



**UNIVERSIDADE FEDERAL DA BAHIA  
INSTITUTO DE BIOLOGIA**

**ANÁLISE *IN SILICO* DE PROTEÍNAS: MODELO BASEADO NA  
APOLIPOPROTEÍNA-E RELACIONADA À DOENÇA DE ALZHEIMER**

por

**FERNANDA ORPINELLI RAMOS DO REGO**

TCC apresentado ao Instituto de Biologia da Universidade Federal da Bahia, como exigência para obtenção do grau de Bacharel em Ciências Biológicas.

Salvador  
2012

# AVALIAÇÃO DA BANCA EXAMINADORA

Data da Defesa: 29/06/2012

**BANCA EXAMINADORA**

---

Dra. FLORA MARIA DE CAMPOS FERNANDES  
UNIVERSIDADE FEDERAL DA BAHIA

---

Dr. GILBERTO CAFEZEIRO BOMFIM  
UNIVERSIDADE FEDERAL DA BAHIA

---

Dr. ARTUR TRANCOSO LOPO DE QUEIROZ  
CENTRO DE PESQUISA GOLÇALO MONIZ – FIOCRUZ-BA

*"Nunca deixe que alguém te diga que não pode fazer algo. Nem mesmo eu. Se você tem um sonho, tem que protegê-lo. As pessoas que não podem fazer por si mesmas, dirão que você não consegue. Se quer alguma coisa, vá e lute por ela. Ponto final."*

*[Will Smith, em À procura da felicidade]*

## RESUMO

---

A possibilidade de analisar sequências de macromoléculas como DNA, RNA e proteínas em menor tempo e custo, comparados aos métodos experimentais, contribuem para a utilização dos métodos *in silico*. Além de análises de sequências genômicas ou proteômicas, as aplicações da bioinformática se estendem às análises funcionais e estruturais. No presente trabalho, utilizaram-se ferramentas computacionais para o estudo estrutural de proteínas, através da modelagem computacional e de análises de propriedades importantes para a função proteica, tendo como modelo a apolipoproteína-E relacionada à Doença de Alzheimer. A presença de polimorfismos no gene da apoE gera três isoformas, sendo uma delas o maior fator de risco genético para desenvolvimento da doença (apoE4), assim estabelecendo-se a relação entre a doença e as alterações na proteína. Com intuito de avaliar *in silico* as diferenças estruturais existentes entre as três isoformas e analisar as modificações das propriedades exibidas por cada uma, foi feita a modelagem da proteína pelo método de similaridade (*threading*), ou reconhecimento de dobras a partir da obtenção das sequências de aminoácidos em bancos de dados de proteínas. Análises dos ângulos torsionais mostraram maior compactação estrutural da isoforma apoE4, em comparação com as outras, o que evidencia a existência da interação de domínios somente nesta isoforma. A instabilidade gerada pela interação de domínios é refletida na formação defeituosa de complexos com outras moléculas, *in silico*, apresentada pela E4. O potencial eletrostático e a carga das isoformas, determinados computacionalmente, também mostraram diferenças entre as isoformas e podem estar relacionadas às características patogênicas da variante E4. Análises de polaridade demonstraram que as isoformas E2 e E3 foram mais hidrofóbicas que a E4. As alterações causadas pela troca de um ou dois aminoácidos, de uma isoforma para outra, alteram as propriedades de interação entre elas e outras moléculas. Estas alterações acarretam desde mudanças conformacionais ao comprometimento funcional. Assim, os métodos *in silico* proporcionaram o estudo da apolipoproteína-E, através de modelagem computacional, e os dados das análises proteicas corroboraram dados descritos experimentalmente por outros autores.

## ABSTRACT

---

The ability to analyze sequences of macromolecules such as DNA, RNA and proteins in the shortest time and cost, compared to experimental methods, contribute to recurrence *in silico* methods. In addition to genomic or proteomics sequence analyses, bioinformatics applications spanning the functional and structural analyses. In this work, computational tools have been used for the structural study of proteins, by computational modeling and analysis of important properties for the protein function, taking as a model the apolipoprotein-E related to Alzheimer's disease. The relationship between the disease and the protein is given by the presence of polymorphisms in the gene apoE, which generate three isoforms, one being the highest genetic risk factor for developing the disease (apoE4). In order to assess *in silico* structural differences existing between the three isoforms and analyze the changes of properties displayed by each one, the protein was modeled by the similarity method (threading) or fold recognition from the sequences of amino acids obtained in protein databases. Analyses of torsion angles showed greater structural compression of the apoE4 isoform, in comparison with the other, which evidences the existence of domain interaction only in this isoform. The instability generated by the interaction of the defective areas is reflected in the complexes formation with other molecules, *in silico*, presented by E4. The electrostatic potential and the charge of isoforms, determined computationally, also showed differences between the isoforms and may be related to pathogenic characteristics of the E4 variant. Polarity analysis showed that isoforms E2 and E3 were more hydrophobic than the E4. The differences caused by exchanging one or two amino acids, from one isoform to another, change the properties of the interaction between them and other molecules. These amendments involve since conformational changes until functional impairment. Thus, *in silico* methods provided the study of apolipoprotein-E, through computational modeling, and protein analysis corroborate experimentally data described by other authors.

## AGRADECIMENTOS

---

A Deus, por ter me dado coragem para enfrentar todos os desafios e por ter guiado meu caminho durante toda a minha vida.

Aos meus pais, Paulo e Edna, pelo apoio, amor e por terem me dado as melhores oportunidades para conquistar o que conquistei até hoje;

Aos meus avós, Sylvino, Regina e Cida, por terem rezado por mim, me dado força e transformado cada reencontro saudoso em uma imensa alegria de estar perto;

Ao meu irmão, Guilherme, por ser um exemplo em determinação e dedicação;

A minha orientadora, Flora, por ter me dado uma oportunidade desde a primeira conversa que tivemos, por ter acreditado em mim, por ter me incentivado, por todo o apoio... nunca vou esquecer;

A Artur Trancoso, pela paciência, estímulo e por tornar mais leves e engraçadas todas as apreensões deste momento;

A Gilberto Bomfim e Carlos Vaccari, pela atenção e discussões, sempre muito pertinentes;

A Ayling, Marcela e Flávia, obrigada por me acolherem e terem enfrentado as maiores e melhores ciladas comigo! Que nossa amizade esteja somente no começo e que a gente ainda passe por muita coisa juntas, rindo e chorando;

Aos colegas do PANGEA, por todos os momentos, desde os mais engraçados aos mais difíceis;

Aos colegas da UFBA, por todos os anos de convivência e aprendizado;

A Lucas, meu namorado, por todo o amor, carinho, compreensão e por sempre ter acreditado em mim, principalmente quando eu mesma não acreditava... Você é essencial;

A todos aqueles que, de alguma forma, estiveram presentes durante a realização trabalho.

Obrigada a todos, de coração!

*Dedico este trabalho aos meus pais, que  
dedicaram suas vidas à minha.*

# SUMÁRIO

---

RESUMO

ABSTRACT

AGRADECIMENTOS

|  |      |
|--|------|
| SUMÁRIO .....  | 8    |
| 1. INTRODUÇÃO .....  | 9    |
| 1.1 Estudo <i>in silico</i> : bioinformática.....            | 9    |
| 1.2 Aplicação da Bioinformática no estudo de proteínas ..... | 9    |
| 1.3 Análises estruturais das macromoléculas.....             | 11   |
| 1.4 Modelo de estudo: apolipoproteína-E .....                | 14   |
| 2. OBJETIVOS.....  | 17   |
| 3. METODOLOGIA.....  | 18   |
| 3.1 Sequências de aminoácidos da apolipoproteína-E .....     | 18   |
| 3.2 Edição das sequências .....                              | 18   |
| 3.3 Construção dos modelos proteicos.....                    | 19   |
| 3.4 Qualidade dos modelos preditos .....                     | 19   |
| 3.5 Análise dos modelos proteicos da apolipoproteína-E ..... | 19   |
| 4. RESULTADOS .....  | 22   |
| 5. DISCUSSÃO .....   | 35   |
| 5.1 MODELOS .....  | 35   |
| 5.2 RAMACHANDRAN .....                                       | 36   |
| 5.3 ÂNGULOS DE TORSÃO .....                                  | 36   |
| 5.4 COMPLEXOS apoE-A $\beta$ .....                           | 3838 |
| 5.5 PERFIL FÍSICO-QUÍMICO.....                               | 39   |
| 5.6 POTENCIAL ELETROSTÁTICO E CARGAS .....                   | 4040 |
| 6. CONCLUSÕES.....   | 41   |
| REFERÊNCIAS BIBLIOGRÁFICAS .....                             | 422  |



# 1. INTRODUÇÃO

---

## 1.1 Estudo *in silico*: bioinformática

O avanço da biologia nas últimas décadas, em especial na área molecular, associado a avanços tecnológicos, com o desenvolvimento de computadores velozes e com grande capacidade para armazenamento de informações, proporcionou o surgimento de uma das mais promissoras áreas de estudo, a bioinformática. Esta nova ciência surge da necessidade de se analisar o crescente volume de dados gerados a partir dos investimentos em novas tecnologias moleculares. Assim, a bioinformática pode ser definida como a união de áreas já conhecidas e indispensáveis à biologia moderna, como matemática, estatística e ferramentas computacionais, com a biologia molecular para as mais diversas aplicações (GOLDING, 2003; XIONG, 2006).

As aplicações da bioinformática podem se dividir em três vertentes principais que correspondem à análise de sequências de macro (DNA e proteínas) e micromoléculas (transcritos e metabólitos), análises funcionais e análises estruturais. As análises de sequências de macromoléculas englobam comparações genômicas, estudos filogenéticos, alinhamento e edição de sequências, identificação de regiões mutadas, identificação de epítomos, reconstrução de redes haplotípicas, inferências de coalescência, entre outros. As análises funcionais correspondem à predição de vias metabólicas através de interações entre proteínas, essenciais para o entendimento da relação entre os genes, seus produtos e, quando necessário, sua correlação com doenças, análises de expressão gênica e identificação de sítios ativos. Análises estruturais compreendem predições de estruturas secundárias, terciárias de macromoléculas, comparação e classificação de proteínas, DNA e RNA (TSOKA; OUZOUNIS, 2000; XIONG, 2006).

## 1.2 Aplicação da Bioinformática no estudo de proteínas

Dentre as áreas de estudo da bioinformática, as análises proteômicas estão em destaque, devido a importância das proteínas nas funções químicas e biológicas essenciais ao metabolismo celular. Compreender a relação entre a estrutura primária, a sequência de aminoácidos das proteínas e sua estrutura tridimensional torna-se um dos maiores objetivos da bioinformática estrutural, principalmente porque a função destas

macromoléculas é fortemente influenciada por sua estrutura (GOLDING, 2003; MOUNT, 2004; XIONG, 2006).

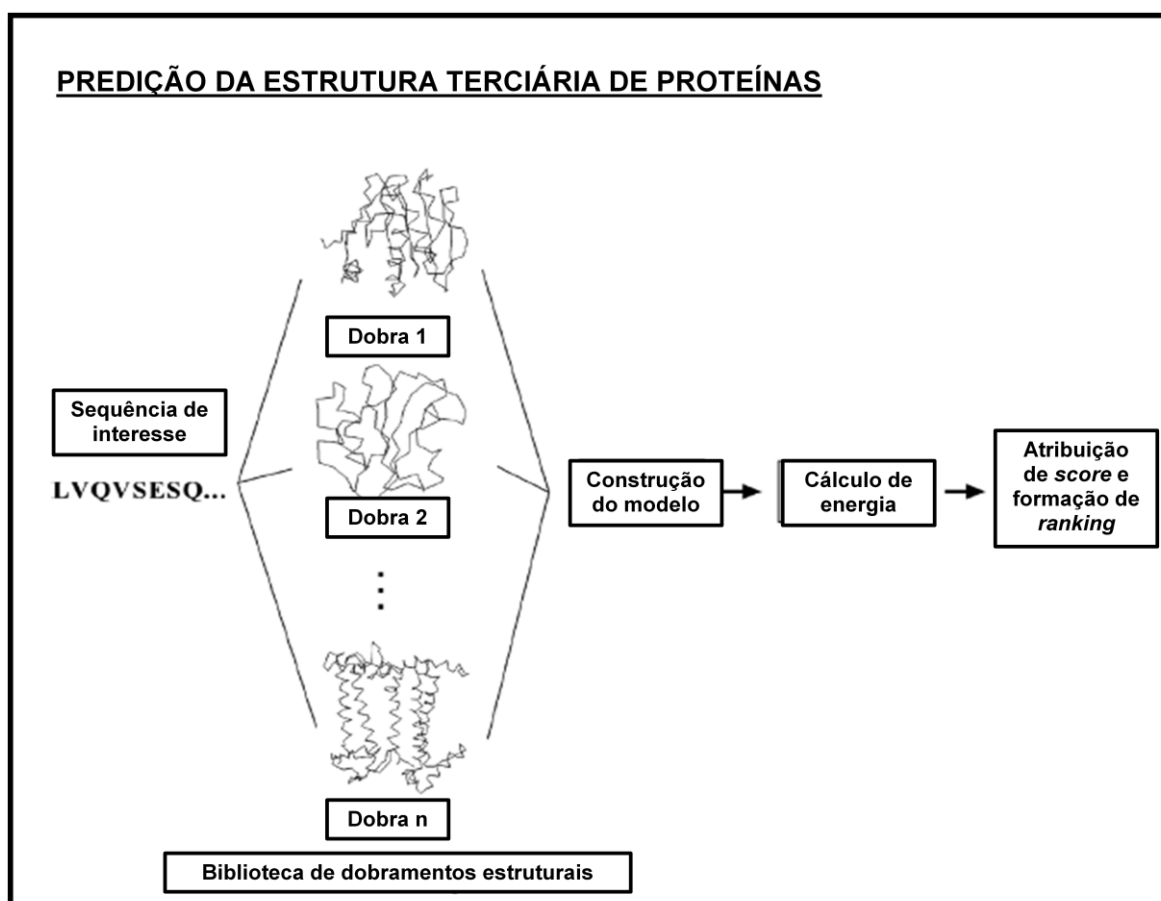
Macromoléculas, como proteínas, apresentam diferentes níveis de organização, que correspondem às hierarquias estruturais, que podem ser primárias, secundárias, terciárias e quaternárias (LEHNINGER; NELSON; COX, 2003). Esta organização estrutural em níveis hierárquicos é o que permite a predição computacional das estruturas de proteínas nos seus diferentes arranjos.

