

Caracterização estrutural *in silico* de proteínas recombinantes Neuraminidase-RFP

In silico structural characterization of recombinant protein Neuraminidase-RFP

Lídia P. Faustino^{1,2}, Angélica B. Cunha¹, Carlos A. Barbosa¹, Alexandre M. V. Machado², Douglas V. Pires², Andrea F. Ribeiro¹

¹ Instituto de Ciências Biológicas e da Saúde, Pontifícia Universidade Católica de Minas Gerais. Campus Betim, Rua do Rosário, 1081, Bairro Angola, CEP 32604-115, Betim, Minas Gerais. lidia.faustino@sga.pucminas.br

² Fundação Oswaldo Cruz. Instituto René Rachou, Av. Augusto de Lima, 1715, Barro Preto, CEP 30190-009, Belo Horizonte, Minas Gerais. lidia.faustino@fiocruz.minas.br

ABSTRACT: According to the World Health Organization – WHO, Influenza viruses cause respiratory infections that affect between 5 and 10% of the world population, being a public health issue. Influenza A virus belongs to the *Orthomyxoviridae* family and its surface has the glycoprotein neuraminidase which is associated to virus infection process. Considering the complexity of its structures, influenza viruses constitute an interesting platform for generation of recombinant viruses, with heterologous proteins. In this way, the present study characterized *in silico* three heterologous proteins which consisted in different portion of neuraminidase coupled to RFP. Prediction of 2D structure and the analysis of the physicochemical properties were realized. Moreover, the modeling of the 3D structures was made by threading. The results demonstrated that the conservation of different portions of neuraminidase changes physicochemical properties of the molecule such as pI, hydropathy index, extinction coefficient. As shown, these parameters also influence the dihedral angles, steric clashes, and consequently, the folding and the surface of the protein. However, more modeling studies involving energy minimization and molecular dynamics are necessary for the obtention of 3D models with better quality. Furthermore, the experimental validation has great importance to better comprehension of the functional role of the structures modeled *in silico*.

Keywords: Modeling; Physical-chemical characterization; Neuraminidase.

RESUMO: Segundo a Organização Mundial de Saúde – OMS, os vírus Influenza causam infecções respiratórias que acometem entre 5 e 10% da população mundial, sendo um problema de saúde pública. O vírus Influenza A é pertencente à família *Orthomyxoviridae* e apresenta em sua superfície a glicoproteína neuraminidase, que está associada ao processo de infecção pelo vírus. Considerando a complexidade de suas estruturas, os vírus influenza constituem uma interessante plataforma para a geração de vírus recombinantes para a expressão de proteínas heterólogas. Dessa forma, o presente estudo buscou realizar a caracterização *in silico* de três

proteínas heterólogas, que consistiam em diferentes porções da neuraminidase acopladas à RFP. Foram feitas as análises das propriedades físico-químicas e predição da estrutura 2D das mesmas. Também foi realizada a modelagem das estruturas 3D por *threading*. Os resultados demonstraram que a conservação de diferentes porções da neuraminidase altera as propriedades físico-químicas da molécula tais como pI, índice de hidropatia e coeficiente de extinção. Esses parâmetros também influenciam os ângulos diedros, as sobreposições estereoquímicas e, conseqüentemente o enovelamento da proteína e sua superfície. Contudo, mais estudos envolvendo modelagem com minimização de energia e dinâmica molecular são necessários para a obtenção de modelos 3D de melhor qualidade. Além disso, a validação experimental desses achados é de suma importância para uma melhor compreensão do papel funcional das estruturas modeladas *in silico*.

Palavras-chave: Modelagem; Caracterização Físico-química; Neuraminidase.

INTRODUÇÃO

Segundo a Organização Mundial de Saúde – OMS (2018), os vírus Influenza causam infecções respiratórias que acometem entre 5 e 10% da população mundial, sendo um problema de saúde pública. Quando pessoas incluídas em grupos de risco são infectadas, severas complicações podem surgir, inclusive a morte. Embora a eficácia e a efetividade da vacina contra o influenza confira moderada proteção, esta pode ser aumentada em populações mais jovens, ou reduzida ou ausente a depender da estação do ano e se o indivíduo for idoso (OSTERHOLM *et al.*, 2012). Nesse contexto, há a necessidade de muitos estudos voltados para o desenvolvimento de novas vacinas com maior eficácia contra esse patógeno (KUMAR *et al.*, 2018).

