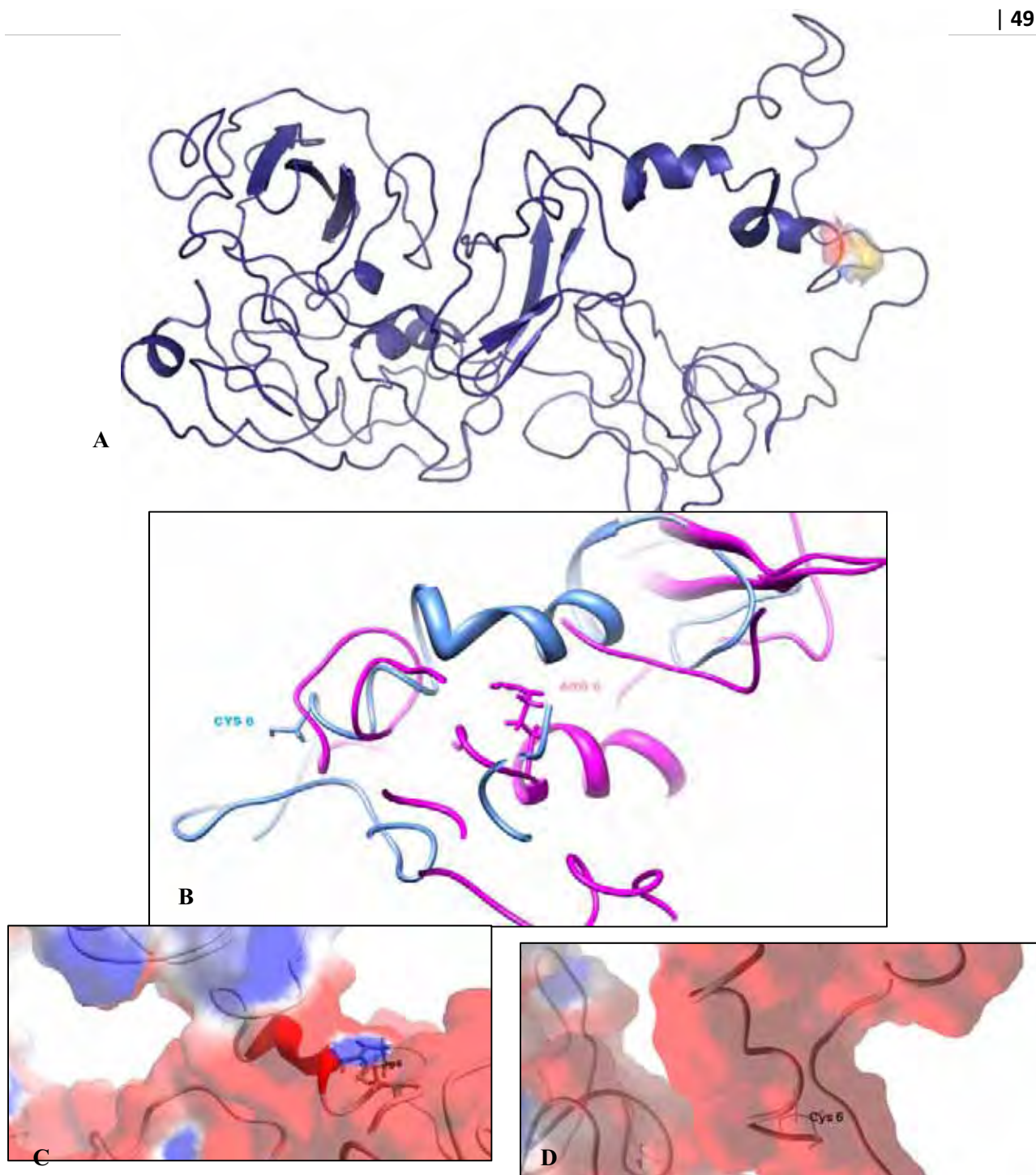


Figura 26. Esquemas da alteração na posição 06. A) Localização, B) Comparação da mutação em azul com a forma nativa da referência D17763 em rosa, C) Padrão de cargas na forma nativa e D) forma mutante (azul – básico, vermelho – ácido, branco – neutro).

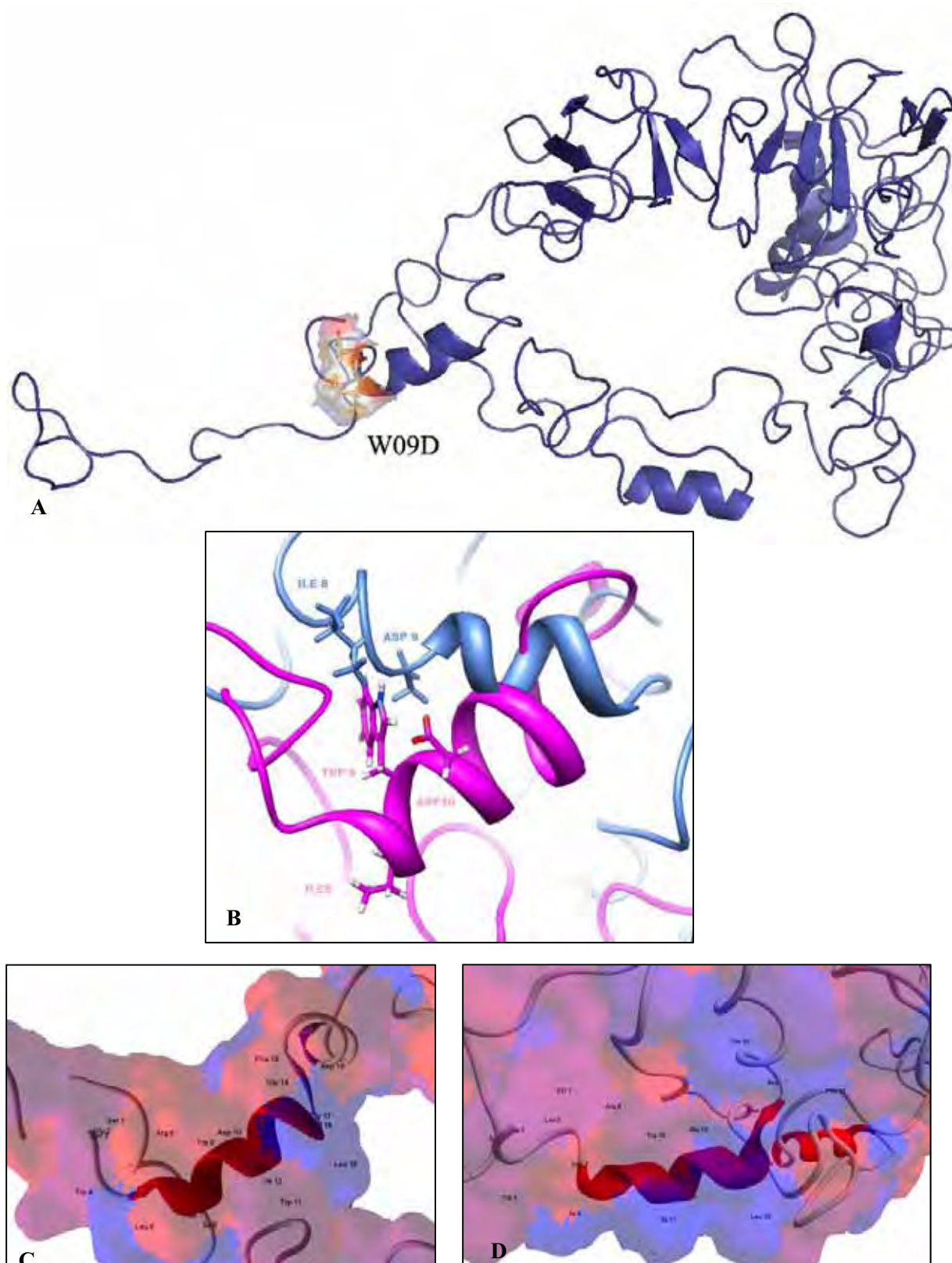


Figura 27. Esquemas da alteração na posição 09. A) Localização, B) Comparação da mutação em azul com a forma nativa da referência AF0069606 em rosa, C) Padrão de hidrofobicidade na forma nativa e D) forma mutante (azul – hidrofóbico, vermelho – hidrofílico, roxo – pouco hidrofóbico).