Existem hoje muitos preditores, principalmente de estruturas secundárias e terciárias de proteínas, que se diferenciam pela metodologia e algoritmos utilizados. A predição da estrutura secundária, resultante de pontes de hidrogênios formadas entre os aminoácidos, é fundamental para a predição da estrutura tridimensional, uma vez que é a partir das interações entre os aminoácidos que as proteínas adquirem as dobras características da estrutura terciária (MOUNT, 2004).

A modelagem da estrutura terciária de proteínas pode ser feita a partir de três diferentes metodologias, que são: *ab initio*, modelagem por homologia e *threading* de proteínas. O método *ab initio* corresponde à predição a partir do início, sem um molde para se basear. Este método é utilizado no caso de proteínas que não apresentam moléculas homólogas de estrutura conhecida, sendo os modelos preditos utilizando-se apenas informações físico-químicas dos resíduos. Os *softwares* que utilizam a homologia como ferramenta para modelagem permitem a obtenção dos modelos através de informações de similaridade entre a sequência da proteína de interesse e estruturas depositadas em bancos de dados que foram determinadas experimentalmente para, posteriormente, extrair informações de ancestralidade comum entre elas. Há ainda outros *softwares* que permitem a realização da modelagem de macromoléculas por reconhecimento de dobras das proteínas, através do método de *threading* de proteínas ou *structural fold recognition*. Este método parte do princípio de que existem muitas proteínas de sequências que não são relacionadas evolutivamente e apresentam padrões de dobramentos similares, e de que, além disso, a quantidade de sequências existente é muito superior à quantidade de dobramentos possíveis (XIONG, 2006).

O *threading* de proteínas refere-se à similaridade de sequências de aminoácidos de proteínas, cuja estrutura tridimensional é desconhecida, com estruturas de proteínas conhecidas para estimar o arranjo estrutural da proteína de interesse (POLANSKI; KIMMEL, 2007; LATHROP et al., 1998). Para isto, é realizada a seleção de uma proteína

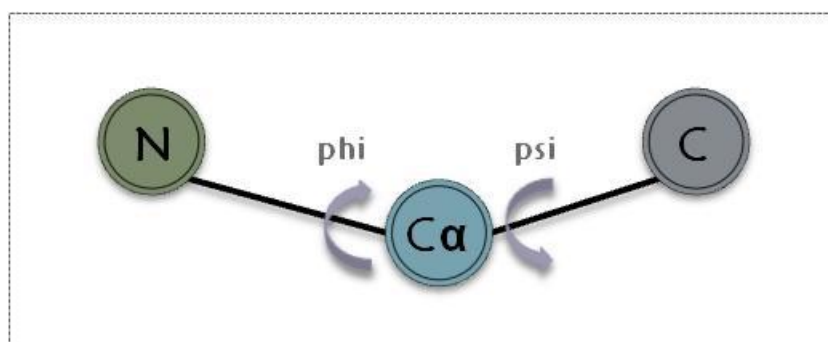
de interesse e a comparação desta macromolécula, a qual não possui estrutura tridimensional definida, com proteínas de estruturas conhecidas disponíveis em bibliotecas de bancos de dados estruturais (figura 1). Posteriormente, são realizados encaixes entre as proteínas de estruturas conhecidas (proteínas de referência) e a sequência da proteína de interesse. Por fim, a construção do modelo é feita a partir de dados dos dobramentos exibidos pelo molde de referência, além do uso de informações da estrutura secundária previstas para a sequência de interesse (LATHROP et al., 1998). Desta forma, o *threading* baseia-se nas semelhanças estruturais apresentadas pelas proteínas, mesmo que não haja informações de ancestralidade comum entre elas (LESK, 2006; XIONG, 2006), ou seja, este é um método baseado em similaridade de topologias proteicas.



**Figura 1.** Fluxograma do método *threading* de predição da estrutura tridimensional de proteínas. Fonte: figura adaptada de XIONG, 2006.

### 1.3 Análises estruturais das macromoléculas

A realização da modelagem das estruturas proteicas requer a validação dos modelos através de análises de qualidade. Uma forma de avaliar a qualidade é através dos ângulos dos dobramentos da estrutura terciária de macromoléculas, uma vez que estes ângulos são essenciais para a determinação da estrutura terciária e da conformação nativa das proteínas. Os ângulos torsionais (figura 2) correspondem ao ângulo phi (ângulo entre o carbono- $\alpha$  e o nitrogênio do grupamento amino) e psi (ângulo formado entre o carbono- $\alpha$  e o carbono do grupamento carboxi) de cada aminoácido, que são plotados no gráfico de Ramachandran, o qual fornece informações acerca da qualidade do arranjo estrutural (MOUNT, 2004).



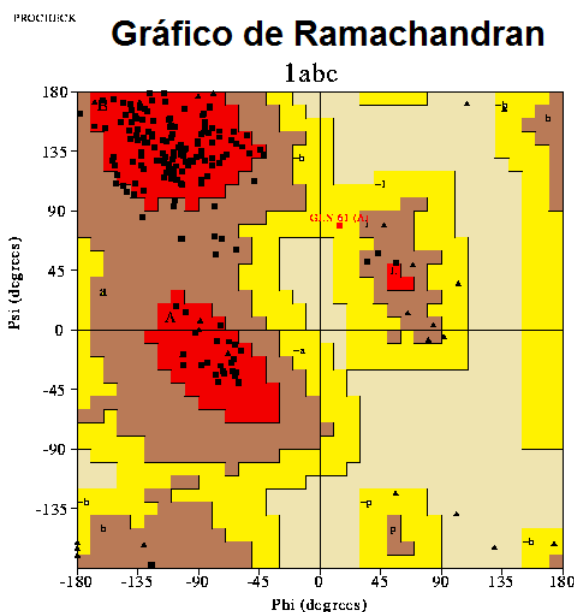
**Figura 2.** Representação dos ângulos de torção phi e psi formados entre o nitrogênio do grupamento amino e o carbono-alfa, e o carbono do grupamento carboxi e o carbono-alfa, respectivamente.

No gráfico de Ramachandran (figura 3), as regiões A, B e L representam as regiões dos valores mais favoráveis, dentre os possíveis, obtidos para os ângulos de torção (phi e psi) nas estruturas secundárias alfa-hélice, folha-beta e *loops*, respectivamente. Um modelo predito de excelente qualidade possui mais de 90% dos resíduos de aminoácidos localizados nas regiões mais favoráveis (A, B e L) (MORRIS et al., 1992).

Existem outras regiões neste gráfico que podem conter os aminoácidos, as regiões adicionais mais permissivas e as regiões permissivas. A primeira contempla os aminoácidos com valores próximos aos mais favoráveis à estrutura tridimensional da proteína, sendo, portanto, uma região secundariamente permitida. A região permissiva apresenta os valores que são menos favoráveis, diminuindo a qualidade da predição estrutural.

Há ainda no gráfico de Ramachandran uma região não permissiva, que possui os valores dos ângulos torsionais inadequados para a constituição da proteína modelada, no entanto, resíduos de glicina e prolina são exceções para esta área. Estes dois resíduos apresentam variações na cadeia lateral que conferem maior rigidez, no caso da prolina, e

maior flexibilidade, no caso da glicina, podendo assim assumir angulações não esperadas. Desta forma, são aceitos nas regiões não permissivas do gráfico de Ramachandran (RAMACHANDRAN; RAMAKRISHNAN; SASISEKHARAN, 1963 ).



**Figura 3.** Exemplo de gráfico de Ramachandran. Os aminoácidos estão representados em preto, as regiões mais favoráveis estão representadas em vermelho, as regiões adicionais mais permissivas são representadas em marrom, as regiões permissivas aparecem em amarelo e as regiões não permissivas correspondem à cor bege no gráfico acima. Fonte: figura adaptada de MORRIS et al., 1992.

A partir de modelos validados qualitativamente, outras análises *in silico* podem ser realizadas para elucidação de propriedades essenciais às funções biológicas das macromoléculas, interações com outras moléculas e dinâmica metabólica. Dentre estas análises pode-se citar: *docking* molecular (interação com outras proteínas para formação de complexos), cálculo do potencial eletrostático, carga total e polaridade da proteína.

O *docking* molecular, também conhecido como *design* racional de drogas, visa encontrar o melhor ajuste de encaixe entre duas moléculas, ou seja, consiste na predição da orientação preferencial de uma molécula com relação à outra, quando ambas formam um complexo. A orientação assumida pelas moléculas em um complexo depende da complementaridade em tamanho, forma, distribuição de cargas, polaridade e potencial de interações, e da afinidade entre as moléculas (POLANSKI; KIMMEL, 2007).

O *docking* de receptores com seus ligantes ocorre em três passos: 1. Marcação de regiões prováveis para acontecerem ligações em ambas as moléculas; 2. Sobreposição destas regiões e 3. Análise estatística e *scores* para obtenção dos melhores complexos. A estatística dos encaixes é dada pontuando-se a formação de cada complexo e, esta estimativa é feita de modo comparativo, em que cada complexo formado é avaliado com

um *score*, que ao final será posicionado em um *ranking* (SCHNEIDMAN-DUHOVNY et al., 2005). Análises de *docking* podem ser úteis para a localização de sítios ativos da proteína, locais preferenciais para ligação de drogas, medida da força de interação entre elas e como as alterações estruturais podem afetar a formação do complexo, através de análises comparativas com outras moléculas mutadas e fornecer subsídios para testes experimentais.

O potencial eletrostático é uma força não covalente, de atração entre grupos de aminoácidos de cargas opostas e, estas interações representam forças estabilizadoras da estrutura de uma proteína. Este tipo de interação ocorre quando o excesso de cargas negativas de uma região da proteína é neutralizado pelo excedente de cargas positivas de outra região, o que resulta na formação de pontes de sal entre resíduos de cargas elétricas opostas, sendo que a força de interação entre elas é maior conforme há o aumento da carga e a diminuição das distâncias entre os sítios de interação (XIONG, 2006). Diferenças no potencial eletrostático podem influenciar a eficiência de interação entre duas moléculas e, conseqüentemente, afetar a função das mesmas.

A polaridade também é uma das propriedades importantes das proteínas, advinda especificamente das cadeias laterais dos aminoácidos que as compõem. Esta propriedade caracteriza a interação com outros componentes químicos e moléculas de água (POLANSKI; KIMMEL, 2007). Cadeias laterais polares podem interagir com outras cadeias laterais polares, com a cadeia principal, ou com moléculas de água, formando ligações de hidrogênio, enquanto que as cadeias apolares somente interagem com moléculas formadas por átomos de mesma eletronegatividade (LESK, 2006).

#### **1.4 Modelo de estudo: apolipoproteína-E**

Para testar estas análises possibilitadas por ferramentas computacionais, foi escolhida uma proteína, a apolipoproteína-E relacionada à doença de Alzheimer. A escolha desta proteína como objeto de estudo se deu devido a um projeto maior em andamento no Laboratório de Genética de Populações e Evolução Molecular da Universidade Federal da Bahia, sob orientação da professora Dra. Flora Maria de Campos Fernandes, que constitui o desenvolvimento de um algoritmo para análise de doenças multifatoriais.

A doença de Alzheimer é caracterizada, principalmente, pelas perdas cognitivas e perdas de memória funcional que acometem os pacientes. Os principais achados histológicos em pacientes com a doença são a presença de placas senis, que se formam

no meio extracelular no cérebro e emaranhados neurofibrilares que se formam no interior de neurônios. As placas senis são resultado da aglomeração do peptídeo  $\beta$ -amilóide, produzido por clivagens específicas da APP, proteína precursora amilóide, por enzimas da família das secretases. Acredita-se que os efeitos das placas e emaranhados no cérebro são responsáveis pela interrupção de sinapses, neurotoxicidade, morte de neurônios e neurodegeneração, característicos da doença (GOEDERT; SPILLANNTINI, 2006).

A relação entre a progressão da doença de Alzheimer e a formação das placas senis está amplamente evidenciada (ROYCHAUDHURI et al., 2009), no entanto, alguns estudos mostram evidências que são encontrados fragmentos da apolipoproteína-E, juntamente ao peptídeo beta-amilóide, na composição das placas senis (WISNIEWSKI; FRANGIONE, 1992; NAMBA, et al., 1991).