O vírus Influenza A é pertencente à família *Orthomyxoviridae* e se caracteriza pela polaridade negativa dos segmentos de RNA que compõem seu genoma viral. Seus subtipos são segregados conforme o tipo de glicoproteínas em sua superfície, a Hemaglutinina (HA) e a Neuraminidase (NA) (BOUVIER; PALESE, 2008). A NA é uma exosialidase (EC 3.2.1.18) cuja sequência de aminoácidos é codificada a partir do sexto segmento do RNA do vírus Influenza e compreende 470 aminoácidos. A estrutura tridimensional (3D) da NA se divide em quatro domínios: o citoplasmático, o transmembranar, a cabeça e a haste, sendo que essa última conecta a cabeça ao domínio transmembranar (BOUVIER; PALESE, 2008; NAYAK *et al.*, 2009).

Considerando a complexidade de suas estruturas, os vírus influenza constituem uma interessante plataforma para a geração de proteínas recombinantes a serem empregadas como vetores a fim de servir para a criação de vacinas contra patógenos de amplo interesse. Como exemplo, temos a utilização da proteína sinalizadora *Red Fluorescent Protein* (RFP). A estrutura da RFP consiste em um barril β e uma hélice coaxial, possuindo 225 aminoácidos. Geralmente, ela suporta a fusão N e C- terminal com variados tipos de proteínas e é conhecida por expressar em células compatíveis com a marcação fluorescente (YARBROUGH *et al.*, 2017). A recombinação com proteínas fluorescentes vem sendo realizada há aproximadamente 20 anos agregando efetividade às técnicas de expressão *in vitro* (WALSH *et al.*, 2000; EKATERINA *et al.*, 2015) e melhorando as análises voltadas para a produção de novas vacinas.

Considerando as análises *in silico*, é possível obter informações da estrutura da proteína (i.e., primária, secundária e terciária) por meio da utilização de recursos computacionais. Tais técnicas permitem avaliar as características físico-químicas e seus possíveis efeitos na estrutura (secundária e terciária) e função da molécula, obter modelos 2D (bidimensional) e 3D da estrutura da mesma, visando reconhecer regiões de maior ou menor interesse biológico. Dentre as técnicas de modelagem 3D temos por homologia, *threading* e *ab initio*. O princípio da modelagem por homologia fundamenta-se na teoria da evolução, onde proteínas com a mesma ancestralidade tenderiam a apresentar enovelamentos similares. Em outras palavras, sequências similares apresentam estruturas semelhantes mesmo que o ancestral comum seja distante (CHOTHIA; LESK, 1986). Essa é uma técnica mais rigorosa quanto a análise do enovelamento (i.e., *folding*) e só pode ser utilizada entre proteínas com alta similaridade estrutural. A modelagem por *threading* é menos restrigente permitindo que estruturas de proteínas com baixa similaridade possam ser preditas. Por fim, a técnica por *ab initio* é válida para proteínas pequenas sem moldes disponíveis, onde por meio de simulação de Monte Carlo os possíveis enovelamentos são preditos (ZHANG; SKOLNICK, 2003). Os moldes utilizados nas análises de modelagem estrutural *in silico* são obtidos por meio das técnicas experimentais de cristalografia por difração de raio-x, ressonância nuclear magnética ou microscopia eletrônica 3D, cuja precisão contempla os critérios de identidade e cobertura (WEBB; SALI, 2017), permitindo assim conhecer ou resolver a estrutura 3D das proteínas. Essas estruturas são depositadas em bancos de dados públicos como RCSB PDB (*Research Collaboratory for Structural Bioinformatics Protein Data Bank* – <http://rcsb.org>).

Desse modo, com o objetivo de caracterizar as estruturas de proteínas recombinantes fusionadas às porções conservadas da haste da NA, foram utilizados recursos computacionais

que fornecem informações acerca dos aminoácidos bem como as suas concentrações nas proteínas de interesse. Como a região da NA até o resíduo 84 ainda não foi cristalizada (BOSSART-WHITAKER et al., 1993; WAN et al., 2015), a modelagem foi feita por *threading* e *ab initio*. Tal abordagem contribui para o conhecimento acerca da natureza química da proteína e pode servir de base para futuras análises *in silico*, como minimização de energia e dinâmica molecular. Além disso, futuramente, podem ser empregadas no planejamento de experimentos voltados para o desenvolvimento de vírus recombinantes como vetores de antígenos heterólogos e, posterior aplicação na geração de vacinas.

MATERIAIS E MÉTODOS

Sequências proteicas: Foram utilizadas três sequências proteicas que são provenientes da tradução *in silico* de plasmídeos que codificam um segmento truncado da NA, os quais podem possuir os primeiros 50, 100 ou 150 nucleotídeos codificantes da região 3' do segmento da NA (*strain A/WS/1933 H1N1- UniprotKB – P03470*), seguido pela inserção da RFP (*PAmCherry1 protein - UniprotKB – D1MPT3*), sendo denominadas Neu1 (246 aminoácidos), Neu2 (263 aminoácidos) e Neu3 (279 aminoácidos), respectivamente. Essas sequências de DNA recombinante foram traduzidas no *software Mega 7* (<https://www.megasoftware.net/>) (KUMAR; STECHER; TAMURA, 2016). Tais sequências foram gentilmente cedidas pelo Prof. Dr. Alexandre de Magalhães Vieira Machado, também autor do presente trabalho.