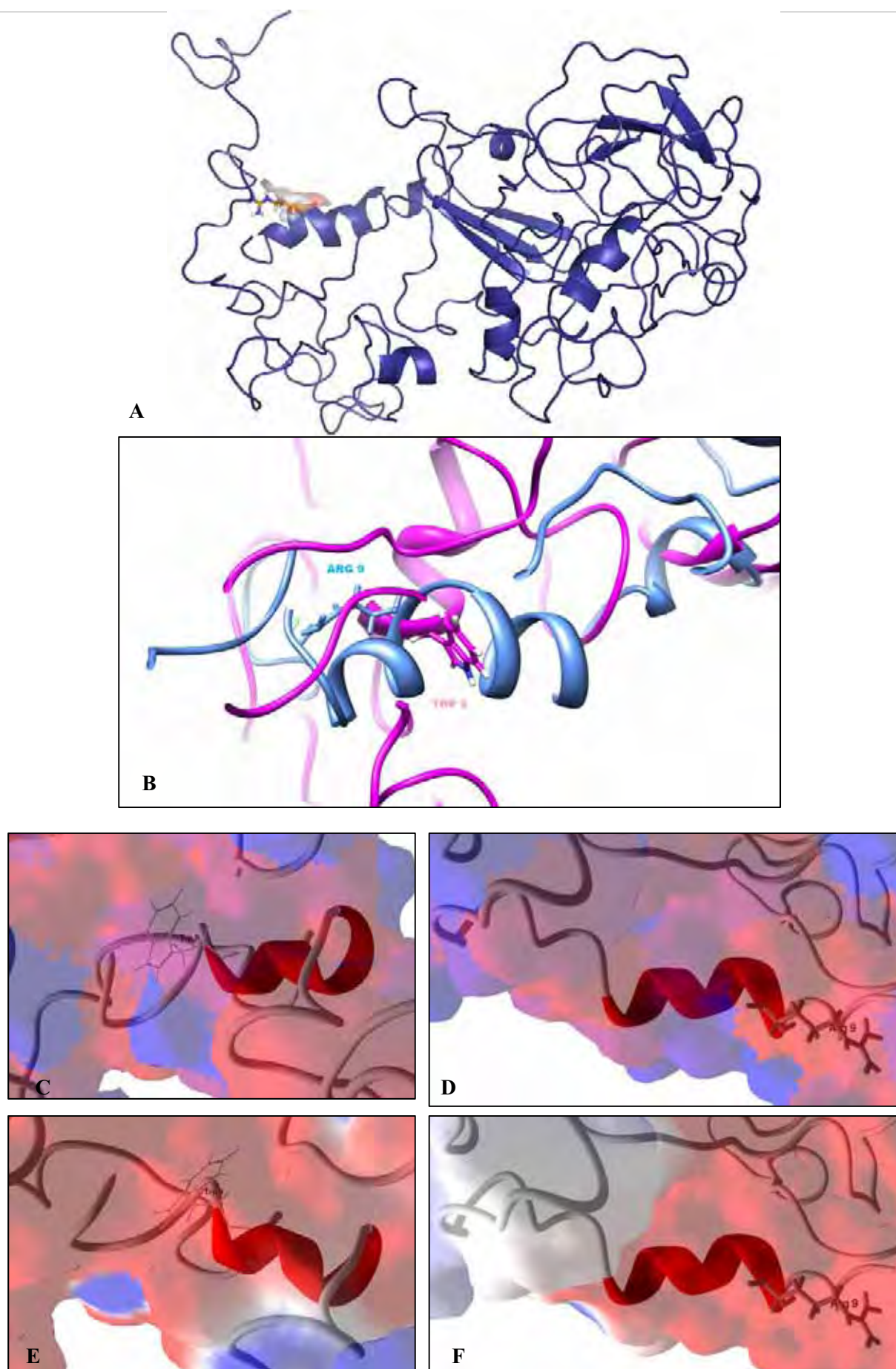


Figura 28. Esquemas da alteração na posição 09. A) Localização, B) Comparação da mutação em azul com a forma nativa da referência D17763 em rosa, C) Padrão de hidrofobicidade na forma nativa e D) forma mutante (azul – hidrofóbico, vermelho – hidrofílico, roxo – pouco hidrofóbico), E) Padrão de cargas na forma nativa e F) forma mutante (azul – básico, vermelho – ácido, branco – neutro).

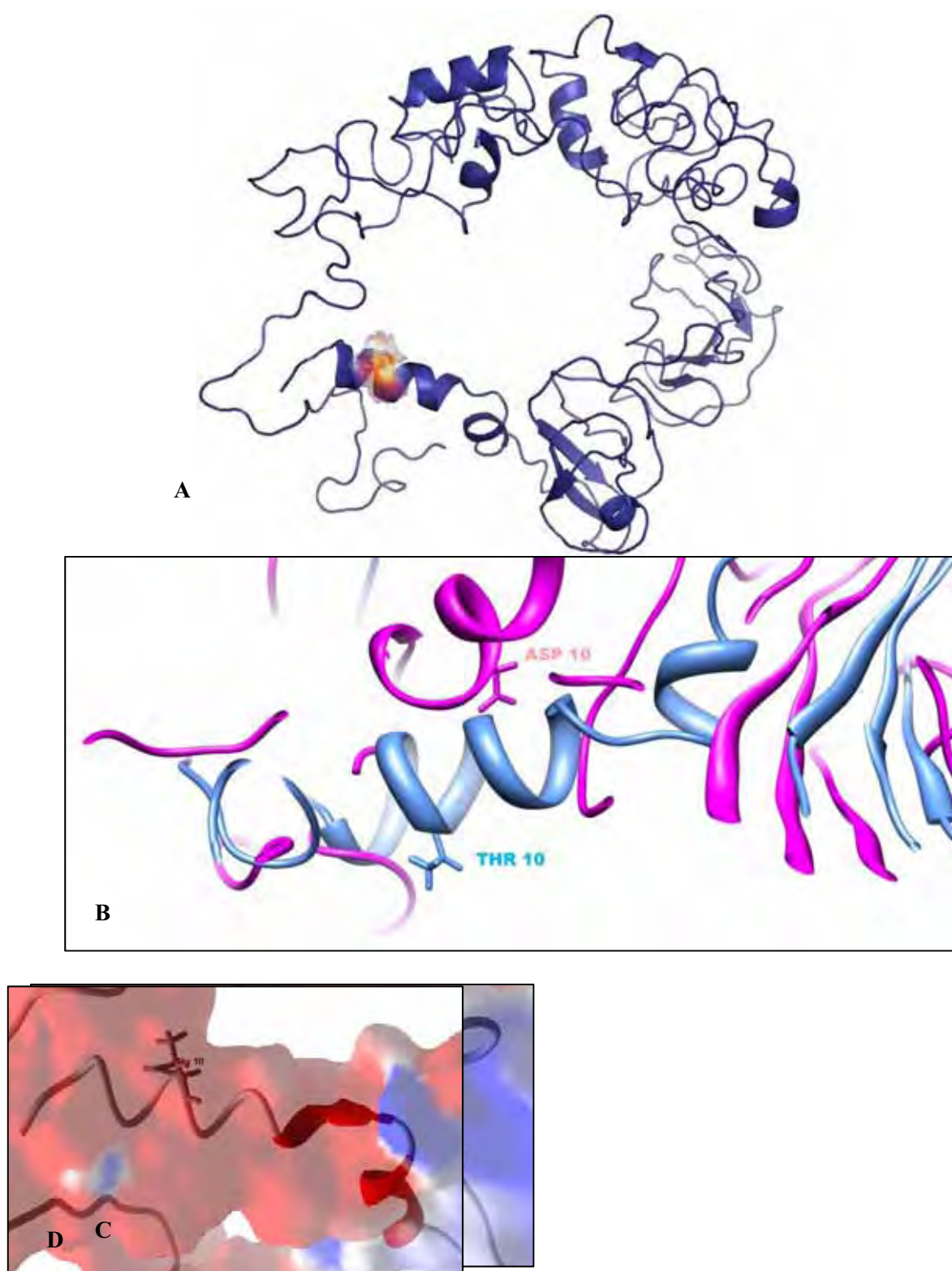


Figura 29. Esquemas da alteração na posição 10. A) Localização, B) Comparação da mutação em azul com a forma nativa da referência D17763 em rosa, C) Padrão de cargas na forma nativa e D) forma mutante (azul – básico, vermelho – ácido, branco – neutro).

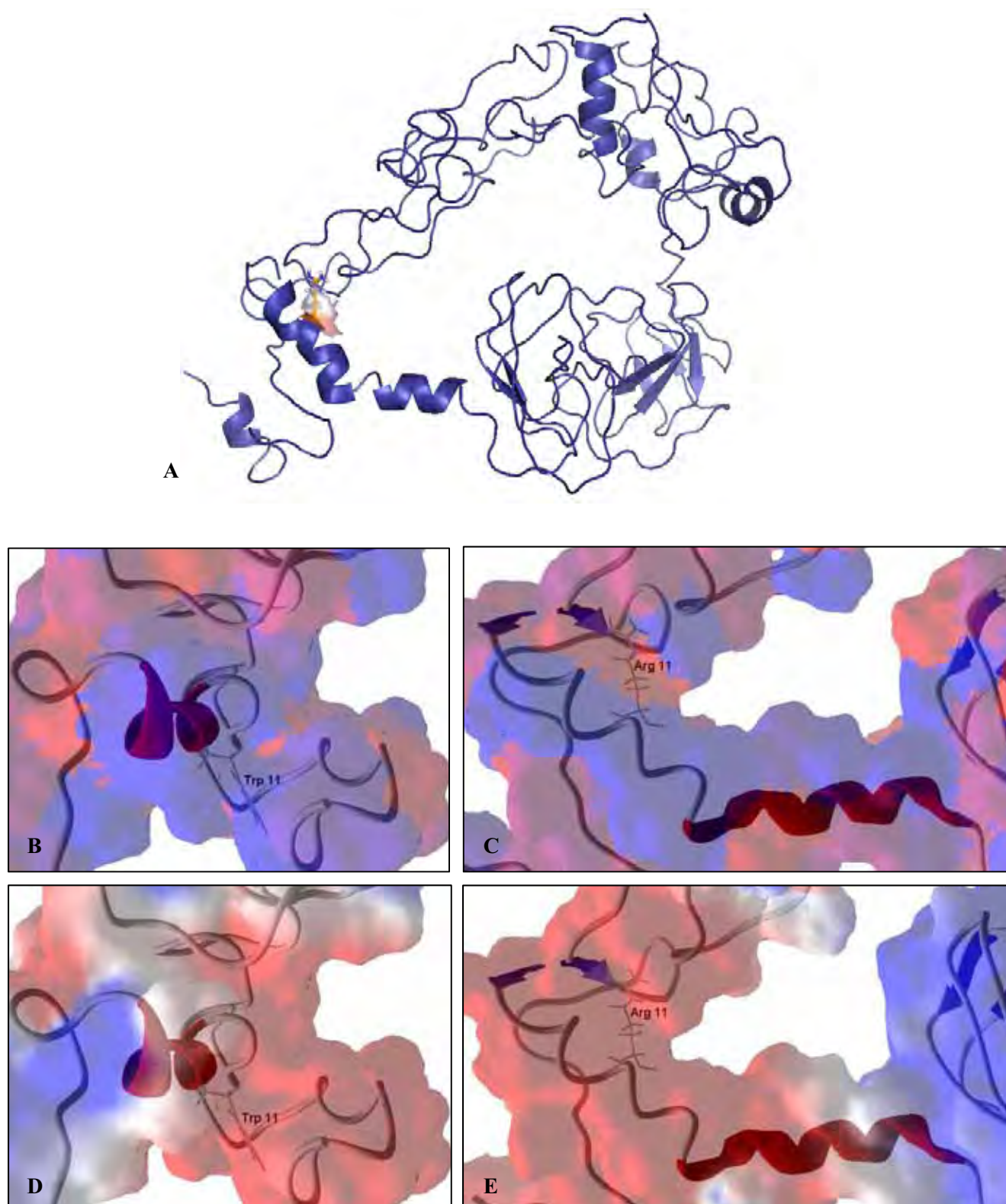


Figura 30. Esquemas da alteração na posição 11. A) Localização, B) Comparação da mutação em azul com a forma nativa da referência D17763 em rosa, C) Padrão de hidrofobicidade na forma nativa e D) forma mutante (azul – hidrofóbico, vermelho – hidrofílico, roxo – pouco hidrofóbico), E) Padrão de cargas na forma nativa e F) forma mutante (azul – básico, vermelho – ácido, branco – neutro).