A apolipoproteína-E possui 299 aminoácidos, sendo um polipeptídeo da família das apolipoproteínas solúveis e possui papel-chave no transporte de lipídios no plasma e no sistema nervoso central (HATTERS; PETERS-LIBEU; WEISGRABER, 2006). O gene da apoE é polimórfico e os polimorfismos de nucleotídeo único mais comuns provocam alterações na sequência codificadora nas posições 112 e 158, resultando nas isoformas: apoE2, apoE3 e apoE4, sendo que a isoforma E3 é considerada o tipo selvagem devido sua alta frequência alélica na população humana e falta de associação com patologias (MAHLEY; RALL, 2000). As variantes E2 e E4 estão associadas a diferentes patologias, sendo que a apoE2 é fortemente associada à Hiperlipidemia do tipo III, doença que acomete o metabolismo de colesterol do organismo. A variante E4 é considerada o maior fator de risco genético para o desenvolvimento da doença, podendo levar à diminuição da idade de início de manifestações clínicas (CORDER et al., 1993; STRITTMATTER et al., 1993).

As diferenças encontradas entre as isoformas ocorrem pela troca de resíduos de cisteína por arginina nas posições residuais 112 e 158, ou seja, a apoE2 possui cisteína-112 e cisteína-158, a apoE3 possui cisteína-112 e arginina-158 e a apoE4 possui arginina-112 e arginina-158. As diferenças de um ou dois aminoácidos entre as variantes da APOE mudam dramaticamente as estruturas e função da apolipoproteína-E (MAHLEY; HUANG, 2006).

Estudos demonstraram que a apolipoproteína-E está relacionada à agregação e ao *clearance* (“limpeza”) do peptídeo  $\beta$ -amilóide do meio extracelular no cérebro, sendo que

falhas nesta última função aumentam a deposição do peptídeo neurotóxico (HOLTZMAN; HERZ; BU, 2012; KIM; BASAK; HOLTZMAN, 2009). As isoformas E2 e E3 parecem executar esta função formando complexos estáveis com o peptídeo  $\beta$ -amilóide, enquanto que a E4 não, possivelmente por apresentar falhas na atividade de ligação com este peptídeo (WEISGRABER; MAHLEY, 1996). Assim, a eficiência na formação do complexo e execução das funções primordiais das isoformas é inversa ao risco de desenvolvimento da doença de Alzheimer (TOKUDA et al., 2000).

Considerando o exposto anteriormente, no presente trabalho procurou-se propor novos modelos baseados em metodologias *in silico* da apolipoproteína-E para o estudo desta proteína e suas interações moleculares. Análises empregando-se ferramentas computacionais, guiadas pela bioinformática, podem funcionar como uma triagem, no caso de estudos de proteômica, bastante específica, com grande acurácia de resultados, otimizando o volume de dados a serem testados experimentalmente. É importante informar que as moléculas atuantes na via metabólica da progressão da doença de Alzheimer foram eleitas, apenas, como modelo para tal objetivo.



## 2. OBJETIVOS

---

### **Geral**

Desenvolver modelos estruturais da apolipoproteína-E, relacionada à Doença de Alzheimer, baseados em metodologias *in silico*.

### **Específicos**

- 2.1 Comparar os modelos preditos considerando a distribuição de cargas, valores dos ângulos de torção e alterações do potencial eletrostático;
- 2.2 Analisar estruturalmente a formação dos complexos entre as isoformas modeladas da apoE e o peptídeo  $\beta$ -amilóide.

### 3. METODOLOGIA

---

#### 3.1 Sequências de aminoácidos da apolipoproteína-E

A sequência de aminoácidos da apolipoproteína-E foi obtida no banco de dados *Universal Protein Resource* – UniProt (<http://www.uniprot.org>) (The UniProt Consortium, 2012; MAGRANE, 2011). O UniProt é um produto da colaboração entre o Instituto Europeu de Bioinformática (EBI), Instituto Suíço de Bioinformática (SBI) e o Recurso de Informação de Proteínas (PIR), e contém dados abrangentes de sequências e anotações proteicas fornecidas por diferentes bancos de dados que o compõem, como: *UniProt Knowledgebase* (UniProtKB) o *UniProt Reference Clusters* (UniRef), e o *UniProt Archive* (UniParc).

O localizador da sequência da proteína de interesse no UniProt é P02649 referente à APOE\_HUMAN descrita e analisada por Rall et al. (1982). Esta sequência representa a isoforma E2 devido a presença de resíduos de cisteína nas posições 112 e 158 do polipeptídeo (figura 4).

---

> Apolipoproteína-E

KVEQAVETEPEPELRQQTEWQSGQRWELALGRFWDYLRWVQTLSEQVQEELLSSQVTQELRALMDETMKELKAYKSE  
 LEEQLTPVAEETRARLSKELQAAQARLGADMEDVCGRLVQYRGEVQAMLGQSTEELRVRLASHLRKLRKLLRDADDLQ  
 KCLAVYQAGAREGAERGLSAIRERLGPLVEQGRVRAATVGS LAGQPLQERAQAWGERLRARMEEMGSRTDRDLDEVK  
 EQVAEVRAKLEEQAQQIRLQAEAFQARLKSWFEPLVEDMQRQWAGLVEKVQAAVGTSAAPVPSDNH

---

**Figura 4.** Sequência da isoforma E2 da apolipoproteína-E disponível no banco de dados UniProt através do localizador P02649. Os resíduos de cisteína das posições 112 e 158, que caracterizam a isoforma E2, estão destacados em negrito e sublinhados.

#### 3.2 Edição das sequências

A sequência obtida a partir do UniProt foi utilizada como referência para a obtenção das outras isoformas, uma vez que elas se diferenciam apenas em duas posições residuais. A sequência referência, correspondente a isoforma E2, foi inserida em triplicata no programa BioEdit versão 7.0.5.3 (HALL, 1999), o que gerou um arquivo no formato fasta com três sequências iguais da apoE2. Estas três sequências foram editadas manualmente utilizando-se o programa BioEdit. A edição consistiu na mudança dos

resíduos das posições 112 e 158 que caracterizam cada isoforma. Para a obtenção da isoforma E3, o resíduo da posição 112 foi mantido como cisteína e houve a troca do resíduo de cisteína para arginina na posição 158. Para a obtenção da isoforma E4, os resíduos de cisteína nas posições de interesse foram substituídos por resíduos de arginina.

### **3.3 Construção dos modelos proteicos**

Para a construção dos modelos proteicos de cada isoforma da Apolipoproteína-E, foi utilizado o *software* de plataforma *online* disponibilizado pelo grupo de pesquisa de Yang Zhang do Departamento de Medicina Computacional e Bioinformática da Universidade de Michigan – EUA. O recurso utilizado para a predição da estrutura terciária da apolipoproteína-E foi o Multi-Sources Threader – MUSTER (WU; ZHANG, 2008). Os parâmetros foram mantidos segundo o *default* do programa.

### **3.4 Qualidade dos modelos preditos**

A fim de avaliar a qualidade dos modelos gerados, utilizou-se o Procheck versão 3.6.2 disponível *online* através do PDBsum-EBI (LASKOWSKI *et al.*, 1993). O Procheck permite realizar análises de qualidade do modelo com base no gráfico de Ramachandran.

### **3.5 Análise dos modelos proteicos da apolipoproteína-E**

Os modelos obtidos *in silico* foram submetidos a análises estruturais comparativas através de diferentes ferramentas computacionais, descritas a seguir:

#### PSIPRED

O PSIPRED é um programa que permite a predição da estrutura secundária de macromoléculas (JONES, 1999), disponível *online* através da *University College London*.

#### ANGLOR

O ANGLOR é um programa, em plataforma *online*, disponibilizado pelo grupo de pesquisa Yang Zhang (WU; ZHANG, 2008). Utilizando-se o ANGLOR, os valores dos ângulos phi e psi de cada resíduo, nas diferentes isoformas modeladas, foram

determinados. A predição dos valores dos ângulos torsionais foi feita para todas as posições residuais das isoformas da apolipoproteína-E. Nesta abordagem, foram analisadas as posições 61, 112, 158 e 255, as quais apresentam interações e são essenciais para a caracterização e diferenciação das variantes E2, E3 e E4.

### PatchDock

O PatchDock é um programa de plataforma *online* que permite a avaliação da interação entre duas moléculas através do *docking* molecular (DUHOVNY; NUSSINOV; WOLFSON, 2002; SCHNEIDMAN-DUHOVNY et al., 2005). As interações analisadas foram entre a apolipoproteína-E e o peptídeo  $\beta$ -amilóide, também relacionado à doença de Alzheimer. A estrutura do peptídeo  $\beta$ -amilóide, determinada experimentalmente, está disponível no banco de dados de proteínas (PDB) através do localizador 1IYT.

### RasMol, PyMol e YASARA

Para a sobreposição das estruturas terciárias, visualização dos complexos e mensuração das distâncias entre a apolipoproteína-E e o peptídeo  $\beta$ -amilóide no complexo utilizaram-se os programas YASARA (KRIEGER, 2002), RasMol versão 2.7.3 (SAYLE; MILNER-WHITE, 1995) e PyMol versão 1.3 (SCHRÖDINGER) disponível para fins estudantis.

### EMBOSS

As análises de carga e hidrofobicidade foram realizadas utilizando-se o programa EMBOSS empregando-se a ferramenta PEPSTATS, a qual realiza o cálculo das propriedades de proteínas (RICE; LONGDEN; BLEASBY, 2000).

### DNASTAR

Análises de estrutura secundária, hidrofobicidade, flexibilidade e acessibilidade foram feitas empregando-se o Protean, uma ferramenta do *software* DNASTAR (BURLAND, 1999).

### Physico-Chemical Profiles

Análises físico-químicas foram feitas nas isoformas utilizando-se a plataforma *online* NPS@ disponível pelo PBIL Lyon-Gerland, do *Pôle Bioinformatique Lyonnais* do Instituto de Biologia e Química de Proteínas – Lyon, França ([http://npsa-pbil.ibcp.fr/cgi-bin/npsa\\_automat.pl?page=/NPSA/npsa\\_pcprof.html](http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_pcprof.html)). O NPS@ é um servidor *web* interativo que realiza análises de sequências proteicas. Neste caso, foram analisadas as sequências primárias das proteínas de interesse. Nestas análises foram calculadas a flexibilidade, a acessibilidade e a hidrofobicidade das proteínas através do *physico-chemical profiles* (COMBET et al.,2000).

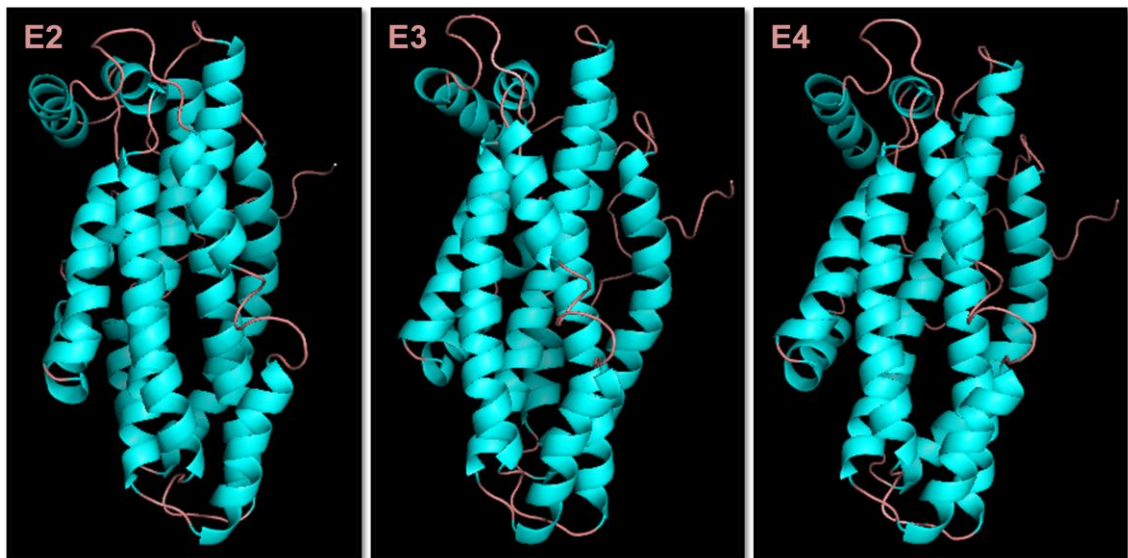
### DeepView

O programa DeepView (GUEx; PEITSCH, 1997), disponível para *download* na plataforma *online* do Swiss Bioinformatics Institute, foi utilizado para cálculo do potencial eletrostático dos modelos proteicos das isoformas da apoE.

## 4. RESULTADOS

---

Os modelos preditos pelo MUSTER, a partir das sequências de aminoácidos das três isoformas, foram gerados baseados na estrutura da apolipoproteína-E3 depositada no Banco de Dados de Proteínas (PDB - <http://www.rcsb.org>), cujo localizador é 2L7B. As estruturas preditas para as isoformas da apolipoproteína-E completa foram geradas com base em metodologias *in silico* (figura 5).