Caracterização físico-química da estrutura primária das proteínas recombinantes: Para mensurar o número absoluto e percentual (%) dos aminoácidos e determinar suas propriedades físico-químicas foram utilizados o *ProtParam Tool* (<http://web.expasy.org/protparam>) e o *ProtScale* (<http://web.expasy.org/protscale>), ferramentas *online* gratuitas e disponíveis no Portal de recursos de Bioinformática ExPASy do Instituto Suíço de Bioinformática (*Swiss Institute of Bioinformatics*) (ARTIMO et al., 2012; WALKER, 2005).

Predição da estrutura secundária das proteínas recombinantes: Para prever os possíveis enovelamentos das sequências de interesse utilizamos o *PSIPRED (PSI-BLAST based secondary structure PREDiction –* <http://bioinf.cs.ucl.ac.uk/psipred>), um servidor *online* para proteínas recombinantes (McGUFFIN; BRYSON; JONES, 2000). Os arquivos de entrada com sequências proteicas recombinantes foram inseridos no formato FASTA e os arquivos de saída no formato gráfico e no formato texto continham as informações da estrutura.

Identificação de regiões de desordem da estrutura primária das proteínas recombinantes: O software online PONDR – Predição de Regiões Naturalmente Desordenadas – foi utilizado para a identificação das regiões de desordem nas proteínas recombinantes (BOMMA et al., 2012). A predição é realizada por meio de três redes neurais estruturadas que avaliam a desordem em regiões longas, amino e carboxi terminais, alimentada por dados cristalográficos obtidos por difração de raios-X.

Modelagem da estrutura 3D das proteínas recombinantes: As estruturas foram obtidas pelo software online e gratuito I-TASSER (YANG; ZHANG, 2015). A partir da sequência primária no formato FASTA, o software identifica os possíveis modelos 3D no formato .pdb. O método *threading* por iteração usado é feito por meio do alinhamento perfil-perfil (*profile-profile alignment*, PPA), onde a avaliação da similaridade das sequências próximas a de interesse também é considerada na modelagem e as regiões sem similaridade são modeladas por *ab initio*. Dessa forma, é possível prever o enovelamento de proteínas pequenas cujas sequências não têm uma relação tão próximas. O arquivo de saída gerado é no formato .pdb.

Avaliação dos modelos 3D das proteínas recombinantes: Para validar o modelo 3D das proteínas foi utilizado o PROCHECK (LASKOWSKI et al., 1993). O programa verifica a estereoquímica de cada aminoácido na estrutura do modelo alvo comparada às estruturas resolvidas experimentalmente na mesma resolução (Å) e ainda destaca quais regiões não estão adequadas (LASKOWSKI et al., 1993).

Visualização dos modelos 3D das proteínas recombinantes: Os modelos gerados foram visualizados com o *Pymol*, um sistema gráfico capaz de gerar imagens com alta definição das estruturas de macromoléculas (DELANO, 2002).

RESULTADOS E DISCUSSÃO

Caracterização físico-química da estrutura primária das proteínas recombinantes

O ponto isoelétrico (pI) diferiu as proteínas recombinantes Neu1 e Neu2 da Neu3 indicando que seja possível separá-las experimentalmente (TAB. 02C). Conforme esperado, a massa molecular aumenta conforme o número de resíduos da proteína, ou seja, a massa de Neu1 é a menor, a de Neu2 é intermediária e a de Neu3 é a maior (TAB. 01C e 02C). Neu1 e Neu2 apresentam a mesma quantidade dos aminoácidos glutamato, aspartato, cisteína, tirosina,

histidina, lisina e arginina, sendo estes diferentes na Neu3. A depender da fração da NA nativa, o *pI* varia entre 5,2 a 6,5 (KENDAL et al., 1973).

O Índice Alifático (IA) é uma medida de termoestabilidade da proteína e se correlaciona com o volume ocupado pelas cadeias laterais alifáticas encontradas nos resíduos de alanina, valina, isoleucina e leucina. Esses grupos tendem a se agrupar no interior da molécula por meio de interações hidrofóbicas. Normalmente, quanto maior o valor desse índice maior a termoestabilidade dessa proteína (AMBLEE; JEFFERY, 2015; GASTEIGER et al., 2005). Para a Neu1 e Neu2 foi observada maior presença de resíduos alifáticos em sua sequência (22, 4% e 24,8%, respectivamente) e, portanto, maior IA (66,54 e 77,07, respectivamente). É importante ressaltar que existem outros fatores que interferem com a termoestabilidade das proteínas, como as pontes de hidrogênio e/ou íons de metais (IKAI, 1980), os quais não foram estimados nessa análise.