Figura 31. Esquemas a forma nativa da referência D17763 em rosa com a forma mutante que apresenta deleção da CRS em azul. As áreas circuladas em vermelho correspondem ao início da cadeia polipeptídica da forma nativa (em rosa claro) e da forma mutante (em azul escuro). Observe que este aminoterminal mutante não apresenta hélice alfa.

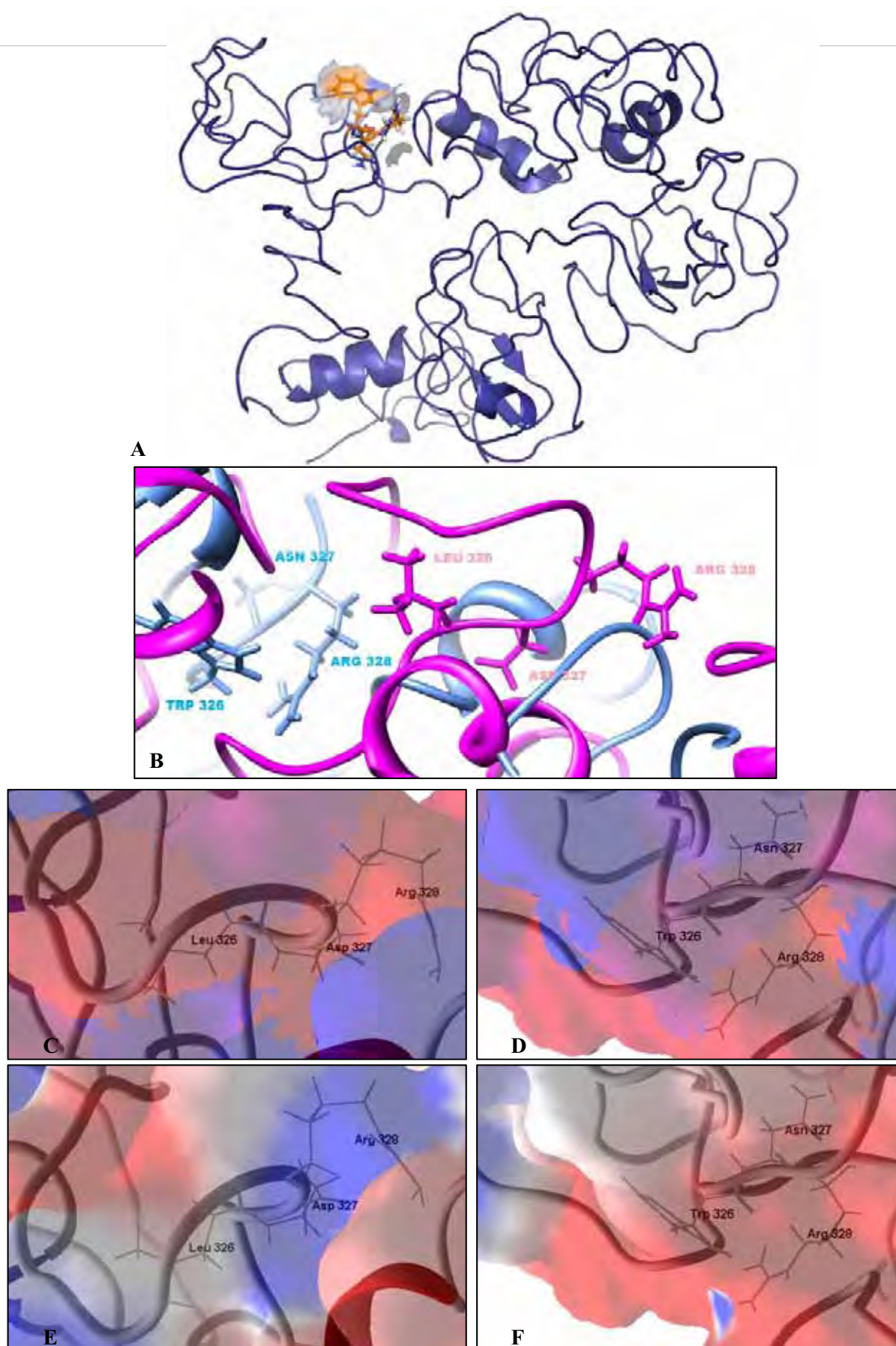


Figura 32. Esquemas da alteração na posição 326-328 A) Localização, B) Comparação da mutação em azul com a forma nativa da referência D17763 em rosa, C) Padrão de hidrofobicidade na forma nativa e D) forma mutante (azul – hidrofóbico, vermelho – hidrofílico, roxo – pouco hidrofóbico), E) Padrão de cargas na forma nativa e F) forma mutante (azul – básico, vermelho – ácido, branco – neutro).

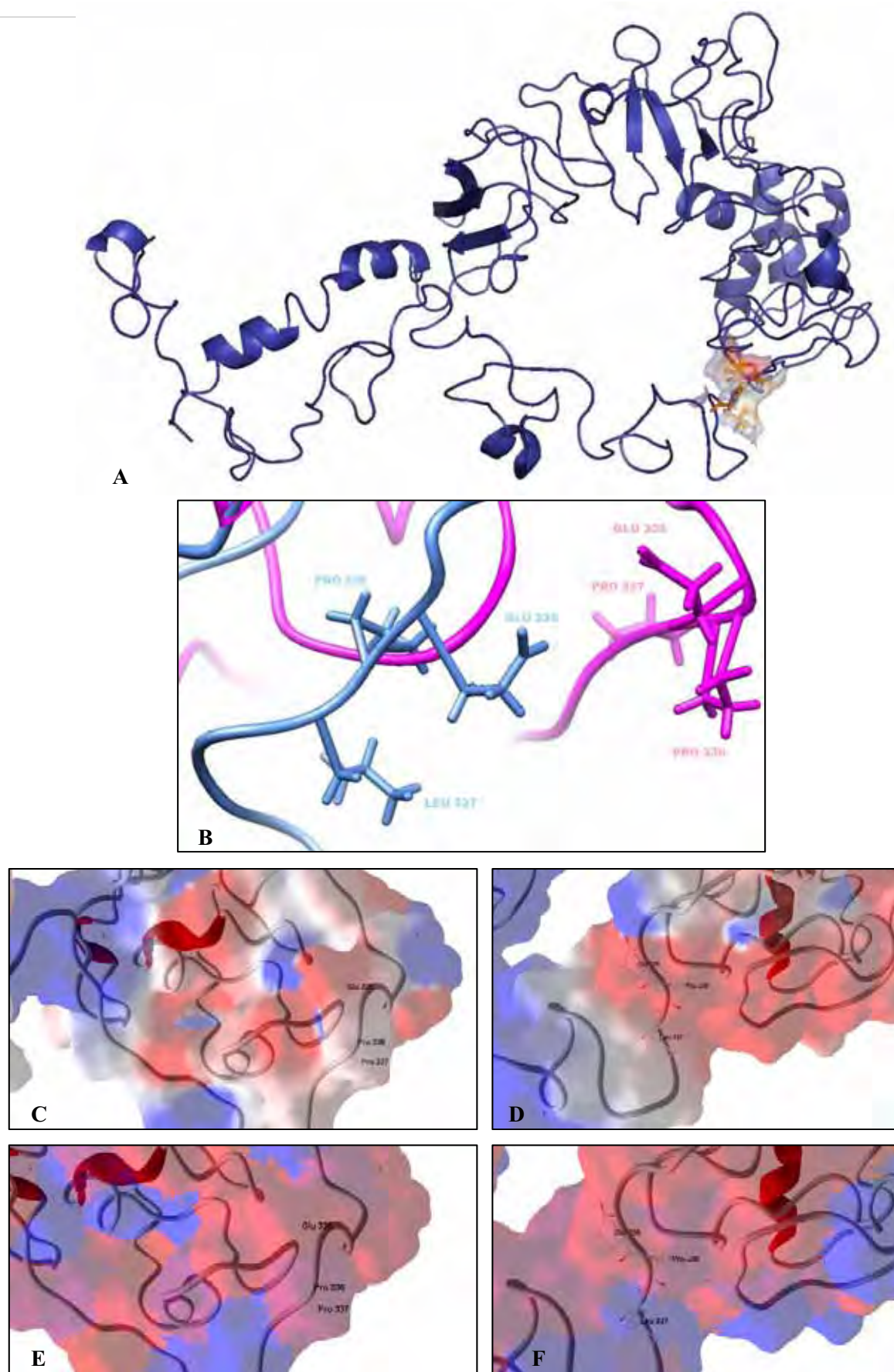


Figura 33. Esquemas da alteração na posição 335-337. A) Localização, B) Comparação da mutação em azul com a forma nativa da referência AF009606 em rosa, C) Padrão de hidrofobicidade na forma nativa e D) forma mutante (azul – hidrofóbico, vermelho – hidrofílico, branco – pouco hidrofóbico), E) Padrão de cargas na forma nativa e F) forma mutante (azul – básico, vermelho – ácido, branco – neutro).