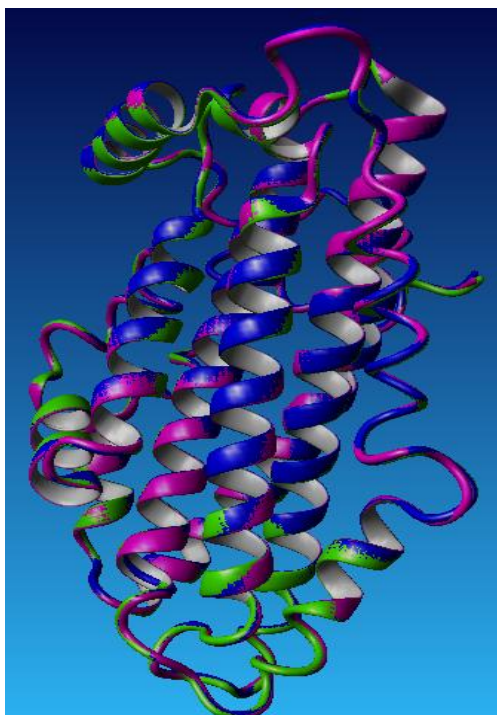


**Figura 5.** Modelos das três isoformas da apolipoproteína-E obtidos através do MUSTER e visualizados pelo PyMol. A primeira imagem representa a isoforma APOE-2, a segunda imagem representa a APOE-3 e ao final a APOE-4. Todas imagens foram obtidas com as estruturas posicionadas no mesmo ângulo para melhor visualização e comparação entre elas.

Os modelos foram preditos tanto no seu arranjo secundário, quanto no terciário. As estruturas secundárias preditas de todas as isoformas, inclusive para a estrutura disponibilizada no PDB (2L7B), através do PSIPRED, são constituídas por estruturas regulares de alfa-hélices e elementos conectivos, os *loops* ou *coils*. A parcela de alfa-hélices formadoras da estrutura secundária da proteína 2L7B representou 58% da constituição total da macromolécula, enquanto que os 42% restantes foram de *coil*. A isoforma E2 apresentou a mesma distribuição de estrutura secundária da proteína 2L7B. A isoforma E3 exibiu 60% da sua estrutura formada por alfa-hélices e 40% de *coil*. O

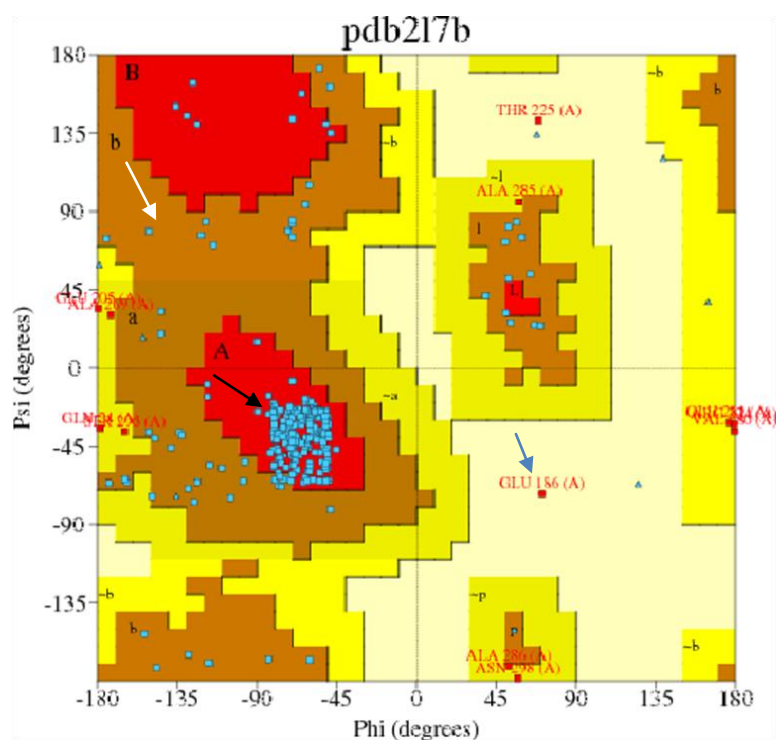
percentual de alfa-hélices para a variante E4 obteve o valor de 62%, enquanto que o valor de *coil* caiu para 38%.

Os modelos tridimensionais das proteínas, que representam a estrutura terciária, são compostos por alfa-hélices e *loops*, e visualmente não apresentam notáveis diferenças estruturais, principalmente em relação às isoformas E3 e E4. A sobreposição das três estruturas geradas *in silico* foi feita através do programa YASARA e evidencia a similaridade estrutural entre as isoformas (figura 6).



**Figura 6.** Sobreposição das três isoformas utilizando-se o *software* YASARA. Cada isoforma da apoE está representada por uma cor diferente, sendo E2 em azul, E3 em rosa e E4 em verde.

Os modelos obtidos foram submetidos à análise de qualidade através do gráfico de Ramachandran utilizando-se o *software* de plataforma *on-line* Procheck e os resultados foram comparados com a estrutura proteica 2L7B disponibilizada no PDB. A distribuição dos resíduos da proteína 2L7B pode ser vista no gráfico de Ramachandran (figura 7), e os valores para cada região do gráfico estão dispostos na tabela 1.



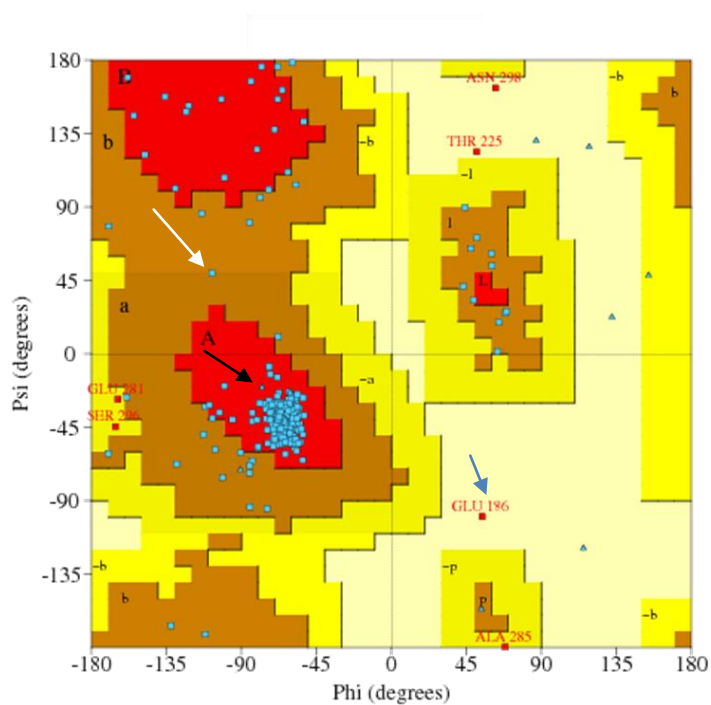
**Figura 7.** Gráfico de Ramachandran da proteína 2L7B. A seta preta indica os resíduos localizados na região mais favorável (vermelho), a seta branca corresponde aos resíduos localizados na região adicional mais permissiva (marrom) e a seta azul aponta para a região não permissiva (bege), na qual dois resíduos foram encontrados.

**Tabela 1.** Resultados do gráfico de Ramachandran da proteína 2L7B.

| Regiões                     | Siglas        | Nº de resíduos | %    |
|-----------------------------|---------------|----------------|------|
| Mais favoráveis             | [A,B,L]       | 204            | 75,0 |
| Adicionais mais permissivas | [a,b,l,p]     | 56             | 20,6 |
| Permissivas                 | [-a,-b,-l,-p] | 10             | 3,7  |
| Não permissivas             | [XX]          | 2              | 0,7  |

A distribuição dos resíduos da isoforma E2 pode ser vista no gráfico de Ramachandran (figura 8), e os valores para cada região do gráfico estão dispostos na tabela 2.



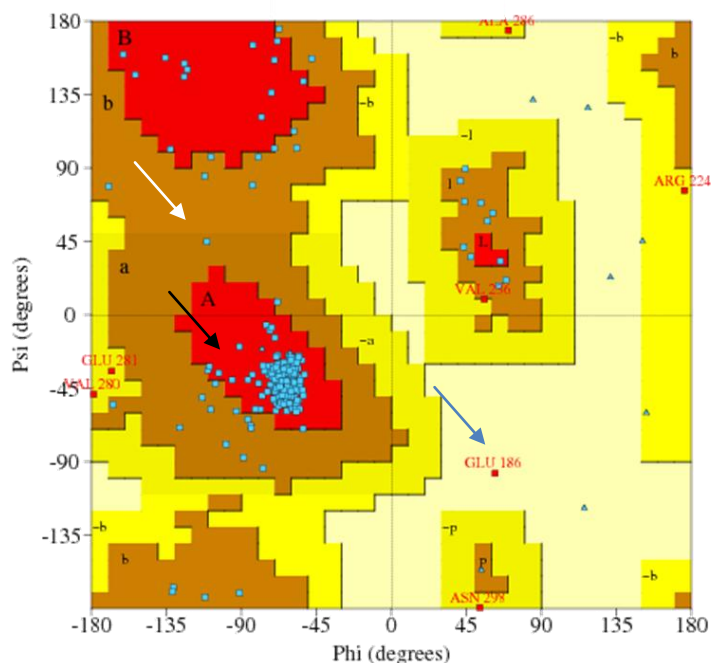


**Figura 8.** Gráfico de Ramachandran da isoforma E2. A seta preta indica os resíduos localizados na região mais favorável (vermelho), a seta branca corresponde aos resíduos localizados na região adicional mais permissiva (marrom) e a seta azul aponta para a região não permissiva (bege), na qual três resíduos foram encontrados.

**Tabela 2.** Resultados do gráfico de Ramachandran da isoforma E2.

| Regiões                     | Siglas        | Nº de resíduos | %    |
|-----------------------------|---------------|----------------|------|
| Mais favoráveis             | [A,B,L]       | 233            | 85,7 |
| Adicionais mais permissivas | [a,b,l,p]     | 33             | 12,1 |
| Permissivas                 | [~a,~b,~l,~p] | 3              | 1,1  |
| Não permissivas             | [XX]          | 3              | 1,1  |

A distribuição dos resíduos da isoforma E3 pode ser vista no gráfico de Ramachandran (figura 9), e os valores para cada região do gráfico estão dispostos na tabela 3.

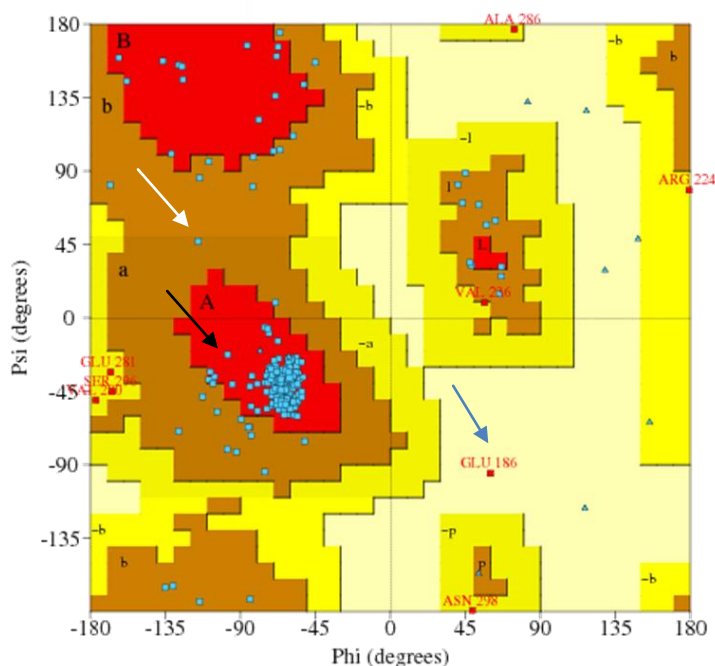


**Figura 9.** Gráfico de Ramachandran da isoforma E3. A seta preta indica os resíduos localizados na região mais favorável (vermelho), a seta branca corresponde aos resíduos localizados na região adicional mais permissiva (marrom) e a seta azul aponta para a região não permissiva (bege), na qual um resíduo foi encontrado.

**Tabela 3.** Resultados do gráfico de Ramachandran da isoforma E3.

| Regiões                     | Siglas        | Nº de resíduos | %    |
|-----------------------------|---------------|----------------|------|
| Mais favoráveis             | [A,B,L]       | 229            | 84,2 |
| Adicionais mais permissivas | [a,b,l,p]     | 36             | 13,2 |
| Permissivas                 | [~a,~b,~l,~p] | 6              | 2,2  |
| Não permissivas             | [XX]          | 1              | 0,4  |

A distribuição dos resíduos da isoforma E4 pode ser vista no gráfico de Ramachandran (figura 10), e os valores para cada região do gráfico estão dispostos na tabela 4.



**Figura 10.** Gráfico de Ramachandran da isoforma E4. A seta preta indica os resíduos localizados na região mais favorável (vermelho), a seta branca corresponde aos resíduos localizados na região adicional mais permissiva (marrom) e a seta azul aponta para a região não permissiva (bege), na qual um resíduo foi encontrado.

**Tabela 4.** Resultados do gráfico de Ramachandran da isoforma E4.

| Regiões                     | Siglas        | Nº de resíduos | %    |
|-----------------------------|---------------|----------------|------|
| Mais favoráveis             | [A,B,L]       | 229            | 84,2 |
| Adicionais mais permissivas | [a,b,l,p]     | 35             | 12,9 |
| Permissivas                 | [~a,~b,~l,~p] | 7              | 2,6  |
| Não permissivas             | [XX]          | 1              | 0,4  |

Ainda com relação aos ângulos de torção do esqueleto protéico, utilizando-se o ANGLOR foi possível obter os valores preditos para os ângulos phi e psi de cada posição dos resíduos das proteínas (tabela 5). Na posição ARG-61, a isoforma E4 apresentou o maior valor de psi = 171,3 e o menor valor de phi = -74,2. Na posição 112, a qual corresponde a uma cisteína nas isoformas E2 e E3, e a uma arginina na isoforma E4, o maior valor de psi = 59,6 e o menor valor de phi = -63,9, também corresponderam a variante E4.