Todas as proteínas se mostraram estáveis, ou seja, com Índice de Estabilidade indicando uma meia-vida *in vivo* maior que 16 horas ($IE < 40$), conforme a classificação proposta por Rogers e cols. (1986), (TAB. 02C). De acordo com Guruprasad e cols. (1990) a estabilidade das proteínas sofre influência em grande parte pela presença dos aminoácidos metionina, glutamina, prolina, glutamato e serina, sendo estes encontrados em menor quantidade na Neu2 (26,9%), seguido da Neu1 (28,6%) e Neu3 (28,95%). Dessa forma, quanto menor a concentração desses aminoácidos na proteína maior seria a estabilidade da mesma. Vale lembrar que a meia-vida das proteínas também sofre influência de outros parâmetros não computados nessa análise, como os envolvidos na resposta desencadeada no organismo frente a presença de elementos exógenos.

A análise do Coeficiente de Extinção Molar (CEM) indica o comportamento ótico da molécula traduzido pela capacidade dos aminoácidos aromáticos (triptofano, tirosina e fenilalanina) de absorver luz ultravioleta (GILL; HIPPEL, 1989; GASTEIGER et al., 2005). Não foi observada diferença entre o CEM da forma oxidada ou não da cisteína para as proteínas Neu1 e Neu2 (TAB 02C). Já para a Neu3 foi observado maior CEM na presença de cisteína oxidada indicando a influência das ligações dissulfeto para absorção de luz nessa proteína (TAB. 02C). Com base nesse resultado seria possível analisar experimentalmente os espectros de absorbância das proteínas em questão, quer seja para a identificação ou quantificação das mesmas, por exemplo.

A hidropatia foi avaliada por meio da Grande Média do Valor de Hidropatia (GRAVY), caracterizada pela razão entre a hidropatia dos aminoácidos e o número total de resíduos da

sequência. Esse índice revela o potencial dos resíduos para interagir com os constituintes de regiões membranares ou então para estabilizar a proteína por meio de interações entre seus domínios (KYTE; DOOLITTLE, 1982). Foi observado que a Neu1 apresenta menor hidropatia comparada a Neu2 e Neu3 (FIG. 01C e TAB. 02C).

Predição da estrutura secundária e identificação de regiões transmembranares e de desordem das proteínas recombinantes

Os resíduos da estrutura selvagem da NA (P03470) estão divididos em domínio citoplasmático (1-6), domínio transmembranar - Hélice (7-27) e superfície viral – Haste e porção globular (28-453), (UniProtKB). Os resultados sugerem que os primeiros resíduos das sequências (1-6) não formam estruturas secundárias (FIG. 02C, 04C e 06C) e também não formam domínios transmembranares (FIG. 08C). Por outro lado, os resíduos que seguem essa região tendem a formar uma folha, sucedida de uma hélice, exceto na sequência Neu1, que se finda nessa região. A Neu3 (FIG. 06C) apresenta uma hélice menor quando comparada a Neu2 (FIG. 04C). Os resíduos envolvidos na formação da hélice nas proteínas recombinantes estão em posições diferentes quando comparados à NA selvagem. Isso sugere que a presença dos grupos carboxílicos, carregados negativamente, dos resíduos de glutamato (E) traduzidos a partir do sítio de clonagem inserido entre a NA e a RFP podem interferir na formação da hélice.

Na região próxima a fusão da RFP à NA foi observada a presença de desordenamento e reduzida confiabilidade na predição da estrutura secundária em todas as proteínas recombinantes estudadas (região de desordem 1, FIG. 03C, 05C e 07C). Segundo Bomma et al. (2012), regiões de desordem são caracterizadas por não possuírem uma conformação característica (i.e., hélice ou folha), mas podem exercer importantes funções como reconhecimento e ativação na proteína, ou também aumentarem a hidropatia, flexibilidade e formação de folhas beta (UVERSKY, 2011). Os aminoácidos normalmente encontrados nessas regiões são alanina, arginina, glicina, glutamina, serina, glutamato, lisina e prolina (UVERSKY, 2011), os quais também foram observados nas regiões desordenadas das proteínas estudadas, bem como, folhas beta (FIG. 02C, 04C e 06C) e hidropatia elevada (FIG. 01C), quando comparadas a outras porções das sequências das mesmas. As demais regiões de desordem (2 e 3) observadas nas estruturas das proteínas recombinantes também se localizavam na RFP, porém sem impacto na predição da estrutura secundária nesses locais. A Neu1 apresenta menos resíduos transmembranares comparada à Neu2 e Neu3 (FIG. 08C). Esse dado está de

acordo com os índices de hidropatia (FIG. 01C) e hidrofobicidade (TAB. 02C) observados anteriormente.