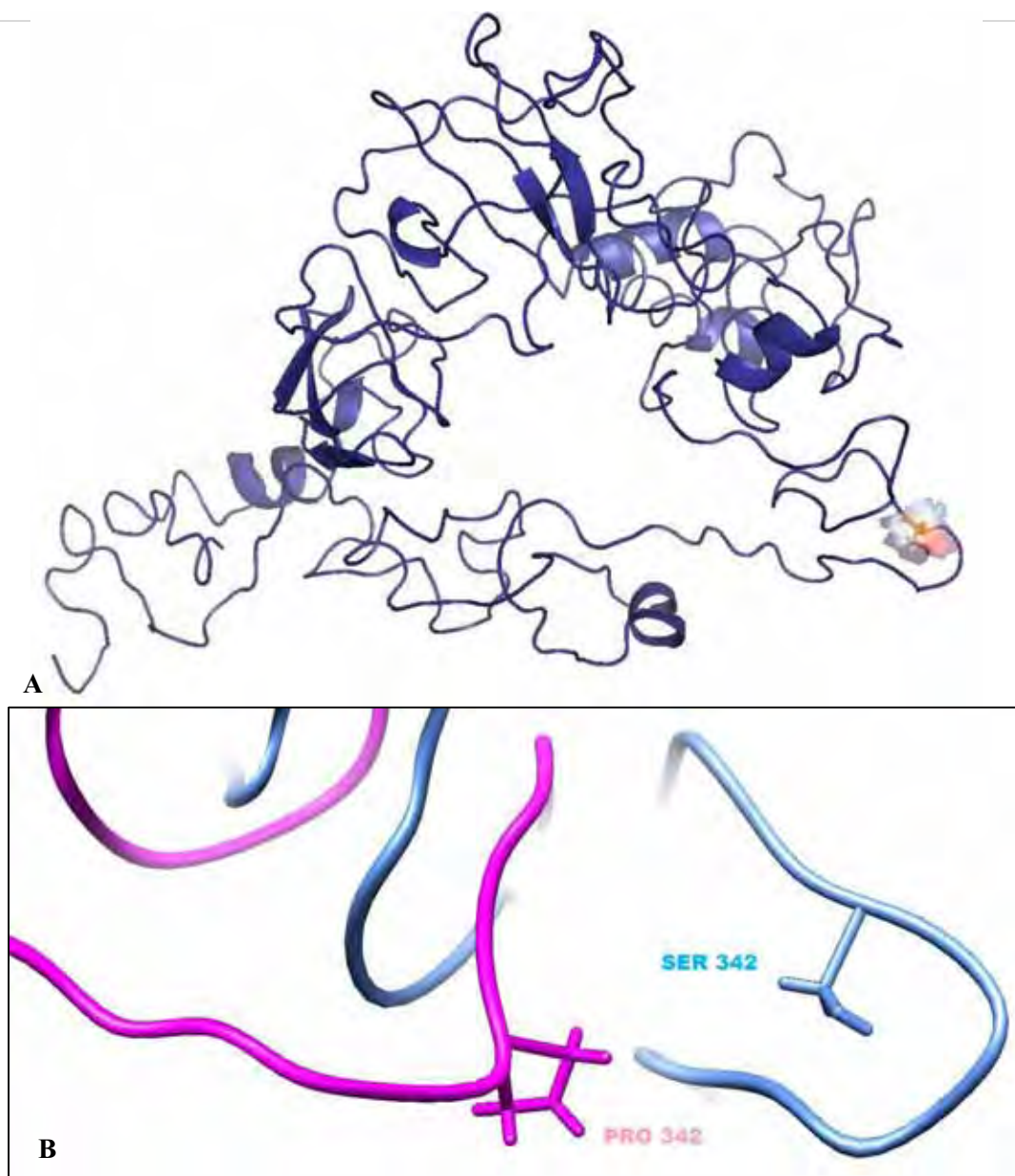


Figura 34. Esquemas da alteração na posição 342. A) Localização, B) Comparação da mutação em azul com a forma nativa da referência AF009606 em rosa, C) Padrão de hidrofobicidade na forma nativa e D) forma mutante (azul – hidrofóbico, vermelho – hidrofílico, roxo – pouco hidrofóbico), E) Padrão de cargas na forma nativa e F) forma mutante (azul – básico, vermelho – ácido, branco – neutro).

5. Discussão

É estimado que aproximadamente 170 milhões de pessoas estejam infectadas com o vírus da Hepatite C. A infecção pode causar uma doença hepática crônica e progressiva. Atualmente, a terapia mais utilizada para da Hepatite C é baseada na administração de Peginterferon combinado à Ribavirina (MORENO et al., 2006; KLENERMAN et al., 2009). No entanto, esse tratamento ainda não é totalmente eficiente: o genótipo 1 apresenta cerca de 45% a 55% de resposta virológica sustentada, enquanto que o genótipo 3 apresenta de 75% a 80%. As bases biológicas para essa diferença ainda não são conhecidas, pois os genomas destes genótipos são tão variáveis que esta diferença pode ser resultado de alterações múltiplas e complexas (KLENERMAN et al., 2009). Além disso, interações tanto fatores virais como os do hospedeiro parecem ser importantes para a resposta virológica (HOOFNAGLE et al., 2009). Os fatores virais mais estudados em relação à resposta a terapia são o genótipo, a carga viral e a composição de *quasispecies* (MORENO-OTERO, 2005). Embora os genótipos 1a e 1b sejam preditivos de uma baixa resposta ao tratamento, nenhuma diferença significativa nas taxas de resposta virológica entre os subtipos do HCV foi reportada (WOHNSLAND et al., 2007). Além do genótipo, a alta taxa de variação do genoma de RNA pode estar relacionada à persistência da infecção viral. A alta variabilidade pode ser uma forma importante de escape do vírus frente ao sistema imune do hospedeiro e ao tratamento utilizado através do mecanismo de seleção das variantes resistentes. Essa característica torna o desenvolvimento de vacinas e novas drogas eficientes para a Hepatite C em um grande desafio (MORADPOUR et al., 2007; HANOULLE et al., 2009).

Métodos experimentais como a difração de raios X em cristais ou Ressonância Magnética Nuclear (RMN), possuem um papel indispensável na correta compreensão dos mecanismos bioquímicos das proteínas, na elaboração de modelos experimentais, na construção de relações filogenéticas, no desenho de fármacos, no desenvolvimento de vacinas, dentre vários outros. Entretanto, muitas proteínas não possuem sua estrutura tridimensional determinada, como

acontece com a proteína NS5A do HCV. Por isso as ferramentas computacionais têm grande utilidade na predição de estruturas tridimensionais de proteínas que ainda não foram resolvidas por métodos experimentais (Marti-Renom et al., 2000).

No presente estudo por ferramentas computacionais, não foram identificados regiões de variação de aminoácidos ou de estrutura que possam estar relacionados a resposta do paciente ao tratamento. Entretanto, tanto a composição de diversos aminoácidos, como a composição de estrutura secundária (Figuras 3-4) apresentaram uma diferença quanto ao genótipo. Na literatura, não foram encontrados dados sobre esses dois tipos de composição protéica, entretanto, sabe-se que o genótipo 1 é associado à manifestações mais severas da Hepatite C e menor taxa de resposta ao tratamento (WOHNSLAND et al., 2007). Como a diferença encontrada foram todos em aminoácidos localizados na superfície da proteína e na quantidade de *coil*, podemos sugerir que a NS5A apresenta diferenças na interação desta com outras proteínas quanto ao genótipo. Desta forma, esta diferença poderia refletir na resistência viral, na replicação e na interação da NS5A com outras proteínas virais e do hospedeiro que se apresenta diferente nos genótipos.

A partir do *template* quimérico montado foi possível elaborar modelos teóricos da proteína NS5A dos genótipos 1 e 3 dentro dos parâmetros estereoquímicos e geométricos permitidos (Figura 5, Tabelas 1 e 2), a partir das sequências referência do *Genbank*. Essas estruturas geradas por modelagem molecular foram utilizadas como ponto de partida para o estudo dinâmico de proteínas utilizando-se a técnica de dinâmica molecular (DM), a qual nos permite acompanhar a trajetória de cada um dos átomos de uma proteína ao longo do tempo. Desta forma, é possível observar as mudanças que ocorrem na estrutura da macromolécula. Assim, ambas as técnicas (MM e DM) foram utilizadas nesse trabalho para que fossem obtidas as estruturas tridimensionais da NS5A completa do vírus da Hepatite C. Durante a dinâmica, os modelos gerados por MM estabilizaram a sua estrutura por volta dos 700ps (Figura 7) que pode ser atribuído às alterações ambientais entre o experimental (sólido) e o de simulação (líquido). Após esta fase, podemos notar pequenas flutuações na conformação decorrentes de acomodações dos átomos. Além disso, pelo

gráfico pela comparação das estruturas antes e após a DM (Figuras 8-10), é possível observar que houve uma compactação da estrutura, ou seja, dentre as estruturas geradas por predição da NS5A a forma mais estável é mais compacta em relação à estrutura gerada por MM. Desta forma, a partir deste modelo gerado por MM e DM será possível realizar outros estudos sobre a estrutura/função da NS5A, como por exemplo para o estudo de inibidores.

A análise pelo PROSITE mostrou que a NS5A possui diversos potenciais sítios de modificações co- e pós-traducionais, altamente conservados nos genótipos 1 e 3, todos localizados na superfície da proteína. Embora a localização dos sítios de N-glicosilação tenha sido variável, todas as sequências apresentaram pelo menos dois sítios (Figura 12). A ligação de carboidratos à proteína pode exercer diversas funções, uma delas é de promover epítomos adicionais de reconhecimento para proteínas receptoras (DRICKAMER, TAYLOR, 1993; PETRESCU et al., 2004). Estes eventos de reconhecimento estão envolvidos em diversos processos celulares, como o tráfego de proteínas, início inflamatório e defesa do hospedeiro (CROCKER, 2002). Não há estudos que descrevam glicosilação na proteína NS5A, mas este tipo de modificação é comumente encontrado nas famílias protéicas. Em vírus, as proteínas glicosiladas são essenciais nos processos de reconhecimento e adesão celular das proteínas do envelope ou capsídeo, inclusive em outros *Flavivirus* como o vírus da Dengue (TAN et al., 2007), o vírus da Encefalite Japonesa (KIM et al., 2008). Além de atuar na superfície celular, estudos com o capsídeo do vírus da Hepatite E demonstrou que mutações no sítio de glicosilação podem inibir a montagem da partícula viral, pois a sua ausência poderia alterar a estrutura da proteína (GRAFF et al., 2007). Helenius (1994) demonstrou que a glicosilação promovia um aumento na solubilidade e possivelmente na interação com chaperonas no Retículo Endoplasmático, portanto atuavam no enovelamento e estabilização da forma funcional protéica. Deste modo, estas proteínas poderiam se agregar irreversivelmente ou saírem do RE sem a associação a oligômeros se a glicosilação fosse inibida (HELENIUS, 1994). Seguindo este mecanismo, a glicosilação da proteína NS5A pode ser essencial para manter a sua estrutura funcional, uma vez que a presença de sítios desta modificação foram presentes em