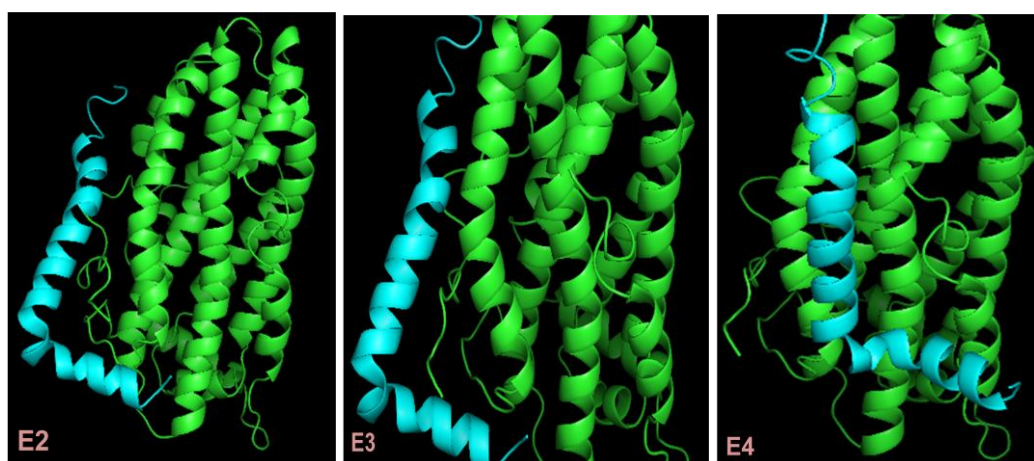
A posição 158 na isoforma E2 corresponde a uma cisteína, enquanto que nas isoformas E3 e E4 tem-se uma arginina. A isoforma E4 apresentou o maior valor de psi = 85,6 e o menor valor de phi = -65,8, em comparação às outras duas variantes, E2 e E3. O ácido glutâmico da posição 255 é o mesmo em todas as isoformas, e a isoforma E2 apresentou o maior valor de psi = 89,6 e menor valor de phi = -63,9.

**Tabela 5.** Valores dos ângulos de torção de acordo com as posições e resíduos de cada isoforma.

| ISOFORMA | POSIÇÃO | PHI   | PSI   |
|----------|---------|-------|-------|
| E2       | ARG-61  | -65,8 | 89,3  |
| E3       | ARG-61  | -65,3 | 86,2  |
| E4       | ARG-61  | -74,2 | 171,3 |
| E2       | CYS-112 | -62,8 | 57    |
| E3       | CYS-112 | -62,8 | 53,9  |
| E4       | ARG-112 | -63,9 | 59,6  |
| E2       | CYS-158 | -64,3 | 78,9  |
| E3       | ARG-158 | -64,3 | 82,6  |
| E4       | ARG-158 | -65,8 | 85,6  |
| E2       | GLU-255 | -63,9 | 89,6  |
| E3       | GLU-255 | -63,1 | 72    |
| E4       | GLU-255 | -63,1 | 82,6  |

Os modelos proteicos das isoformas da apolipoproteína-E também foram submetidos ao *docking* molecular. Através desta análise, pode-se prever as regiões de ligação com o ligante, no caso, de ligação da apoE com o peptídeo  $\beta$ -amilóide (apoE-A $\beta$ ).

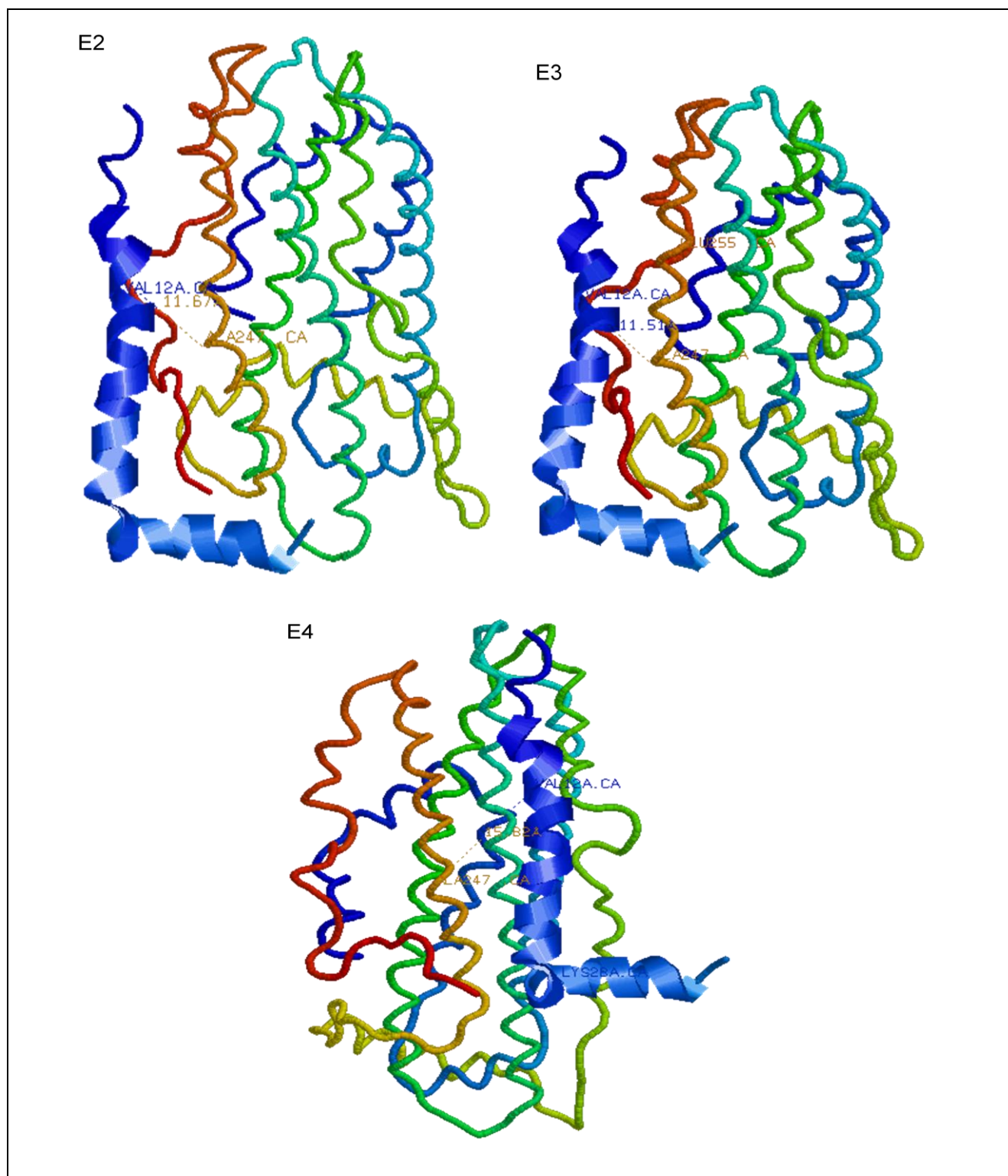
O modelo no formato pdb de cada isoforma foi carregado no PatchDock juntamente com a estrutura proteica do peptídeo  $\beta$ -amilóide (1IYT – PDB) para simular a ligação entre eles que, possivelmente, ocorre na Doença de Alzheimer (figura 11).



**Figura 11.** Complexos APOE-A $\beta$  formados com as diferentes isoformas. A estrutura da apolipoproteína-E representada em verde e a estrutura do peptídeo  $\beta$ -amilóide em azul. As imagens foram obtidas pelo PyMol e as estruturas encontram-se no mesmo ângulo de visão.

A partir da formação dos complexos pelo PatchDock, pôde-se analisar a eficiência da interação entre o sítio de ligação do peptídeo  $\beta$ -amilóide e a apolipoproteína-E pelo

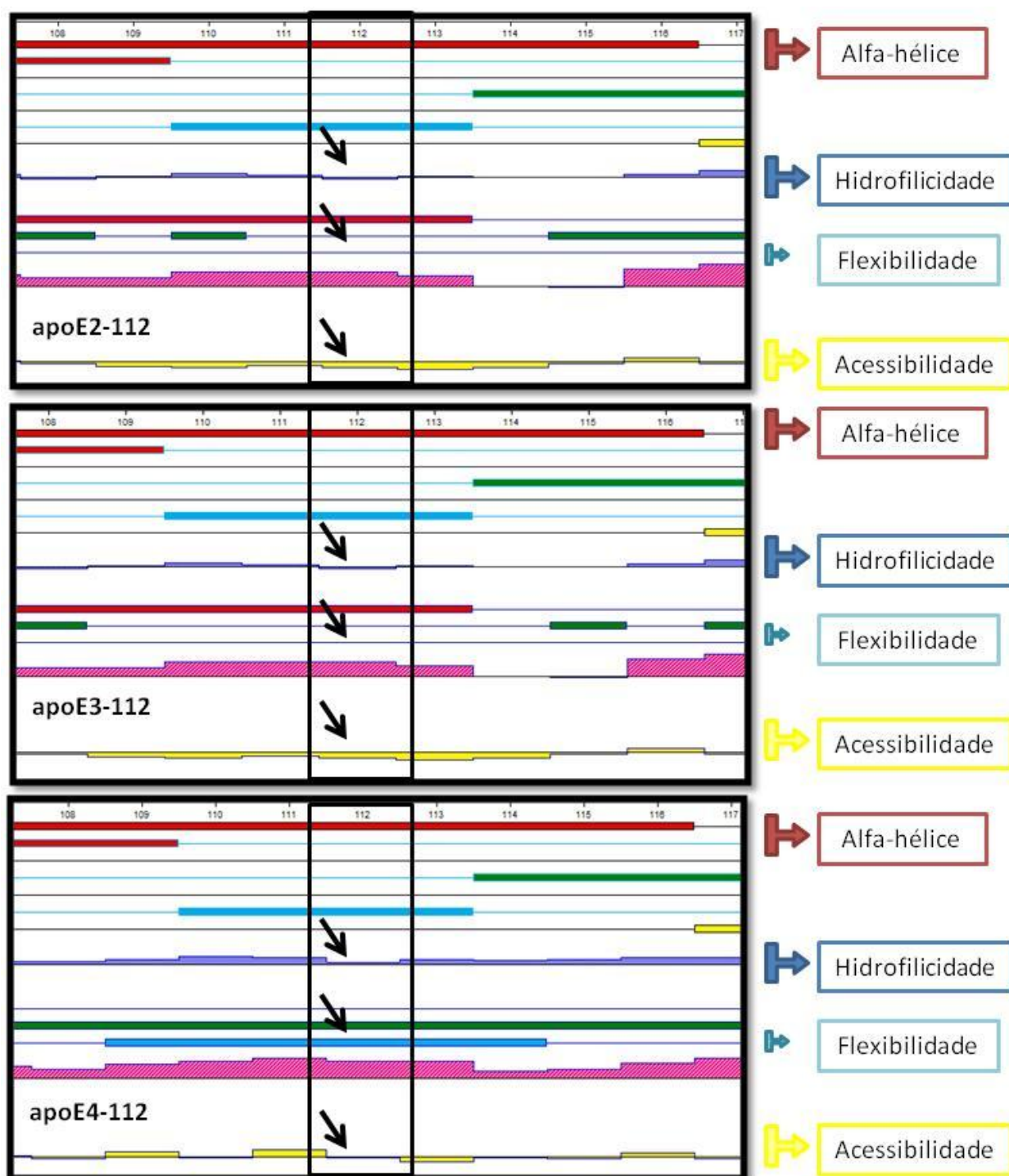
cálculo das distâncias entre as moléculas (figura 12). A distância entre os resíduos das moléculas do complexo apoE-A $\beta$  na isoforma E2 foi de 11,67 Å. Essa distância foi menor para a isoforma E3, sendo de 11,51 Å. A maior distância de interação no complexo foi calculada na isoforma E4 com o valor de 15,82 Å entre os resíduos selecionados.