Juntos, os resultados demonstraram que não somente a RFP exerce influência no enovelamento das porções conservadas da NA, como essas porções também interferem na sua estrutura secundária. Contudo, foi observado que na Neu1 a estrutura da RFP não sofreu modificações quando comparada a sua predição individual (dados não mostrados). Por outro lado, na Neu2 (FIG. 04C) duas folhas menores, a primeira formada por Val175 e Met176 e a segunda formada por G180 e W181 foram substituídas por uma única alça (168 ao 190 resíduo). Além disso, na Neu3 a primeira folha após a hélice está constituída de um resíduo a mais (Gln40) e uma das folhas menores, formadas por Val189 e Met190 também foi substituída pela alça. É importante observar que a última alteração está presente em um local de desordem (FIG. 07C).

Modelos 3D

A NA não foi enovelada conforme a estrutura secundária predita (FIG. 01). Por outro lado, a estrutura 3D da RFP apresentou poucas modificações quando comparada à sua estrutura secundária. Deve ser lembrado que a predição da estrutura secundária é fundamentada apenas na estrutura primária, enquanto que a modelagem 3D também tem como base as coordenadas atômicas e os ângulos diedros (WEBB; SALI, 2017).

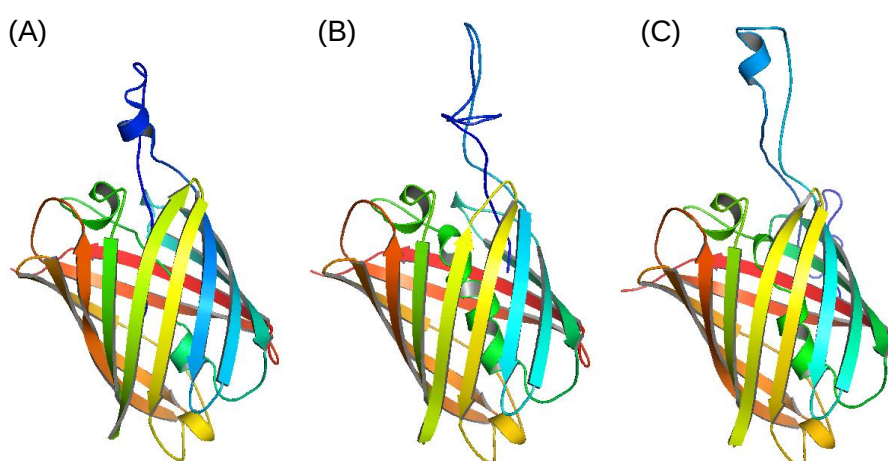


FIGURA 01 – MODELOS GERADOS DAS PROTEÍNAS RECOMBINANTES

Nota: As imagens dos modelos 3D foram geradas no Pymol – Delano, 2002. Onde A: Neu1, B: Neu2 e C: Neu3.

A porção inicial da NA apresenta cargas predominantemente neutras em sua superfície (FIG. 09C). Como a Neu3 é a única que possui resíduos o suficiente para atravessar a membrana, a sua porção final se mostra carregada negativamente. Os pontos de sítio de clonagem (fusão da NA e RFP) também apresentaram cargas negativas em sua superfície em função da presença dos resíduos de glutamato. Além disso, as superfícies carregadas negativamente das proteínas recombinantes se localizam nas mesmas regiões de desordem (FIG. 03C, 05C e 07C) e podem influenciar na conformação da estrutura 3D proteica. A Neu1 apresenta menor quantidade de resíduos hidrofóbicos e o inverso é observado para a proteína Neu2 (FIG. 10C). Como foi demonstrado anteriormente a Neu1 possui menor hidropaticidade, índice de estabilidade e região transmembranar, quando comparada a Neu2, indicando coerência com os dados obtidos no modelo 3D. Também deve ser ressaltado que há maior concentração de resíduos hidrofóbicos na porção de NA correspondente a região transmembranar (FIG. 08C) sendo esta conservada nas proteínas recombinantes. Young e cols. (1994) descreve que a superfície hidrofóbica está relacionada com as interações inter e intramoleculares da proteína, contudo, não foi realizada a análise da interação entre as estruturas 3D obtidas e as moléculas de água do meio, sendo essa uma etapa importante para a compreensão dessas interações. Em conjunto a compreensão das características eletrostáticas e hidrofóbicas pode ser aplicada na elucidação do papel biológico da molécula bem como na identificação de sítios de ligação que podem ter importância no desenvolvimento de novas terapias (NICOLAU *et al.*, 2014; VLACHAKIS *et al.*, 2014).