todas as sequências. Vários potenciais sítios de fosforilação também foram identificados ao longo da proteína NS5A (Figuras 13 a 16). Assim como os sítios de N-glicosilação, estes foram presentes em todas as sequências, embora variassem quanto ao número e localização. A fosforilação e hiperfosforilação da NS5A são descritas como essenciais para a montagem do vírus. A ausência da fosforilação pode inibir a interação com a proteína do *core*, processo que é essencial para a montagem da partícula viral (MASAKI, et al., 2008). A fosforilação é um processo de modificação pós-traducional reversível, sendo assim um dos fatores que permitem a multifuncionalidade de proteínas como a NS5A. Diversas proteínas celulares foram identificadas no processo de fosforilação da NS5A e vários estudos sugerem que esta modificação pode atuar na regulação da replicação viral e na modulação das vias de sinalização intracelular do hospedeiro (HUANG et al., 2007). As proteínas quinases celulares com interação identificadas com a NS5A foram a AKT, p70s6K, MEK, MKK1, CKI, CKII e Syk (HUANG et al., 2007; TELLINGHUISEN et al., 2008; INUBUSHI et al., 2008). Não foram encontrados estudos da fosforilação por PKC ou cAMP, entretanto como o processo de fosforilação na NS5A ainda não foi totalmente elucidado, a PKC e o cAMP seriam também candidatos neste processo, uma vez possíveis sítios de interação com estas substâncias foram encontrados em quase todas as sequências de ambos os genótipos.

Além de vários sítios de N-glicosilação e fosforilação, foram identificados diversos possíveis sítios de N-miristoilação que também mostrou-se qualitativamente conservada, embora o número e a localização fossem variáveis mesmo em sequências provenientes do mesmo paciente (Figura 17). A ligação covalente de miristato é um processo geralmente irreversível e altera a hidrofobicidade da proteína. A miristoilação em proteínas de diversos vírus está relacionada a inúmeras funções como a localização celular da proteína e a interação proteína-proteína (GORDON et al., 1991; FARAZI et al., 2001). O processo de miristoilação não é descrito para a proteína NS5A do vírus da Hepatite C. Entretanto, como todas as *quasispecies* analisadas apresentaram estes sítios, se dados experimentais comprovarem a existência desta modificação na

NS5A, pode-se sugerir que a miristoilação é um importante processo na estabilização da estrutura e função da NS5A.

Deste modo, se dados experimentais corroborarem com os resultados de predição de sítios de modificação co- e pós-traducionais, a inibição da glicosilação, da fosforilação e/ou de miristoilação da NS5A seriam possíveis mecanismos de atuação para desenhos de novos antivirais.

Outro sítio interessante identificado pelo PROSITE foi o sítio de ligação celular (RGD), que se mostrou altamente conservada nos genótipos 1 e 3, tanto quanto ao número quanto à localização. Todas as sequências apresentaram esse *motivo* no resíduo 48-50, apenas (Figura 18). Esta região está dentro da região transmembrânica que foi demonstrada pela predição do programa MEMSAT3 (Figura 11). A sequência RGD (Arginina, Glicina, Ácido Aspártico) é uma sequência de ligação a integrinas, proteínas de superfície celular que atuam na ligação célula-célula e célula-matrix extracelular (RUOSLAHTI, PIERSCHBACHER, 1986). Além disso, essa sequência foi identificada como essencial no processo de dobramento correto no Retículo Endoplasmático e transporte intracelular para glândulas secretórias da proproteína convertase 1, mostrando que a sequência RGD também pode possuir função intracelular (ROVÈRE et al., 1999).

Estudos com micelas de peptídeos cíclicos RGD transfectados em célula HeLa, mostraram que estes se concentraram na região perinuclear através de microscopia confocal a laser (OBA et al., 2008). Além disso, experimentos com a proteína L1-CAM humana, uma proteína superexpressada em diversos carcinomas e regula a expressão de vários genes, mostraram que mutantes com a sequência RGE infectados em célula HEK293 não translocaram para o núcleo como ocorre com a forma nativa RGD, evidenciando a importância deste sítio na sinalização nuclear de L1 e sugerindo que esta é dependente de integrinas (GAST et al., 2008).

Deste modo, a sequência RGD encontrada nas sequências da NS5A dos genótipos 1 e 3 poderiam ter papel em (1) no dobramento e transporte intracelular desta e/ou (2) na localização

nuclear e perinuclear. A NS5A possui um sinal de localização nuclear (NLS) funcional no seu C-terminal (IDE et al., 1996). Entretanto, a forma completa da proteína é encontrada predominantemente no citoplasma (MANABE et al., 1994; IDE et al., 1996; KIM et al., 1999) e/ou na região perinuclear (TANJI et al., 1995). Apenas formas contendo deleção no N-terminal da NS5A (resíduos 1-38) apresentavam localização nuclear predominante (SONG et al., 2000). As formas com deleção no N-terminal ocorrem naturalmente durante a infecção, pela ação de caspases (HIDAJAT et al., 2004; KALAMVOKI et al., 2006). A função das formas nucleares da NS5A ainda não é esclarecida, mas estas podem ser potenciais reguladores de transcrição (TANIMOTO et al., 1997; FUKUMA et al., 1998; SAUTER et al., 2009). Além disso, um estudo recente demonstrou que a proteína c-Raf, uma proteína de transdução de sinal celular que interage com o C-terminal da NS5A, é detectada no núcleo em células infectadas com formas da NS5A sem o N-terminal (resíduos 1 a 31). Além disso, nesta situação a replicação do HCV é comprometida, mostrando que a c-Raf pode ser essencial no complexo replicativo que se localiza no citoplasma (SAUTER et al, 2009).

Deste modo, podemos sugerir que os fatores que possivelmente participam na localização nuclear da NS5A são: (1) a deleção da região N-terminal que atuaria como “mascarador” da região NLS; (2) presença da NLS intacta; (3) a possível interação da sequência RGD com integrinas que direcionariam a NS5A para a região perinuclear; (4) a possível região transmembrânica (resíduos 32-52) que atuaria no transporte através da membrana nuclear. Como foi demonstrado que as formas nucleares inibem a replicação viral, também podemos sugerir que a indução da produção de formas nucleares da NS5A pode ser um potencial mecanismo de ação para novos antivirais. Entretanto, como demonstrado pela proteína c-Raf, outras proteínas que interagem com a NS5A e importantes em diversas vias celulares poderiam ser “carregadas” para o núcleo e por isso mais estudos sobre os efeitos destes mutantes são necessários.

A predição de função do *ProtFun* demonstrou que há provavelmente variação funcional da NS5A de acordo com o domínio. O domínio I poderia atuar como enzima, enquanto que o domínio III não teria função enzimática. Além de possuir uma região de ligação à membrana no N-terminal, o domínio I possui uma região de ligação ao zinco e possivelmente pode interagir com RNA (LOVE et al., 2009; TELLINGHUISEN et al., 2005), entretanto nenhuma função enzimática foi reportada até o momento. O domínio III é uma região altamente flexível que pode interagir com diversas proteínas, mas assim como os resultados deste estudo, nenhuma função enzimática foi identificada.

A predição funcional da NS5A também variou de acordo com os domínios. A sequência completa da NS5A foi classificada principalmente na categoria funcional de metabolismo central intermediário (Figura 19), enquanto que a ontologia do gene fora classificada em regulação da transcrição. Entretanto, quando analisamos os três domínios da proteína separadamente, nos três domínios predomina a predição da categoria de GO de fator de crescimento (Figuras 20, 23 e 25). Embora esses dados pareçam conflitantes, há estudos que indicam que de fato a NS5A pode ser um regulador na transcrição no caso das formas nucleares como citado acima e que ela também se associa com receptores de fatores de crescimentos, como a proteína Grb2, podendo interferir em diversos processos celulares como proliferação, diferenciação, metabolismo, apoptose, dentre outros (TAN et al., 1999). Deste modo, podemos inferir que estes dados não são conflitantes, mas sim adicionais. Além disso, outras categorias como hormônio (Figura 25) e transcrição (Figura 21), também podem ser candidatos funcionais da NS5A.

Na predição de categoria funcional, as sequências do domínio I apresentaram em metabolismo energético, o que condiz com a função relacionada a fator de crescimento. O domínio III apresentou 99% das sequências na categoria de transporte e ligação (Figura 24). Embora não seja muito clara, podemos inferir que esta função deve-se a presença da NLS no domínio três, além de sítios de interação com proteínas de diversas vias celulares, como a c-Raf

descrita anteriormente. Apenas o domínio II apresentou sequências com diferentes predições em tradução (60%) e transporte e ligação (30%) (Figura 23). Embora não fora encontrada estudos de interação com proteínas diretamente relacionadas à tradução, sabe-se que assim como o domínio III, o domínio II pode interagir com diversas proteínas que participam de inúmeras vias celulares como diversas quinases, proteínas que possuem domínios SH3, dentre várias outras (LIANG et al., 2007). Deste modo, embora novas funções da NS5A não tenham sido encontradas, a análise do *ProtFun* reforça a teoria de multifuncionalidade da NS5A, mesmo dentro de um mesmo domínio.

O presente estudo visou analisar estruturalmente e funcionalmente a NS5A do vírus da Hepatite C genótipos 1 e 3, comparando tanto quanto ao genótipo quanto à resposta ao tratamento observada nos pacientes. A partir das ferramentas de bioinformática, foi possível a elaboração de um modelo da estrutura protéica completa da NS5A, o que não consta na literatura, uma vez que experimentalmente apenas o domínio I teve sua estrutura determinada. Além disso, esta estrutura passou por vários programas de avaliação que mostraram que suas características estereoquímicas e geométricas estão de acordo com o permitido nos parâmetros já observados.

Não foi possível estabelecer uma possível relação entre a resposta ao tratamento e a estrutura da NS5A. Entretanto, como inúmeros estudos apontam, vários fatores virais e dos hospedeiros seriam em conjunto responsáveis pela persistência viral. Desta forma, estudos considerando os resultados da NS5A viral obtidos neste trabalho juntamente com outros fatores do hospedeiro, poderiam ser mais elucidativos quanto à predição da resposta ao tratamento com Interferon por meio da análise da NS5A.