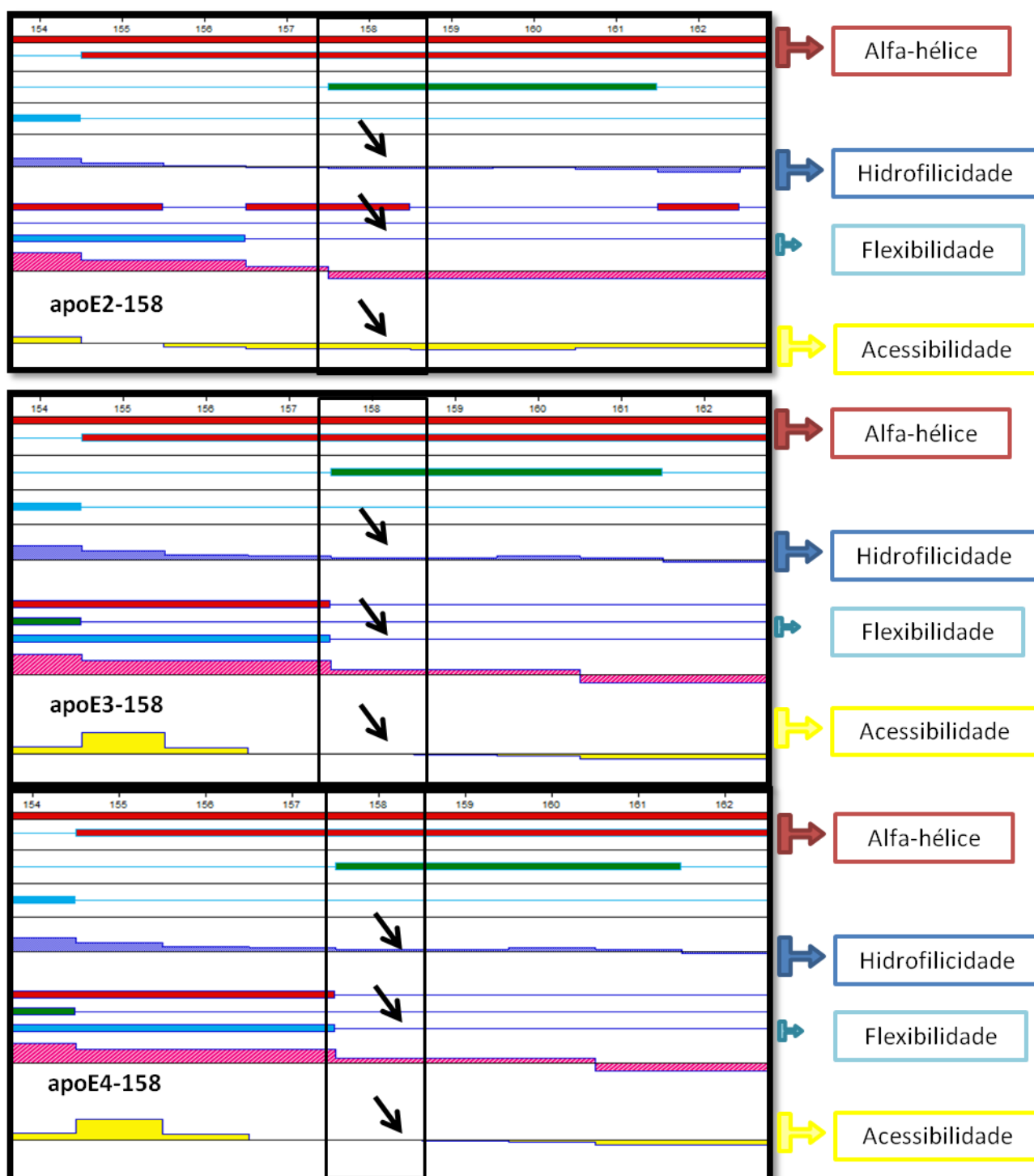
**Figura 12.** Complexos APOE-A $\beta$  e as distâncias medidas em ångstron, em cada isoforma da apolipoproteína-E. O ângulo de visão para a apoE4 difere das outras isoformas de forma a facilitar a mensuração da distância entre os resíduos.

O perfil físico-químico das estruturas foi analisado utilizando-se o *software* de plataforma *on-line* Physico-Chemical Profiles e as imagens obtidas foram a partir do Protean, uma ferramenta do DNASTAR. Os dados obtidos referem-se à flexibilidade, acessibilidade, hidrofobicidade e hidrofobicidade das diferentes isoformas com base na sua sequência de aminoácidos.

As diferenças ocorreram especificamente nas posições nas quais as isoformas se diferenciam, 112 (figura 13) e 158 (figura 14). Com relação à acessibilidade e flexibilidade, a isoforma E4 se mostrou mais acessível e flexível em ambas as posições em comparação com as isoformas E2 e E3. As isoformas E2 e E3 apresentaram maior hidrofobicidade, ou seja, a isoforma E4 obteve os maiores valores de hidrofobicidade, sendo menos hidrofóbica em relação às outras duas.



**Figura 13.** Resultado obtido pelo Protean para as isoformas E2, E3 e E4, com relação à hidrofilicidade (azul-escuro), flexibilidade (azul-claro) e superfície acessível (amarelo), calculados para a posição 112.



**Figura 14.** Resultado obtido pelo Protean para as isoformas E2, E3 e E4, com relação à hidrofilicidade (azul-escuro), flexibilidade (azul-claro) e superfície acessível (amarelo), calculados para a posição 158.

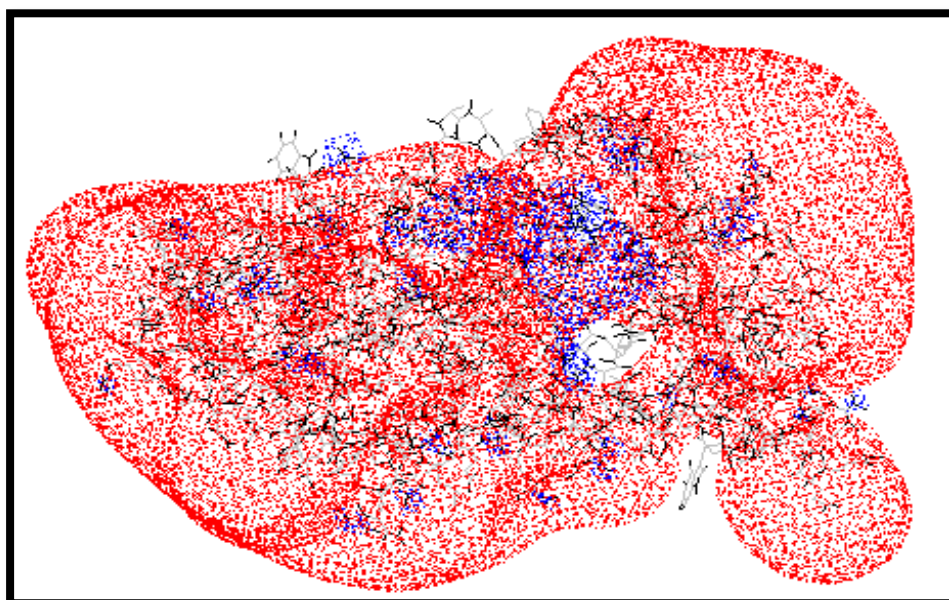
Tratando-se ainda da hidrofobicidade das isoformas, a análise do HHMOMENT, realizada através do EMBOSS, apresenta valores numéricos de hidrofobicidade baseados na estrutura primária dos polipeptídeos. A isoforma E2 apresenta o valor médio entre as



posições 112 e 158 de 0,3494 de hidrofobicidade, enquanto a E3 apresenta o valor médio de 0,3481 e a E4 possui valor médio entre as posições de interesse de 0,3407.

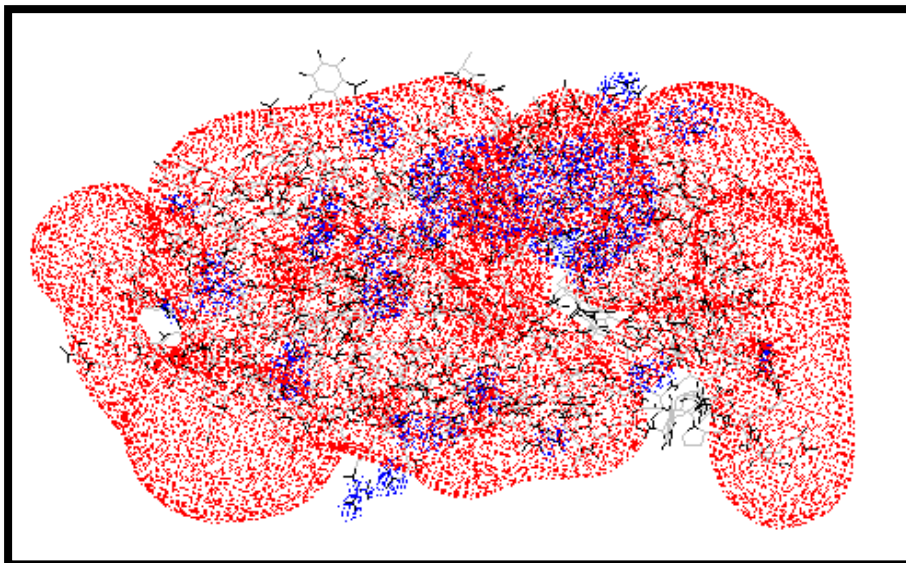
O potencial eletrostático para os modelos de cada isoforma gerados *in silico* foi avaliado utilizando-se o programa DeepView. Além do potencial eletrostático, que mostra graficamente a distribuição das cargas positivas e negativas, a carga total das proteínas foi determinada através do EMBOSS utilizando-se a ferramenta PEPSTATS.

O modelo da isoforma 2 da apolipoproteína-E (figura 15) exibiu uma nuvem densa de eletronegatividade e alguns pequenos pontos de eletropositividade, na região mais interna da molécula. A carga calculada para esta isoforma teve o valor de -5,0, sendo a mais negativa quando comparada com as outras isoformas.



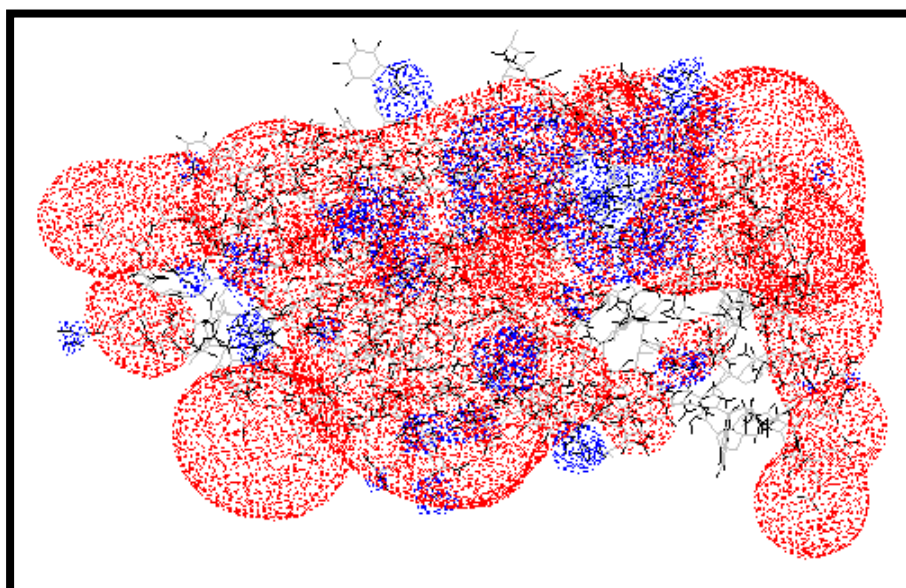
**Figura 15.** Potencial eletrostático da apolipoproteína-E, isoforma 2. A nuvem eletronegativa é representada em vermelho e a eletropositividade está representada em azul.

A isoforma E3 (figura 16) exibiu uma grande nuvem eletronegativa, porém menor que a nuvem formada na isoforma E2. Pode-se observar também que as regiões eletropositivas ficaram mais evidentes. A carga total calculada para a isoforma E3 teve o valor de -4,0.



**Figura 16.** Potencial eletrostático da apolipoproteína-E, isoforma 3. A nuvem eletronegativa é representada em vermelho e a eletropositividade está representada em azul.

A diminuição da nuvem eletronegativa é mais evidente analisando-se a isoforma E4 (figura 17). Nesta isoforma, existem regiões que deixaram de apresentar eletronegatividade e outras em que a eletropositividade ficou mais exposta. A carga total calculada para o modelo da apolipoproteína-E4 teve o valor de -3,0.



**Figura 17.** Potencial eletrostático da apolipoproteína-E, isoforma 4. A nuvem eletronegativa é representada em vermelho e a eletropositividade está representada em azul.

## 5. DISCUSSÃO

---

### 5.1 MODELOS

Existe um número muito maior de sequências proteicas descritas do que os dobramentos que as proteínas podem assumir, ou seja, do que o número de estruturas tridimensionais. Isto permite a utilização das estruturas tridimensionais, que são mais conservadas que as sequências de aminoácidos, para modelagem de macromoléculas com estruturas, parcial ou totalmente, desconhecidas (XIONG, 2006).

Análises estatísticas da estrutura secundária das isoformas da apolipoproteína-E mostraram que há, aproximadamente, em sua composição 60% de hélices e 40% de regiões de união entre elas, os *loops*. Os dados da estrutura secundária conferem maior robustez ao modelo tridimensional, uma vez que, apesar da estrutura tridimensional ser mais conservada que a estrutura primária, a estrutura secundária é mais conservada em termos evolutivos, do que a terciária.

As proteínas modeladas utilizando-se o *software* MUSTER tiveram como referência a estrutura da apolipoproteína-E3 humana, de identificador 2L7B depositada no PDB. O esqueleto protéico dos modelos foi bastante similar à estrutura depositada no banco de dados, a qual foi obtida experimentalmente por ressonância magnética nuclear (CHEN et al., 2011).

Wilson et al. (1991) determinaram o domínio N-terminal por cristalografia de raio-X, e sua composição foi descrita como um conjunto de 4 a 5 alfa-hélices principais alinhadas em sentidos opostos e unidas por regiões de *looping*, concordando com os resultados obtidos *in silico* (figura 4).