Validação dos modelos 3D

O diagrama de Ramachandran é utilizado para avaliar a qualidade das estruturas 3D por meio dos ângulos diedros ϕ (phi) e ψ (psi), que determinam a rotação da ligação entre o carbono alfa e o nitrogênio ($C_{\alpha} - N$) e a ligação entre o carbono alfa e o carbono vizinho ($C_{\alpha} - C$), respectivamente (RAMACHANDRAN; RAMAKRISHNAN; SAAISEKHARAN, 1961).

A Neu1 apresenta os resíduos Val16, Glu19, Met22, Asp136, Ser152, Asp175 em regiões de desordem (FIG. 11C). Os três primeiros resíduos citados têm carga e são polares (FIG. 03C), podendo influenciar na conformação da proteína (UVERSKY, 2011). O resíduo de Asp175 pode ser um possível local sob influência do grupo planar de resíduos adjacentes (Tyr172 – 0,038 Å e Glu174 – 0,029 Å), (FIG. 14C). O mesmo se observa para a Ser152 (FIG. 14C) devido a presença de um grupo planar do resíduo vizinho Asp153 (0,020 Å). A ligação covalente do átomo de carbono varia de acordo com os ângulos de torção (ϕ e ψ) e

consequentemente pode impactar a geometria da estrutura 3D da proteína (BERKHOLZ et al., 2009). Esses dados mais a hidrofobicidade dessa região (FIG. 01C) também podem contribuir para a geração de forças de repulsão e de *Van der Waals* (pontes de hidrogênio e dissulfeto) com possível impacto no arranjo da proteína (RAMACHANDRAN et al., 2011).

A Neu2 apresenta os resíduos Ile7, Met15, Val17, Gln25, Arg74, Asp153 e Thr261 em regiões desfavoráveis (FIG. 12C). sendo que os quatro primeiros resíduos podem sofrer alterações geométricas devido às distorções nas ligações e ângulos interatômicos e intermoleculares (FIG. 15C e 16C). A Ile7 tem a ligação entre o C α e C β (1,602 Å) maior que o valor ideal (1,540 Å), o que pode interferir no ângulo entre as ligações C β – C α – C (130,0 Å) e N – C α – C (100,5 Å), que está 20,9 Å acima e 10,7 Å abaixo do ideal, respectivamente. Do mesmo modo, o aumento na distância entre a ligação do C α e C (158,2 Å) da Met15 também refletiu no aumento do ângulo entre C β – C α – C (129,8 Å). Assim como o aumento na ligação interatômica C α e C (1,600 Å) e N e C (1,524 Å) refletem no ângulo da ligação entre N – C α – C (126,7 Å). Igualmente, as ligações entre C α e C (1,589 Å) e N e C α (1,510 Å) na Gln25 alteraram o ângulo da ligação N – C α – C (126,7 Å). Ainda, nessa região, a ligação entre N e C α (1,510 Å) mudam o ângulo da ligação entre N – C α – C (124,7 Å). Por outro lado, a Asp153, aparentemente, sofreu alterações resultantes das distâncias interatômicas e angulares dos resíduos vizinhos (Gln 152 e Gly154). Também é possível observar que a Neu2 apresenta mais distorções angulares quando comparada a Neu1 e Neu3, demonstrando que essa proteína está mais susceptível aos efeitos de atração e repulsão resultantes de forças de *Van der Waals*.

Para a Neu3 (FIG. 13C) foi observado que os resíduos Ile13, Asp208, Asp223, Ser257 e His258 estão em regiões não favoráveis. O primeiro resíduo apresenta alteração no ângulo da ligação C β – C α – C (123,2 Å), assim como o resíduo vizinho, Ser12 (FIG. 17C). A Asp223 não está inserida em uma região de desordem, mas, está próxima a resíduos carregados como Lys221 e Glu230. Ainda a Asp208 sofre influência da distorção do grupo planar do resíduo Glu207 (0,022 Å), (FIG. 18C).

A Neu2 possui 73,4% dos resíduos em regiões energeticamente favoráveis, seguida por Neu1 com 71,5% e Neu3 com 66,1% e, Neu2 apresenta 3,2%, Neu1 2,9% e Neu3 2,1% de resíduos em regiões energeticamente desfavoráveis (FIG. 11C, 12C e 13C). Dessa forma, novas análises de modelagem deveriam ser realizadas, buscando-se maior refinamento dos ângulos no modelo gerado, por meio da utilização de minimização de energia, seguida de dinâmica molecular. Juntas, essas estratégias são ferramentas uteis que podem melhorar a qualidade

estereoquímica das estruturas recombinantes, diminuindo os “maus contatos” que não satisfazem as restrições espaciais.