Além disso, os resultados de composição de aminoácidos e estrutura secundária indicam que a NS5A tem diferença estrutural e talvez funcional quanto ao genótipo, uma vez que estas diferenças possam afetar nas interações da NS5A com outras proteínas virais e do hospedeiro. Embora os efeitos das diferenças nas sequências da proteína entre os genótipos ainda não sejam claros, sabe-se que a infecção por alguns genótipos são preditivos de não-resposta ao tratamento

(genótipo 1), enquanto que em outros a taxa de resposta virológica sustentada ultrapassa 90%. Desta forma, a resistência viral observada em alguns genótipos pode estar ligada a estas diferenças na composição estrutura primária e secundária da NS5A.

Quanto às análises funcionais, pode-se observar diversos potenciais sítios de modificações co- e pós-traducionais e um sítio de ligação celular na proteína NS5A que não são descritos na literatura. Ainda que dados experimentais comprovando a existência destes sítios sejam necessárias, a alta conservação observada é um indicativo de que estes sejam essenciais para a NS5A. A análise do *ProtFun* também demonstrou que a NS5A pode estar ligada a outras funções não descritas na literatura.

Além da predição de possíveis funções, este estudo também incluiu a análise de mutações na NS5A que foram descritas *in vitro* como comprometedoras da função e que em nossa população de estudo foram encontrados *in vivo*. Essas alterações foram descritas nos estudos com a CRS por Penin e colaboradores (2004) e com os domínios II e III por Tellinghuisen e colaboradores (2008).

As mutações na CRS foram encontradas em ambos os genótipos. No genótipo 1, houve uma mudança de polaridade e carga na posição 2 e uma deleção de triptofano na posição 9 (Tabela 6). Pela Figura 26, pode-se observar que além desta mudança bioquímica, a alfa-hélice que compõe a âncora de membrana também pode ser desestabilizada por esta alteração no resíduo 2. Estudos com mutantes que desestabilizavam a estrutura da âncora de membrana, mostraram que poderia ocorrer atenuação da interação com a membrana e que mesmo que a alteração ainda mantivesse os requisitos para uma âncora de membrana, a replicação da NS5A poderia ser afetada pelo comprometimento de alguma outra função da CRS ainda desconhecida (PENIN et al., 2004). No genótipo 3, também foram encontradas alterações de hidrofobicidade e carga de diversas regiões da CRS (Figuras 27, 29 a 31) e uma deleção da CRS completa em dois pacientes respondedores ao final do tratamento (Figura 32). Como discutido anteriormente, formas com

deleção no aminoterminal da NS5A tem sido descritas como formas nucleares da proteína. Embora a função destas formas nucleares ainda não sejam bem esclarecidas, um estudo recente determinou que estas podem ser geradas por ações de caspases celulares talvez como um processo de defesa, uma vez que a replicação é inibida *in vitro* em células infectadas com essas formas da NS5A (SAUTER et al., 2009).

Nos domínios II e III, foram encontradas alterações consideradas comprometedoras da função da NS5A em estudos de mutação sítio dirigida (TELLIGHUISEN et al., 2008). Foi encontrada uma alteração no genótipo 3 na tríade dos resíduos 326 a 328 que alterava a hidrofobicidade e a carga da região de *coil* (Figura 33) e duas alterações no genótipo 1 (Figuras 34 e 35), uma na tríade 335 a 337 e outra na Prolina 342. Embora ainda haja poucos estudos sobre a função dos domínios II e III, sabe-se que estes são essenciais na replicação e montagem do vírus (HUGHES et al, 2009), além de interagirem com diversas proteínas virais e celulares. Nas figuras estruturais é possível ainda observar que todas estas alterações foram em locais de superfície da proteína e que portanto, tais alterações seriam mais prováveis de estarem alterando regiões de interação essencial à replicação do que à estabilidade da proteína. Em adição, essas características (acessibilidade estrutural e conservação) tornam estas regiões interessantes alvos para inibidores da NS5A.

Somando a estes resultados, devemos ressaltar que não há estudos que analisem a estrutura e a função da NS5A a partir de um número de sequências tão abrangente quanto este estudo que englobou 345 sequências da proteína completa além de 9 sequências referências retiradas do *Genbank*. Como o HCV é um vírus com alta variabilidade, estudos com pequeno número de sequências podem gerar resultados não confiáveis, uma vez que dados seriam perdidos em variantes dos vírus que não fossem analisadas.

A partir dos resultados deste estudo, os modelos gerados e as análises realizadas poderão ser utilizados para investigação funcional da NS5A e interação desta com outras proteínas

celulares e virais por técnicas empíricas e para o desenho de antivirais específicos para a NS5A, uma vez que esta proteína é essencial para a replicação viral e consequentemente para a infecção do HCV.

6. Conclusão

De acordo com os objetivos traçados neste estudo, podem concluir que:

- Este trabalho traz importante contribuição na compreensão das características da proteína NS5A do HCV dos genótipos 1 e 3;
- A análise por bioinformática da estrutura primária e secundária mostrou que há diferenças entre os genótipos 1 e 3 da NS5A;
- A modelagem molecular comparativa, o *threading* e a dinâmica molecular são poderosas ferramentas na predição de estrutura tridimensional de proteínas de difícil análise experimental como a NS5A, gerando modelos com confiança para a aplicação de diversas análises como a comparação de estruturas mutantes;
- A procura de padrões e *motivos* pelo PROSITE mostrou que a NS5A pode ter modificações pós-traducionais ainda não descritas na literatura como glicosilação e miristoilação, além de um sítio de interação com integrinas;
- A predição de função mostra que pode haver diferenças na função desempenhadas por cada domínio da NS5A;
- Mutações na CRS são comprometedoras da função principalmente por comprometerem a função de âncora de membrana;

- Mutações deletérias do domínio II e III podem ser encontradas *in vivo*, e embora a função destas regiões ainda não sejam certas, é possível que esta desempenhe um papel crucial em interações com outras proteínas virais e/ou celulares.
- Não foram encontradas neste estudo relação entre as características da NS5A em amostras pré-tratamento e a resposta do paciente à terapia;

7. Perspectivas

Considerando os resultados obtidos no trabalho e as informações obtidas com relação à estrutura e função da NS5A, ainda serão realizados:

- Clonagem, expressão e purificação da NS5A para caracterização estrutural da NS5A por meio de estudos experimentais como cristalografia, RMN, SAXS e Espectroscopias;
- Comparação com os dados experimentais e obtidos por estudos em bioinformática;
- Caracterização funcional pelo estudo da NS5A em células infectadas com o vírus da HCV.

8. Referências Bibliográficas

ALLEN, F.H. The Cambridge Structural Database: a quarter of a million crystal structures and rising. **Acta Cryst**, v.B58, p.380-388, 2002.

ALTER, H.J. et al. Detection of antibody to hepatitis C virus in prospectively followed transfusion recipients with acute and chronic non-A, non-B hepatitis. **N Engl J Med**, v.321, n.22, p.1494-1500, 1989.

APPEL, N. et al. Essential role of domain III of nonstructural protein 5A for hepatitis C virus infectious particle assembly. **PLoS Pathog**, v.4, n.3, p.e1000035, 2008.

BAIROCH, A.; BUCHER, P.; HOFMANN, K. The PROSITE database, its status in 1997. **Nucleic Acids Research**, v.25, p.217-221, 1997.

BAYAT, A. Science, medicine, and the future: Bioinformatics. **BMJ**, v. 324, p. 1018-1022, 2002.

BERMAN, H. M. et al. The Protein Data Bank, **Nucleic Acids Research**, v. 28, p. 235-242, 2000.

BRASS, V. et al. An aminoterminal amphipathic alpha-helix mediates membrane association of the hepatitis C virus nonstructural protein 5A. **J Biol Chem**, V.227, P.8130-8139, 2002.

BROWN, R.S.Jr; GAGLIO, P.J. Scope of worldwide hepatitis C problem. **Liver Transpl**, v.9, n.11, p.S10-13, 2003.

BRÜNGER, A.T. The Free R value: A novel statistical quantity for assessing the accuracy of crystal structures. **Nature**, v.355, p.472-474, 1992.

BUJNICKI, J. **Prediction of Protein Structures, Functions, and Interactions**. Ed. John Wiley & Sons, 2008, 302p.

BÜRCKSTÜMMER, T. et al. Raf-1 kinase associates with Hepatitis C virus NS5A and regulates viral replication. **FEBS Lett**, v.580, n.2, 2006.

BURLEY, S.K. et al. Structural genomics: beyond the human genome project. **Nat. Genet.** v. 23, p.151-157, 1999.

CAMPIOTTO, S. et al. Geographic distribution of hepatitis C virus genotypes in Brazil. **Braz J Med Bio Res**, v.38, n.01, p.41-49, 2005.

CHAMBERS, T.J. et al. *Quasispecies* heterogeneity within the E1/E2 region as a pretreatment variable during pegylated interferon therapy of chronic hepatitis C virus infection. **J Virol**, v.79, n.5, p.3071-3083, 2005.

CHOO, Q.L. et al. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. **Science**, v.244, n.4902, p.362-364, 1989.

CONRY-CANTILENA, C. et al. Routes of infection, viremia, and liver disease in blood donors found to have hepatitis C virus infection. **N Engl J Med**, v.334, n.26, p.1691-1696, 1996.

CROCKER, P.R. Siglecs: sialic-acid-binding immunoglobulin-like lectins in cell-cell interactions and signalling. **Curr Opin Struct Biol**, v.12, p.609-615, 2002.

ENGH, R.A.; HUBER, R. Accurate bond and angle parameters for X-ray protein structure refinement. **Acta Cryst.**, v.A47, p.392-400, 1991.

ENOMOTO, N. et al. Mutations in the nonstructural protein 5A gene and response to interferon in patients with chronic hepatitis C virus 1b infection. **N Engl J Med**, v. 334, n. 2, p. 77-81, 1996.,

ESTEBAN, J.I.; SAULEDA, S.; QUER, J. The changing epidemiology of hepatitis C virus infection in Europe. **J Hepatol**, v.49, n.1, p.148-162, 2008.