No presente estudo, utilizando-se as ferramentas computacionais, o domínio C-terminal modelado exibiu hélices unidas por regiões de *looping* na sua composição. Diferentemente da região amino terminal, a região C-terminal ainda não possui a estrutura definida por métodos experimentais (ZHONG; WEISGRABER, 2009; HATTERS; PETERS-LIBEU; WEISGRABER, 2006). Estudos dos perfís bioquímicos dos resíduos que compõem o domínio carboxi terminal e a predição da estrutura secundária propõem que esta região consiste em um conjunto de alfa-hélices anfipáticas (PILLOT et al., 1999).

Os resultados obtidos também corroboram outros estudos, levados a efeito por Segrest et al. (1990; 1992), nos quais, utilizando-se algoritmos computacionais baseados

na sequência de aminoácidos, os autores propuseram que ocorre a formação de alfa-hélices nos resíduos que compõem o domínio C-terminal, e essa região parece estar envolvida na interação proteína-proteína, ou associações da própria molécula. A predição estrutural computacional pode ser considerada uma boa ferramenta para elucidação de estruturas terciárias ainda desconhecidas, uma vez que os resultados são apoiados pelos resultados das análises da estrutura secundária e estão em concordância com métodos experimentais, considerados mais precisos.

## 5.2 RAMACHANDRAN

Os resultados das análises do gráfico de Ramachandran, utilizando-se o *software* de plataforma *on-line* Procheck, mostraram que a qualidade da estrutura depositada no PDB, cujo identificador é 2L7B, foi de 75%. Esta porcentagem refere-se aos resíduos que exibiram valores dos ângulos torsionais dentro das regiões mais favoráveis do gráfico, estando localizados principalmente na área correspondente às alfa-hélices. O valor percentual confere qualidade boa, mas não excelente, à estrutura, mesmo que esta tenha sido obtida por métodos experimentais de ressonância magnética nuclear.

Os resultados obtidos para as isoformas, cujas estruturas foram modeladas computacionalmente, se aproximaram de 90%, valor que é considerado de excelente qualidade (LASKOWSKI et al., 1993). Dos 299 aminoácidos correspondentes à proteína, cerca de 269 estiveram na região com valores de angulação mais favoráveis, e a região preferencial também correspondeu à estrutura em alfa-hélices. Isto sugere a boa qualidade de predição dos modelos, através do *software* utilizado, e maior confiabilidade dos dados obtidos *in silico*.

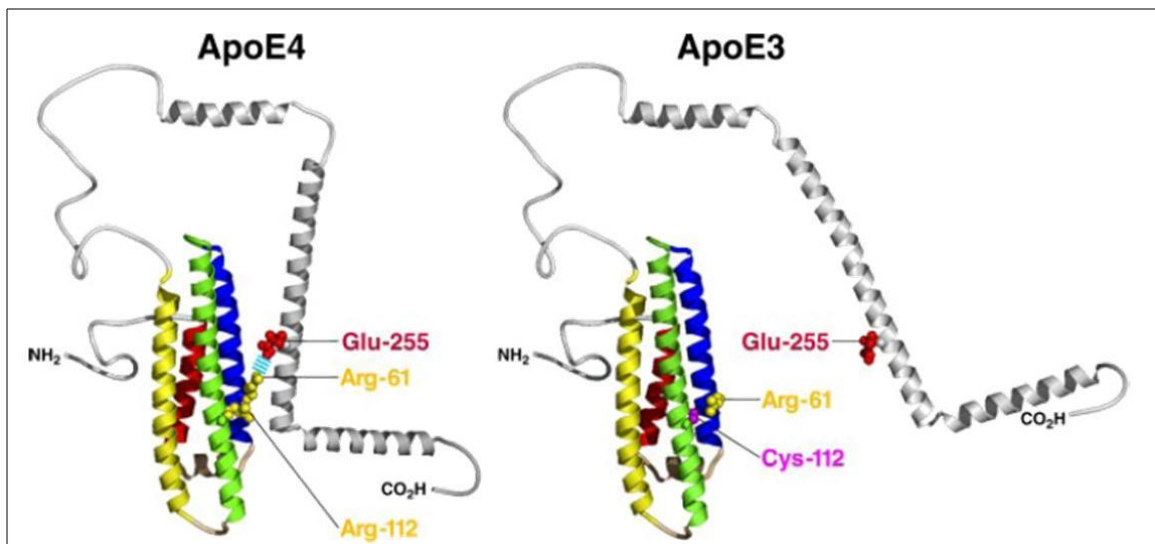
## 5.3 ÂNGULOS DE TORSÃO

Os ângulos de torção, preditos através do ANGLOR, mostraram as diferentes interações entre os domínios das proteínas, e diferenças nas conformações assumidas pelas moléculas. A interação dos domínios amino e carboxi terminais ocorre através de pontes salinas, ou interações iônicas, formadas por resíduos presentes em cada região, que, neste caso, são influenciadas pelo resíduo da posição 112. Estudos baseados em análises computacionais e experimentais mostraram que as pontes de sal podem ser

estabilizadoras (XU et al., 1997; MARQUSEE; SAUER, 1994), tornando a região fixa para o melhor encaixe funcional de moléculas (LOUNNAS; WADE, 1997), ou desestabilizadoras (SUN et al., 1991; HENDSCH; TIDOR, 1994), por provocarem modificações específicas no dobramento e função de proteínas (SUN et al., 1991).

Jones et al. (2011), através de análises da variante E4, demonstraram que esta isoforma adota uma conformação mais compactada, tornando os domínios amino e carboxi terminais mais próximos entre si, do que a conformação adotada pela variante E3. As propriedades de interação de domínio são responsáveis pela baixa estabilidade, maior ocorrência de clivagens exibidas pela apoE-4 e maior propensão a formar agregados (ZHONG; WEISGRABER, 2009). Estes fatores são sugeridos como causas para associação desta isoforma com a doença de Alzheimer, visto que a menor estabilidade da molécula pode resultar em prejuízos funcionais (MORROW et al., 2000; DONG; WEISGRABER, 1996).

O valor predito para o ângulo psi, no resíduo ARG-61 da isoforma E4, foi o maior dentre os exibidos pelas demais variantes, sendo igual a 171,3 (tabela 5). Isso indica uma maior abertura do resíduo de arginina da posição 61 (na região N-terminal), resultando na maior exposição deste resíduo e, conseqüentemente, maior aproximação para interagir com o resíduo GLU-255 (no domínio C-terminal) (figura 18).



**Figura 18.** Estruturas das isoformas E4 e E3. As posições nas quais ocorrem interações estão identificadas. Fonte: Adaptado de Hatters et al., 2006.

A troca do resíduo da posição 112 (de arginina para cisteína) na isoforma E3 resulta na atração entre a arginina da posição 61 e a cisteína da posição 112 por apresentarem cargas opostas. Esta atração resulta na diminuição do ângulo psi do resíduo ARG-61 e,

consequentemente, menor exposição para fora da hélice, que assume valor de 86,2 (tabela 5), dificultando a interação com o resíduo de ácido glutâmico presente na posição 255. Desta forma, a interação de domínio, que só ocorre na isoforma E4, afeta o arranjo conformacional desta variante. Esta interação tem sido relacionada também à produção de fragmentos neurotóxicos, devido ocorrência de clivagens preferenciais nesta variante, que podem resultar em depósitos intracelulares, assim como ocorre com os emaranhados neurofibrilares do peptídeo  $\beta$ -amilóide (BRECHT et al., 2004).

#### 5.4 COMPLEXOS apoE-A $\beta$

Outra diferença importante ocorreu na formação de complexos entre as isoformas da apolipoproteína-E e outras moléculas. Os complexos apoE-A $\beta$ , modelados pelo *software* PatchDock, apresentaram as seguintes distâncias entre os peptídeos constituintes: 11,67 Å, 11,51 Å e 15,82 Å, referentes à apolipoproteína-E2, E3 e E4, respectivamente (figura 11).

Complexos formados pelas variantes E2 e E3 com o peptídeo amilóide são mais estáveis e induzem o *clearance* (“limpeza”) do A $\beta$  do meio extracelular, impedindo a formação de agregados neurotóxicos prejudiciais (PILLOT et al., 1999). A apoE3 parece, inclusive, apresentar maior afinidade ao A $\beta$ , quando comparada à ligação apoE4-A $\beta$ , a qual não é estável e apresenta maior distância entre receptor e ligante (TOKUDA et al., 2000).

A instabilidade particular da apoE4 pode ser enfatizada durante a ligação com outras moléculas, enquanto que as isoformas estáveis E2 e E3 não são afetadas pela formação do complexo e continuam com suas funções inalteradas (LUO et al., 2010; HATTERS; PETERS-LIBEU; WEISGRABER, 2006). A maior instabilidade da isoforma E4, gerada pela interação de domínio, citada anteriormente, somada à ligação instável que ocorre na formação do complexo apoE4-A $\beta$ , sugerem que a função biológica desta proteína é prejudicada, o que eleva o risco de desenvolvimento da doença de Alzheimer. Adicionalmente, as clivagens que ocorrem na região de dobra entre os domínios da apolipoproteína-E4 liberam fragmentos de C-terminal ligados ao  $\beta$ -amilóide, que são encontrados em placas senis dos cérebros de pacientes com a patologia (HARRIS et al., 2003).

## 5.5 PERFIL FÍSICO-QUÍMICO

As análises dos perfis físico-químicos das isoformas mostraram diferenças na acessibilidade, flexibilidade, hidrofilicidade e hidrofobicidade. As diferenças entre os aminoácidos se dão nas cadeias laterais, as quais conferem a cada um suas propriedades físico-químicas (POLANSKI; KIMMEL, 2007).

A variante E4, de acordo com os valores obtidos, é a mais acessível. Corroborando com o proposto por Jones e colaboradores (2011), a isoforma E4 apresenta melhor superfície acessível à interação, apesar de ter menor afinidade com o peptídeo  $\beta$ -amilóide, que é causada pelas propriedades físico-químicas apresentadas pela variante E4.

A apolipoproteína-E4 exibiu maior flexibilidade, quando comparada com as demais. Isto significa que esta isoforma pode assumir diferentes conformações na formação de complexos com moléculas de tamanho e formas variadas, o que já foi descrito por Hatters e colaboradores em 2006.

As variantes E2 e E3 apresentaram-se mais hidrofóbicas do que a variante E4, sendo esta última a mais hidrofílica das isoformas analisadas. Os resultados obtidos pelo HHMOMENT apoiam os resultados anteriores, em que as isoformas E2 e E3 foram consideradas mais hidrofóbicas, com relação à periodicidade de estados polares e apolares dos resíduos. O efeito hidrofóbico refere-se à manutenção da solubilidade de moléculas não-polares na água. As implicações deste efeito podem ser prejudiciais para o comportamento das proteínas, ou seja, as moléculas mais hidrofóbicas se manteriam solúveis em meio aquoso, uma vez que os resíduos mais hidrofóbicos estão dispostos mais no interior da molécula e os resíduos mais hidrofílicos mais expostos para interagir com a água, enquanto aquelas com menor momento hidrofóbico teriam maior tendência a formar agregados (LESK, 2006).

Estes resultados também sugerem que a isoforma E4, por ser menos hidrofóbica, apresenta menor interação com proteínas e lipídios. Assim, conforme descrito por Tokuda e colaboradores (2000), a eficiência na ligação com o peptídeo amilóide é menor (E2>E3>>E4), o que prejudica sua função de remoção desta molécula neurotóxica, podendo inclusive passar a formar os agregados juntamente com o  $\beta$ -amilóide, aumentando a chance de desenvolvimento da doença de Alzheimer, ou até mesmo influenciando a severidade dos sintomas apresentados.

A isoforma E2, associada à doença Hiperlipidemia do tipo III, é a mais hidrofóbica dentre as três isoformas, o que permite sugerir que a isoforma E3, considerada a variante selvagem, possui hidrofobicidade intermediária, portanto não associada a patologias. A presença de resíduos de cisteína, ou resíduos de arginina, nas posições 112 e 158 está fortemente associada a patologias, enquanto que a presença de um resíduo de cisteína na posição 112 e um resíduo de arginina na posição 158, não apresenta relação com qualquer doença descrita.

## **5.6 POTENCIAL ELETROSTÁTICO E CARGAS**

O potencial eletrostático é decisivo para o dobramento específico, especificidade e afinidade da ligação proteína-proteína e função de uma macromolécula (SINHA; SMITH-GILL, 2002). Interações deste tipo interferem na associação entre moléculas e alterações no potencial podem gerar uma menor atração entre o ligante e seu receptor.

A análise do potencial eletrostático da apolipoproteína-E2 revelou a presença de uma grande nuvem eletronegativa e alguns focos de eletropositividade. Esta grande nuvem eletronegativa pode ser justificada pelos dois resíduos de cisteína a mais na isoforma E2, comparada à isoforma E4, o que resulta na carga total de -5,0. A variante E3 apresentou-se intermediária, devido à presença de um resíduo de cisteína igual à isoforma E2 e um de arginina igual à isoforma E4. A diminuição da eletronegatividade e o aumento da eletropositividade fica mais evidente com a perda dos resíduos de cisteína, e isto se reflete também na carga total, sendo de -4,0 para a E3 e -3,0 para a E4. A carga interfere no comportamento e função dos polipeptídeos, interferindo nas propriedades de aglomeração e solubilidade, desta forma, a carga menos negativa apresentada pela E4, com relação a E2 e E3, pode influenciar na interação com outras moléculas, dependendo da carga destas, e alterar sua função biológica, promovendo inclusive formação de aglomerados, que alteram o curso da doença de Alzheimer.