CONCLUSÃO

Foi possível obter as propriedades físico-químicas, além das características estruturais secundárias e terciárias das proteínas recombinantes estudadas. As análises revelaram que a conservação de diferentes porções da neuraminidase altera as propriedades físico-químicas da proteína recombinante tais como pI, índice de hidropatia e coeficiente de extinção. Esses parâmetros também impactam os ângulos de torção, estereoquímica, superfície eletrostática e consequentemente o enovelamento das proteínas recombinantes. A qualidade do modelo 3D foi abaixo do esperado, se fazendo necessário mais estudos envolvendo modelagem com minimização de energia e dinâmica molecular. Também é necessária a validação experimental para uma melhor compreensão do impacto desses achados sobre a função das proteínas recombinantes. Tais resultados podem ser utilizados em futuras análises de expressão proteica, bem como em experimentos voltados para a caracterização experimental das estruturas. Por fim, futuramente o conhecimento gerado pode ser aplicado no desenvolvimento de novas terapias.

Perspectivas

Desenvolver estudos de refinamento nos modelos, como dinâmica molecular e minimização de energia para melhorar a compreensão dos aspectos tridimensionais e aplicar o protocolo em outras construções recombinantes. Além disso, fazer a validação experimental das análises computacionais.

Agradecimentos

Os autores agradecem as instituições de ensino Pontifícia Universidade Católica de Minas Gerais (PUC Minas) e Instituto de Pesquisa René Rachou – Fiocruz Minas (IRR) e a colaboração de Amanda Tabita da Silva Albanaz e M.e. Ana Paula de Faria Gonçalves, alunas do IRR.

REFERÊNCIAS

- AMBLEE, V.; JEFFERY, C. Physical features of intracellular proteins that moonlight on the cell surface. **Plos One**. n. 10. 2015.
- ARTIMO, P.; JONNALAGEDDA, M.; ARNOLD, K.; BARATIN, D.; CSARDI, G.; DE CASTRO, E.; DUVAUD, S.; FLEGEL, V.; FORTIER, A.; GASTEIGER, E.; GROSDIDIER, A.; HERNANDEZ, C.; IOANNIDIS, V.; KUZNETSOV, D.; LIECHTI, R.; MORETTI, S.; MOSTAGUIR, K.; REDASCHI, N.; ROSSIER, G.; XENARIOS, I.; STOCKINGER, H. Expasy: SIB bioinformatics resource portal. **Nucleic Acid Res**. n. 40. 2012.
- BERKHOLZ, D.S.; SHAPOVALOV, M.V.; DUNBRACK, R.L.J.; KARPLUS, P.A. Conformation Dependence of backbone geometry in proteins. **Structure**. vol. 17 n. 10 p. 1316-1325. 2009.
- BERMAN, H.M.; WESTBROOK, J.; FENG, Z.; GILLILAND, G.; BHAT, T.N.; WEISSIG, H.; SHINDYALOV, I.N.; BOURNE, P.E. The Protein Data Bank. **Oxford University Press**. vol. 1 n. 28 p. 235-242. 2000.
- BOMMA, R.; VENKATESH, P.; ANIL KUMAR, D.L.V.N.S.S.S.R.; BABU, A.Y.; RAO, K. PONDR Predictors of Natural Disorder Regions. **IJCTEE**. vol. 2 n. 4. 2012.
- BOSSART-WHITAKER, P.; CARSON, M.; BABU, Y.S.; SMITH, C.D.; LAVER, W.G.; AIR, G.M. Three-dimensional structure of influenza A N9 neuraminidase and its complex with the inhibitor 2-deoxy 2,3-dehydro-N-acetylneuraminic acid. **J Mol Biol**. n. 232 p.1069-83.1993.
- BOUVIER, N. M.; PALESE, P. The biology of influenza viruses. **Vaccine**. vol. 12 n. 24. p. D49-D53. 2008.
- DELANO, W.L. The Pymol Molecular Graphics System. **Delano Scientific**. San Carlos. 2002.
- EKATERINA, K.; CAPRONI, L.J.; TREGONING, J.S. A comparison of red fluorescent proteins to model DNA vaccine expression by whole animal in vivo imaging. **PLoS One**. vol. 10. n. 6. e0120375. 2015.
- GASTEIGER, E.; HOOGLAND, C.; GATTIKER, A.; DUVAUD, S.; WILKINS, M.R.; APPEL, R.D.; BAIROCH, A. Protein identification and analysis tools on the Expasy Server. **The proteomics protocols handbook**. Humana Press p.988. 2005.
- GILL, S. C.; VON HIPPEL, P. H. Calculation of protein extinction coefficients from amino acid sequence data. **Analytical Biochemistry**. n 182. p. 319-326. Institute of molecular biology and department of chemistry, University of Oregon. Eugene. 1989.