FARAZI, T.A.; WAKSMAN, G.; GORDON, J.I. The Biology and Enzymology of Protein N-Myristoylation. **The Journal of Biological Chemistry**, v.276, n.43, p.39501-39504, 2001.

FELD, J.J.; HOOFNAGLE, J.H. Mechanism of action of interferon and ribavirin in treatment of hepatitis C. **Nature**, v.436, n.7053, p.967-972, 2005.

FRIED, M.W.; HADZIYANNIS, S.J. Treatment of chronic hepatitis C infection with peginterferons plus ribavirin. **Semin Liver Dis**, v.24, p.47-54, 2004.

FUKUMA, T. et al. Mutations in the interferon-sensitivity determining region of hepatitis C virus and transcriptional activity of the nonstructural region 5A protein. **Hepatology**, v.28, p.1147-1152, 1998.

GALE Jr., M. et al. Control of PKR protein kinase by hepatitis C virus nonstructural 5A protein: molecular mechanisms of kinase regulation. **Mol Cell Biol**, v.18, n.09, p.5208-5218, 1998.

GALE, M. Jr.; FOY, E.M. Evasion of intracellular host defense by hepatitis C virus. **Nature**, v.436, n.7053, p.939-945, 2005.

GAST, D. et al. The RGD integrin binding site in human L1-CAM is important for nuclear signaling. **Exp Cell Res**, v.314, n.13, p.2411-2418, 2008.

GIBAS, C.; JAMBECK, P. **Desenvolvendo a Bioinformática**, Ed. Campus, Rio de Janeiro, 2001.

GONZALEZ-PERALTA, R.P. Clinical implications of viral *quasispecies* heterogeneity in chronic hepatitis C. **J Med Virol**. v.49, n.03, p.242-247, 1996.

GORDON, J.I. et al. Protein N-Myristoylation. **The Journal of Biological Chemistry**, v.266, n.14, p.8647-8650, 1991.

GRAFF, J. et al. Mutations within potential glycosylation sites in the capsid protein of hepatitis E virus prevent the formation of infectious virus particles. **J Virol**, v.82, n.3, p.1185-1194, 2008.

GRAHOVAC, B. et al. Dynamics of serum hepatitis C virus load and *quasispecies* complexity during antiviral therapy in patients with chronic hepatitis C. **J Clin Virol**. v.20, n.1-2, p.85-89, 2001.

HANOULLE, X. et al. Domain 3 of non-structural protein 5A from hepatitis C virus is natively unfolded. **Biochem Biophys Res Commun**, v.381, n.4, p.634-638, 2009.

HELENIUS, A. How N-linked oligosaccharides affect glycoprotein folding in the endoplasmic reticulum. **Mol Biol Cell**, v.5, n.3, p.253-265, 1994.

HIDAJAT, R. et al. Cleavage of the hepatitis C virus NS5A protein by caspase-3 in the interferon sensitivity-determining region in a sequence-dependent manner. **Kobe J Med Sci**, v.50, p.153-166, 2004.

HIGGINS, D.G.; TAYLOR, W.R. Multiple sequence alignment. **Methods Mol Biol**, v.143, p.1-18, 2000.

HOFMANN, W.P.; ZEUZEM, S.; SARRAZIN, C. Hepatitis C virus-related resistance mechanisms to interferon alpha-based antiviral therapy. **J Clin Virol**, v.32, n.02, p.86-91, 2005.

HOOFNAGLE, J.H. et al. Early changes in hepatitis C virus (HCV) levels in response to peginterferon and ribavirin treatment in patients with chronic HCV genotype 1 infection. **J Infect Dis**, v.199, n.8, p.1112-1120, 2009.

HOOFT, R. W. et al. Errors in protein structures. **Nature**, v. 381, p. 272, 1996.

HUANG, R.H.; HU, K.Q. A practical approach to managing patients with HCV infection. **Int J Med Sci**, v.3, n.2, p.63-68, 2006.

HUANG, Y. et al. Phosphorylation of hepatitis C virus NS5A nonstructural protein: a new paradigm for phosphorylation-dependent viral RNA replication? **Virology**, v. 364, n.1, p.1-9, 2007.

HUGHES, M. et al. A conserved proline between domains II and III of hepatitis C virus NS5A influences both RNA replication and virus assembly, **J Virol**, [PUBMED: submetido a publicação], acesso em 23 de agosto de 2009.

IDE, Y. et al. Characterization of the nuclear localization signal and subcellular distribution of hepatitis C virus nonstructural protein NS5A. **Gene**, v.182, n.1-2, p.203-211, 1996.

INUBUSHI, S. et al. Hepatitis C virus NS5A protein interacts with and negatively regulates the non-receptor protein tyrosine kinase Syk. **J Gen Virol**, v.89, p.1231-1242, 2008.

JENSEN, L.J. et al. Ab initio prediction of human orphan protein function from post-translational modifications and localization features. **J Mol Biol**, v.319, p.1257-1265, 2002.

JONES, D.T. Improving the accuracy of transmembrane protein topology prediction using evolutionary information. **Bioinformatics**, v.23, n.5, p.538-544, 2007.

KALAMVOKI, M.; GEORGOPOULOU, U.; MAVROMARA, P. The NS5A protein of the hepatitis C virus genotype 1a is cleaved by caspases to produce C-terminal-truncated forms of the protein that reside mainly in the cytosol. **J Biol Chem**, v.281, p.13449-13462, 2006.

KEARSLEY, S.K. On the orthogonal transformation used for structural comparisons. **Acta Cryst**, v.45, n.2, p.208-210, 1989.

KIM, J.E. et al. Subcellular localization of hepatitis C viral proteins in mammalian cells. **Arch Virol**, v.144, p.329-343, 1999.

KIM, J.M. et al. A single N-linked glycosylation site in the Japanese encephalitis virus prM protein is critical for cell type-specific prM protein biogenesis, virus particle release, and pathogenicity in mice. **J Virol**, v.82, n.16, p.7846-7862, 2008.

KLENERMAN, P.; FLEMING, V.; BARNES, E. What are the prospects for controlling hepatitis C? **PLoS Med**, v.6, n.6 p.e1000096, 2009.

KRAJCSI, P.; WOLD, W.S. Viral proteins that regulate cellular signalling. **J Gen Virol**, v.79, n.Pt6, p.1323-1335, 1998.

KUIKEN, C.; SIMMONDS, P. Nomenclature and numbering of the hepatitis C virus. **Methods Mol Biol**, v.510, p.33-53, 2009.

LASKOWSKI, R. A. et al. Validation of protein models derived from experiment. **Current Opinion in Structural Biology**, v. 8, p. 631-639, 1998.

LIANG, Y. et al. Domain 2 of nonstructural protein 5A (NS5A) of hepatitis C virus is natively unfolded. **Biochemistry**, v. 46, n.41, p.11550-11558, 2007.

LIN, C. et al. The hepatitis C virus NS4A protein: interactions with the NS4B and NS5A proteins. **J Virol**, v.71, p.6465-6471, 1997.

LINDENBACH, B.D. et al. Complete replication of hepatitis C virus in cell culture. **Science**, v.309, n.5734, p.623-626, 2005.

LIU, S. et al. Insertion and deletion analyses identify regions of non-structural protein 5A of Hepatitis C virus that are dispensable for viral genome replication. **J Gen Virol**, v.87, p.323-327, 2006.

LOVE, R.A. et al. Crystal structure of a novel dimeric form of NS5A domain I protein from hepatitis C virus. **J Virol**, v.83, n.9, p.4395-4403, 2009.

LYRA, A.C.; FAN, X.; DI BISCEGLIE, A.M. Molecular biology and clinical implication of hepatitis C virus. **Braz Med Biol Res**. v.37, n.05, p.691-695, 2004.

MACDONALD, A.; HARRIS, M. Hepatitis C virus NS5A: tales of a promiscuous protein. **J Gen Virol** v.85, p.2485-2502, 2004.

MANABE, S. et al. Production of nonstructural proteins of hepatitis C virus requires a putative viral protease encoded by NS3. **Virology**, v.198, p.636-644, 1994.

MARTI-RENO, M.A. et al. Reliability of assessment of protein structure prediction methods. **Structure**, v.10, n.3, p.435-440, 2002.

MASAKI, T. et al. Interaction of hepatitis C virus nonstructural protein 5A with core protein is critical for the production of infectious virus particles. **J Virol**, v.82, n.16, 7964-7976, 2008.

MCGUFFIN, L.J. et al. High throughput profile-profile based fold recognition for the entire human proteome. **BMC Bioinformatics**, v.7, p.288, 2006.

MORADPOUR, D. et al. Hepatitis C: na update. **Swiss Med Wkly**, v.131, n.21-22, p.291-298, 2001.

MORADPOUR, D.; BLUM, H.E. A primer on the molecular virology of hepatitis C. **Liver Int**. v.24, n.06, p.519-525, 2004.

MORADPOUR, D.; PENIN, F.; RICE, C.M. Replication of hepatitis C virus. **Nat Rev Microbiol**, v.5, n.6, p.435-463, 2007.

MORENO, M.P. et al. Evidence of recombination in *quasispecies* populations of a Hepatitis C Virus patient undergoing anti-viral therapy. **Virology**, v.24, p.3-87, 2006.

MORENO-OTERO, R. Therapeutic modalities in hepatitis C: challenges and development. **J Viral Hepat**, v.12, n.1, p.10-19, 2005.

NANDURI, B. et al. Proteomic analysis using an unfinished bacterial genome: the effects of subminimum inhibitory concentrations of antibiotics on *Mannheimia haemolytica* virulence factor expression. **Proteomics**, v. 5, p.4852–4863, 2005.

NEEDLEMAN, S.B.; WUNSCH, C.D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. **J Mol Biol**, v.48, n.3, p.443-453.

NEUMANN, A.U. et al. Hepatitis C viral dynamics *in vivo* and the antiviral efficacy of interferon-alpha therapy. **Science**, v.282, n.5386, p.103-107, 1998.

OBA, M. et al. Polyplex micelles with cyclic RGD peptide ligands and disulfide cross-links directing to the enhanced transfection via controlled intracellular trafficking. **Mol Pharm**, v.5, n.6, p.1080-1092, 2008.