## 6. CONCLUSÕES

---

1. As ferramentas computacionais foram suficientes para a obtenção dos modelos proteicos das 3 isoformas da apolipoproteína-E com excelente qualidade e com estrutura similar a proteína disponível no PDB;

2. As diferenças apresentadas pelas isoformas da apolipoproteína-E ressaltaram o caráter patológico da apoE4, com relação a Doença de Alzheimer, e da apoE2, com relação a Hiperlipidemia do tipo III. Enquanto que a isoforma E3, manteve-se como tipo selvagem, por não apresentar relação com nenhuma doença descrita;

3. Análises *in silico* possibilitaram o estudo de tais propriedade de forma rápida e não dispendiosa, e os resultados obtidos estiveram de acordo com os dados de outros trabalhos realizados por técnicas experimentais;

4. Os complexos entre apoE-A $\beta$ , gerados computacionalmente, apresentaram características similares aos complexos analisados experimentalmente, o que sustenta a utilização de ferramentas computacionais para o estudo da interação intermolecular;

## REFERÊNCIAS BIBLIOGRÁFICAS

---

- BRECHT, W.J. HARRIS, F.M., CHANG, S., et al. *Neuron-specific apolipoprotein E4 proteolysis is associated with increased tau phosphorylation in brains of transgenic mice*. J. Neurosci. 24, 2527–2534; 2004.
- BURLAND, G. T. *DNASTAR's Lasergene Sequence Analysis Software*. Springer Protocols, vol. 132; pag 71-91; 1999.
- CHEN, J.; LI, Q.; WANG, J. *Topology of human apolipoprotein E3 uniquely regulates its diverse biological functions*. Proc. Nat. Acad. Sci. U. S. A. 108; 2011.
- COMBET, C.; JAMBON, M.; DELEAGE, G.; GEOURJON, C. *NPS@: Network Protein Sequence Analysis*. TIBS March Vol. 25, No 3 [291]:147-150; 2000.
- CORDER, E.H.; SAUNDERS, A.M.; STRITTMATTER, W.J.; SCHMECHEL, D.E., GASKELL, et al. *Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families*. Science 261, 921–923; 1993.
- DONG, L.M.; WEISGRABER, K.H. *Human apolipoprotein E4 domain interaction. Arginine 61 and glutamic acid 255 interact to direct the preference for very low density lipoproteins*. J. Biol. Chem. 271, 19053–19057; 1996.
- DUHOVNY, D.; NUSSINOV, R.; WOLFSON, H.J. *Efficient Unbound Docking of Rigid Molecules*. In Gusfield et al., Ed. Proceedings of the 2'nd Workshop on Algorithms in Bioinformatics(WABI) Rome, Italy, Lecture Notes in Computer Science 2452, pp. 185-200, Springer Verlag; 2002.
- GOEDERT, M.; SPILLANNTINI, M.G. *A century of Alzheimer's disease*. Science, 314, 777-781, 2006.
- GOLDING, B. G. *DNA and the revolutions of molecular evolution, computational biology, and bioinformatics*. Genome, 2003, 46(6): 930-935, 10.1139/g03-108.
- GUEX, N.; PEITSCH, M.C. *SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling*. Electrophoresis 18, 2714-2723; 1997.
- HALL, T.A. *BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT*. Nucl Acids Symp Ser 41: 95-98; 1999.
- HARRIS, F.M. ; BRECHT, W.J. ; XU, Q. ; TESSEUR, I., et al. *Carboxylterminal-truncated apolipoprotein E4 causes Alzheimer's disease-like neurodegeneration and behavioral deficits in transgenic mice*. Proc Natl Acad Sci U S A 100: 10966–10971; 2003.

- HATTERS D.M.; PETERS-LIBEU C.A.; WEISGRABER K.H. *Apolipoprotein E structure: insights into function*. Trends Biochem Sci, 31:445– 454, 2006.
- HENDSCH, Z.S.; TIDOR, B. *Do Salt Bridges Stabilize Proteins? A Continuum Electrostatic Analysis*. Protein Sci., 3, 211-226; 1994.
- HOLTZMAN, D.M.; HERZ, J.; BU, G. *Apolipoprotein E and Apolipoprotein E Receptors: Normal Biology and Roles in Alzheimer Disease*. Cold Spring Harb Perspect Med;2:a006312; 2012.
- JONES, D.T. *Protein secondary structure prediction based on position-specific scoring matrices*. J. Mol. Biol. 292: 195-202; 1999.
- JONES, P.B.; ADAMS, K.W.; ROZKALNE, A.; SPIRES-JONES, T.L., et al. *Apolipoprotein E: Isoform Specific Differences in Tertiary Structure and Interaction with Amyloid- $\beta$  in Human Alzheimer Brain*. PLoS ONE 6(1): 2011.
- KIM, J.; BASAK, J.M.; HOLTZMAN, D.M. *The role of apolipoprotein E in Alzheimer's disease*. Neuron 63, 287–303; 2009.
- KRIEGER, E.; KORAIMANN, G.; VRIEND, G. *Increasing the precision of comparative models with YASARA NOVA—a self-parameterizing force field*. Proteins 47 (3): 393–402; 2002.
- LARKIN, M.A. BLACKSHIELDS, G.; BROWN, N.P.; CHENNA, R., et al. ClustalW and ClustalX version 2.0 Bioinformatics, 23, 2947–2948; 2007.
- LASKOWSKI, R. A., MacARTHUR M. W., MOSS, D. S., THORNTON, J. M. *PROCHECK - a program to check the stereochemical quality of protein structures*. J. App. Cryst., 26, 283-291; 1993.
- LATHROP, R.; ROGRES, R. Jr.; BIENKOWOSKA, J.; BRYANT, B., et al. *Analysis and Algorithms for Protein Sequence-Structure Alignment*. Comp Methods in Molecular Biology, chapter 12, pp. 227-283, 1998.
- LEHNINGER, A. L; NELSON, D. L.; COX, M. M. *Princípios de bioquímica*. 3. ed. São Paulo: Sarvier, 838 p, 2003.
- LESK, M. A. *Introduction to bioinformatics*. 3ed. New York: Oxford. 2006
- LOUNNAS, V.; WADE, R.C. *The exceptionally stable salt-bridges in cytochrome P450cam have functional roles*. Biochemistry, 36,5402-5417; 1997.
- LUO, J.; MARECHAL, J.D.; WARMLANDER, S.; GRASLUND, A.; PERALVAREZ-MARIN, A. *In silico analysis of the apolipoprotein E and the amyloid beta peptide interaction: misfolding induced by frustration of the salt bridge network*. PLoS Comput Biol 6: 2010.

- MAGRANE, M., The UniProt consortium. *UniProt Knowledgebase: a hub of integrated protein data*. Database, 2011.
- MAHLEY R. W.; RALL S. Jr. *Apolipoprotein E: far more than a lipid transport protein*. *Annu Rev Genomics Hum Genet.*1:507-37; 2000.
- MAHLEY R.W.; HUANG Y. *Apolipoprotein (apo) E4 and Alzheimer's disease:unique conformational and biophysical properties of apoE4 can modulate neuropathology*. *Acta Neurol Scand Suppl* 185: 8–14, 2006.
- MARQUSEE, S.; SAUER, R.T. *Contributions of a hydrogen bond/salt bridge network to the stability of secondary and tertiary structure in  $\lambda$  repressor*. *Protein Sci.*, 3, 2217-2225; 1994.
- MORRIS, A.L.; McARTHUR, M.W.; HUTCHINSON, E.G.; THORNTON, J.M. Stereochemical quality of protein structure coordinates. *Proteins*, 12, 345-364; 1992
- MORROW, J.A. et al. *Differences in stability among the human apolipoprotein E isoforms determined by the amino-terminal domain*. *Biochemistry* 39, 11657–11666; 2000.
- MOUNT, D. W. *Bioinformatics*. 2. ed. Cold Spring Harbor: Paperback, 692 p; 2004.
- NAMBA, Y., TOMONAGA, M., KAWASAKI, H., OTOMO, E., IKEDA, K. *Apolipoprotein E immunoreactivity in cerebral amyloid deposits and neurofibrillary tangles in Alzheimer's disease and kuru plaque amyloid in Creutzfeldt–Jakob disease*. *Brain Res.*, 541, pp. 163–166; 1991.
- PILLOT, T.; GOETHALS, M.; NAJIB, J.; LABEUR, C., et al.  *$\beta$ -Amyloid peptide interacts specifically with the carboxy-terminal domain of human apolipoprotein E: relevance to Alzheimer's disease*. *J. Neurochem.* 72 230–237; 1999.
- POLANSKI, A; KIMMEL, M. *Bioinformatics*. New York:Springer. 376 pp; 2007.
- RALL, S.C. Jr.; WEISGRABER, K.H.; MAHLEY, R.W. *Human apolipoprotein E. The complete amino acid sequence*. *J. Biol. Chem.* 257:4171–78; 1982.
- RAMACHANDRAN, G.N.; RAMAKRISHNAN, C.; SASIEKHARAN, V. *Stereochemistry of polypeptide chain configurations*. *Journal of Molecular Biology* 7: 95–9; 1963.
- RICE, P.; LONGDEN, I.; BLEASBY, A. *EMBOSS: The European molecular biology open software suite*. *Trend in Genetics* 16, (6) pp 276-277; 2000.
- ROYCHAUDHURI, R.; YANG, M.; HOSHI, M. M.; TELOW, D. B. *Amyloid beta-protein assembly and Alzheimer disease*. *J Biol Chem*, v. 284, n. 8, p. 4749-4753; 2009.
- SAYLE, R.; MILNER-WHITE, J. E. *RasMol: Biomolecular graphics for all*. *Trends in Biochemical Sciences (TIBS)*, September, Vol. 20, No. 9, p. 374; 1995.

- SCHNEIDMAN-DUHOVNY D.; INBAR, Y.; NUSSINOV, R.; WOLFSON, H.J. *PatchDock and SymmDock: servers for rigid and symmetric docking*. Nucl. Acids. Res. 33: W363-367; 2005.
- SCHRÖDINGER, LLC. *PyMOL-The PyMOL Molecular Graphics System*, Version 1.3.
- SEGREST, J.P.; DE LOOF, H.; DOLMAN, J.G.; BROUILLETTE, C.G.; ANANTHARAMAIAH, G.M. *Amphipathic helix motif: classes and properties*. Proteins 8:103–17; 1990.
- SEGREST, J.P.; JONES, M.K.; DE LOOF, H.; BROUILLETTE, C.G.; VENKATACHALAPATHI, Y.V. ; ANANTHARAMAIAH, G.M. *The amphipathic helix in the exchangeable apolipoproteins: a review of secondary structure and function*. J Lipid Res 33:141–66; 1992.
- SINHA, N.; SMITH-GILL, S. J. *Electrostatics in protein binding and function*. Curr. Protein Pept. Sci. 3, 601–614; 2002.
- STRITTMATTER, W.J.; SAUNDERS, A.M.; SCHMECHEL D.; PERICAK-VANCE, M.; ENGHILD, J.; SALVESEN, G.S.; ROSES, A.D. *Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease*. Proc. Natl. Acad. Sci. USA 90,1977–1981; 1993.
- SUN, D.P.; SAUER, U.; NICHOLSON, H.; MATTHEWS, B.W. *Contributions of engineered surface salt bridges to the stability of T4 lysozyme determined by directed mutagenesis*. Biochemistry, 30, 7142-7153; 1991.
- The UniProt Consortium. *Reorganizing the protein space at the Universal Protein Resource (UniProt)*. Nucleic Acids Res. 40: D71-D75; 2012.
- TOKUDA, T.; CALERO, M.; MATSUBARA, E.; VIDAL, R.; KUMAR, A, et al. *Lipidation of apolipoprotein E influences its isoform-specific interaction with Alzheimer's amyloid beta peptides*. Biochem. J. 348, 359–365; 2000.
- TSOKA, S.; OUZOUNIS,C.A. *Recent developments and future directions in computational genomics*. FEBS Lett., 480,p42–48, 2000.
- WEISGRABER, K.H.; MAHLEY, R.W. *Human apolipoprotein E : the Alzheimer's disease connection*. FASEB J. 10,1485–1494; 1996.
- WILSON, C.; WARDELL, M. R.; WEISGRABERr, K. H.; MAHLEY, R. W.; AGARD, D. A. *Three-dimensional structure of the LDL receptor-binding domain of human apolipoprotein E*. Science 252, 1817–1822; 1991.

- WISNIEWSKI, T.; FRANGIONE, B. *Apolipoprotein E: a pathological chaperone protein in patients with cerebral and systemic amyloid*. *Neurosci. Lett.*, 135, pp. 235–238; 1992.
- WU, S.; ZHANG, Y. *MUSTER: Improving protein sequence profile-profile alignments by using sources of structure information*. *Proteins: Structure, Function, and Bioinformatics*.72: 547-556; 2008.
- WU, S.; ZHANG, Y. *ANGLOR: A composite machine-learning algorithm for protein backbone torsion angle prediction*. *PLoS ONE* 3; 2008.
- XIONG, J. *Essential Bioinformatics*. New York: United States of America by Cambridge University Press, 2006. 339p.
- XU, D.; LIN, S.L.; NUSSINOV, R. *Protein binding versus protein folding: The role of hydrophilic bridges in protein association*. *J. Mol. Biol.*,265, 68-84. Review; 1997.
- ZHONG, N.; WEISGRABER, K. H. *Understanding the association of apolipoprotein E4 with Alzheimer's disease: clues from its structure*. *J. Biol. Chem.* 284, 6027–6031; 2009.