KENDAL, A.P.; KILEY, M.P.; ECKERT E.A. Isoelectric focusing studies of A21957 influenza neuraminidase and its subunits. **Biochimica et Biophysica Acta**. vol. 317. p. 28-33. 1973.

KYTE, J.; DOOLITTLE, R.F. A simple method for displaying the hydropathic character of a protein. **J Mol Bio**. vol. 157 n. 105 p.32.1982.

KUMAR, A.; MELDGAARD, T.S.; BERTHOLET, S. Novel Platforms for the Development of a Universal Influenza Vaccine. **Frontiers In Immunology**. vol. 9. 2018.

KUMAR, S.; STECHER, G.; TAMURA, K. MEGA7: Molecular Evolutive genetics analysis version 7.0 for bigger datasets. **Mol Biol Evol**. vol. 33 n.7 p.1870-1874. 2016.

LASKOWSKI, R.A.; MacARTHUR, M.W.; MOSS, D.S.; THORNTON, J.M. PROCHECK: a program to check the stereochemical quality of protein structures. **J Appl Cryst**. vol. 26 p.283-291. 1993.

MCGUFFIN, L.J.; BRYSON, K.; JONES, D.T. The PSIPRED protein structure prediction server. **Oxford University Press**. vol. 4 n. 16 p. 404-405. 2000.

NAYAK, D.P.; BALOGUN, R.A.; YAMADA, H.; ZHOU, Z.H.; BARMAN, S. Influenza virus morphogenesis and budding. **Virus Res**. vol. 143n. 2 p. 147–161. 2009.

OSTERHOLM, M.T.; KELLEY, N.S.; SOMMER, A.; BELONGIA E.A. Efficacy and effectiveness of influenza vaccines: a systematic review and meta-analysis. **Lancet Infect Dis**. n 12. vol. 1. p. 36-44. 2012.

RAMACHANDRAN; RAMAKRISHNAN; SASISEKHARAN. Stereochemistry of polypeptide chain configurations. **J Mol Biol**. vol. 7. p. 95-99. 1963.

SALI, A.; BLUNDELL, T.L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. **J. Mol. Biol**. n. 234 p. 778-815. 1993.

SU, X.D.; ZHANG, H.; TERWILLIGER, T.C.; LIJAS, A.; XIAO, J.; DONG, Y. Protein crystallography from the perspective of technology developments. **Crystallogr Rev**. vol. 21 n.1-2 p.122-153. 2015.

UVERSKY, V.N. Intrinsically disordered proteins from A to Z. **Int J Biochem Cell Biol**. vol.43 n. 8 1090-1103. 2011.

VLACHAKIS, D.; TSANIRAS, S.C.; TSILIKI, G.; MEGALOOIKONOMOU, V.; KOSSIDA, S. 3D structural analysis of proteins using electrostatic surfaces based on image segmentation. **J Mol Biochem**. vol. 3 n. 1 p. 27-33. 2014.

WALKER, J.M. Protein identification and analysis tools on the ExPASy Server. **Humana Press Inc.** Totowa. NJ. 2005.

WALSH, E.P.; BARON, M.D.; ANDERSON, J.; BARRETT, T. Development of a genetically marked recombinant rinderpest vaccine expressing green fluorescent protein. **Journal of General Virology**. n. 81. p. 709-718. 2000.

WAN, H.; YANG, H.; SHORE, D.A.; GARTEN, R.J.; COUZENS, L.; GAO, J.; JIANG, L.; CARNEY, P.J.; VILLANUEVA, J.; STEVENS, J.; EICHELBERGER, M.C. Structural characterization of a protective epitope spanning A(H1N1)pdm09 influenza virus neuraminidase monomers. **Nat Commun**. n. 6: 6114. 2015.

WEBB, B.; SALI, A. Comparative Protein Structure Modeling Using MODELLER. **Curr. Protoc Bioinformatics**. vol. 54 p.1-37. 2016.

YANG, J.; ZHANG, Y. I-TASSER server: new development for protein structure and function predictions. **Nucleic Acids Research**. vol. 43. 2015.

YARBROUGH, D.; WACHTER, R.M.; KALLIO, K.; MATZ, M.V.; REMINGTON, S.J. Refined crystal structure of DsRed, a red fluorescent protein from coral, at 2.0 Å resolution. **Proc Natl Acad Sci U S A**. n. 98 n. 2 p. 462-467. 2001.

ZHANG, Y.; SKOLNICK, J. Automated structure prediction of weakly homologous proteins on a genomic scale. **PNAS**. vol. 101 n. 20 p.7594-7599. 2003.