OHLSON, T.; WALLNER, B.; ELOFSSON, A. Profile-profile methods provide improved fold-recognition: a study of different profile-profile alignment methods. **Proteins**, v.57, n.1, p.188-197, 2004.

OKADA, H. et al. The degree of variability in the amino terminal region of the E2/NS1 protein of hepatitis C virus correlates with responsiveness to interferon therapy in viremic patients. **Hepatology**, v. 16, n. 3, p. 619-624, 1992.

ORENGO, C.A.; JONES, D.T.; THORNTON, J.M. **Bioinformatics: Genes, Proteins & Computers**. Oxford: BIOS Scientific Publishers Ltd, 2003, 300p.

PENIN, F. et al. Structure and Function of the Membrane Anchor Domain of Hepatitis C Virus Nonstructural Protein 5A. **Journal of Biological Chemistry**, v. 279, p. 40835-40843, 2004.

PETRESCU, S.M. et al. Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. **Glycobiology**, v.14, n.2, p.103-114, 2004.

PHILLIPS, J.C. et al. Scalable Molecular Dynamics with NAMD. **J Comput Chem**, v.26, n.16, p.1781-1802, 2005.

PUIG-BASAGOITI, F. et al. Dynamics of hepatitis C virus NS5A *quasispecies* during interferon and ribavirin therapy in responder and non-responder patients with genotype 1b chronic hepatitis C. **J Gen Virol**, v. 86, p. 1067-1075, 2005.

PYBUS, O.G. et al. The epidemic behavior of the hepatitis C virus. **Science**, v.292, n.5525, p.2323-2325, 2001.

REED, K.E.; XU, J.; RICE, C.M. Phosphorylation of the hepatitis C virus NS5A protein *in vitro* and *in vivo*: properties of the NS5A-associated kinase. **J Virol**, v.71, n.10, p.7187-7197, 1997.

RHEE, S.Y. et al. Use and misuse of the gene ontology annotations. **Nat Rev Genet**, v.9, p.509–515, 2008.

ROST, B.; SANDER, C. Improved prediction of protein secondary structure by use of sequence profiles and neural networks. **Proc Natl Acad Sci U S A**, v.90, n.16, p.7558-7562, 1993.

ROST, B.; YACHDAV, G.; LIU, J. The PredictProtein Server. **Nucleic Acids Research**, v.32(Web Server issue), p.W321-W326, 2004.

ROVÈRE, C. et al. The RGD motivo and the C-terminal segment of proprotein convertase 1 are critical for its cellular trafficking but not for its intracellular binding to integrin $\alpha 5 \beta 1$. **J Biol Chem**, v.274, n.18, p.12461-12467, 1999.

RUOSLAHTI, E. ; PIERSCHBACHER, M.D. Arg-Gly-Asp: a versatile cell recognition signal. **Cell**, v.44, n.4, p.517-518, 1986.

SAITO, I. et al. Hepatitis C virus infection is associated with the development of hepatocellular carcinoma. **Proc Natl Acad Sci U S A**, v.87, n.17, p.6547-6549, 1990.

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

SALI, A.; OVERINGTON, J.P. Derivation of rules for comparative protein modeling from a database of protein structure alignments. **Protein Sci**, v.3, n.9, p.1582-1596, 1994.

SALI, A. et al. Evaluation of comparative protein modeling by *Modeller*. **Proteins**, v.23, n.3, p.318-326, 1995.

SALMERON, J. et al. *Quasispecies* as predictive response factors for antiviral treatment in patients with chronic hepatitis C. **Digestive Diseases and Sciences**, v. 51, n. 5, p. 960-967, 2006.

SÁNCHEZ, R.; ŠALI, A. Advances in comparative protein-structure modeling. **Current Opinion in Structural Biology**, v. 7, p. 206-214, 1997.

SÁNCHEZ, R.; ŠALI, A. Advances in comparative protein-structure modeling. **Current Opinion in Structural Biology**, v. 7, p. 206-214, 1997.

SAUTER, D. et al. Localization determines function: N-terminally truncated NS5A fragments accumulate in the nucleus and impair HCV replication. **J Hepatol**, v.50, n.5, p.861-871, 2009.

SCHWIETERS, C.D.; KUSZEWSKI, J.J.; TJANDRA, N; CLORE, G.M. The Xplor-NIH NMR Molecular Structure Determination Package. **J. Magn. Res.**, v.160, p.66-74, 2003.

SEVERINO, P. et al. Comparative transcriptome analysis of *Listeria monocytogenes* strains of the two major lineages reveals differences in virulence, cell wall, and stress response. **Appl. Environ. Microbiol**, v.73, p.6078–6088, 2007.

SHIROTA, Y. et al. Hepatitis C virus NS5A binds RNA-dependent RNA polymerase NS5B and modulates RNA-dependente RNA polymerase activity. **J Biol Chem**, v.277, p.11149-11155, 2002.

SIMMONDS, P. Genetic diversity and evolution of hepatitis C virus – 15 years on. **J Gen Virol**, v.85, p.3173-3188, 2004.

SOLDAN, K. et al. Estimation of the risk of hepatitis B virus, hepatitis C virus and human immunodeficiency virus infectious donations entering the blood supply in England, 1993-2001. **Vox Sang**, v.84, n.04, p.274-286, 2003.

SONG, J. et al. Nuclear localization and intramolecular cleavage of N-terminally deleted NS5A protein of hepatitis C virus. **Virus Research**, v.69, p.109-117, 2000.

SY, T.; JAMAL, M.M. Epidemiology of hepatitis C virus (HCV) infection. **Int J Med Sci**, v.03, n.02, p.41-46, 2006.

TAN, B.H.; FU, J.L.; SUGRUE, R.J. Characterization of the dengue virus envelope glycoprotein expressed in *Pichia pastoris*. **Methods Mol Biol**, v.379, p.163-176, 2007.

TAN, S.L. et al. NS5A, a nonstructural protein of hepatitis C virus, binds growth factor receptor-bound protein 2 adaptor protein in a Src homology 3 domain/ligand-dependent manner and perturbs mitogenic signaling. **Proc Natl Acad Sci U S A**, v.96, n.10, p.5533-5538, 1999.

TANG, X. et al. Relationship between diversity of hepatitis C virus *quasispecies* and viremia, activity of liver disease and response to interferon therapy. **Zhonghua Shi Yan He Lin Chuang Bing Du Xue Za Zhi**, v.16, n.02, p.128-131, 2002.

TANIMOTO, A. et al. The amino terminal deletion mutants of hepatitis C virus nonstructural protein NS5A function as transcriptional activators in yeast. **Biochem Biophys Res Commun**, v.236, p.360-364.

TANJI, Y.; SATOH, S.; KANEKO, T. et al. Phosphorylation of hepatitis C virus-encoded nonstructural protein NS5A. **J Virol**, v.69, n.07, p.3980-3986, 1995.

TAYLOR, M.E.; DRICKAMER, K. Structural requirements for high affinity binding of complex ligands by the macrophage mannose receptor. **J Biol Chem**, v.268, n.1, p.399-404, 1993.

TELLINGHUISEN, T.L. et al. The NS5A protein of hepatitis C virus is a zinc metalloprotein. **J Biol Chem**, v.10, n.1074, p.363-366, 2004.

TELLINGHUISEN, T.L.; MARCOTRIGIANO, J.; RICE, C.M. Function follows form: the structure of n-terminal domain of HCV NS5A. **Hepatol Elsewhere**, v.42, n.03, p.732-735, 2005.

TELLINGHUISEN, T.L. et al. Identification of Residues Required for RNA Replication in Domains II and III of the Hepatitis C Virus NS5A Protein. **J Virol**, v.82, p.1073–1083, 2008.

TELLINGHUISEN, T.L.; FOSS, K.L.; TREADAWAY, J. Regulation of hepatitis C virion production via phosphorylation of the NS5A protein. **PLoS Pathog**, v.4, n.3, p.e1000032, 2008.

THELU, M.A. et al. Dynamics of viral *quasispecies* during interferon therapy in non responder chronic hepatitis C patients. **J Clin Virol.** v.22, n.01, p.125-131, 2001.

TORRES-PUENTE, M. et al. Genetic variability in hepatitis C virus and its role in antiviral treatment response. **J Viral Hepat,** v.15, n.3, p.188-199.

WASLEY, A.; ALTER M.J. Epidemiology of hepatitis C: geographic differences and temporal trends. **Semin Liver Dis,** v.20, n.1, p.1-16, 2000.

WOHNSLAND, A.; HOFMANN, W.P.; SARRAZIN, C. Viral determinants of resistance to treatment in patients with hepatitis C. **Clin Microbiol Rev,** v.20, n.1, p.23-38, 2007.

WRIGHT, P. E., and DYSON, H. J. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm, **J. Mol. Biol.** v.293, p.321-331, 1999.

YE, S.Q. **Bioinformatics: A Practical Approach.** Boca Raton: Chapman & Hall/CRC Taylor & Francis Group, 2008, 618p.

9. Anexo



FACULDADE DE MEDICINA DE SÃO JOSÉ DO RIO PRETO
AUTARQUIA ESTADUAL - LEI Nº 8899, de 27/09/94
[Reconhecida pelo Decreto Federal nº 74.179, de 14/06/74]

Parecer n.º 087/2004

COMITÊ DE ÉTICA EM PESQUISA

O Protocolo n.º 2025/2004 sob a responsabilidade de Roberta Maria Fachini com o título "Estudo epidemiológico das principais vias de transmissão do vírus da hepatite C, em uma população fechada e descrição da evolução da carga viral com o tratamento segundo a genotipagem" está de acordo com a Resolução CNS 196/96 e foi aprovado por esse CEP.

Lembramos ao senhor(a) pesquisador(a) que, no cumprimento da Resolução 251/97, o Comitê de Ética em Pesquisa (CEP) deverá receber relatórios semestrais sobre o andamento do Estudo, bem como a qualquer tempo e a critério do pesquisador nos casos de relevância, além do envio dos relatos de eventos adversos, para conhecimento deste Comitê. Salientamos ainda, a necessidade de relatório completo ao final do Estudo.

São José do Rio Preto, 10 de maio de 2004.


Prof.ª Dr.ª Patrícia Maluf Cury
Coordenadora do CEP/FAMERP