

SELEÇÃO DE CANDIDATOS A FÁRMACOS PARA TRATAMENTO DE DIVERSOS TIPOS DE TUMORES MALIGNOS

Adonis de Melo LIMA

LIMA, Adonis de Melo. **Seleção de candidatos a fármacos para tratamento de diversos tipos de tumores malignos.** Projeto de investigação científica, do Curso de Biomedicina – Centro Universitário Fibra, Belém, 2020.

O câncer surge no organismo a partir de uma mutação genética que pode ou não estar associada aos proto-oncogenes (inativos em células saudáveis), os quais quando ativos se tornam oncogenes. A razão da ocorrência desta mutação pode estar ligada a fatores genéticos hereditários ou fatores externos (estilo de vida) (INCA, 2019). O aumento na mortalidade do câncer nos últimos tempos é fortemente associado ao envelhecimento e crescimento populacional, mudança na distribuição na prevalência de fatores de risco associados à condição socioeconômica. A estimativa mundial mais recente apresentou como números de novos casos 18 milhões, sendo as mortes de 9,6 milhões; no homem em ordem do mais frequente para o menos foi apresentado o câncer de

pulmão (14,5%), próstata (13,5%), cólon e reto (10,9%), estômago (7,2%) e fígado (6,2%); em mulheres os tipos de câncer mais incidentes foram os de mama (24,2%), cólon e reto (9,5%), pulmão (8,4%) e colo de útero (6,6%). No Brasil, para cada um dos anos do triênio 2020 -- 2022, foi calculado o surgimento de 625 mil novos casos de câncer (INCA, 2020 apud BRAY *et al.*, 2018). A principal característica dessa doença é o crescimento descontrolado de células anormais, que juntas formam os tumores; diversos tecidos tumorais apresentam grande dependência do aminoácido L-asparagina para crescer e se desenvolver, sendo a obtenção desse aminoácido feita do meio extracelular pelas células cancerígenas, pois estas têm silenciamento do gene asparagina sintetase. Alguns medicamentos quimioterápicos agem hidrolisando a asparagina em amônia e ácido aspártico (aspartato). Dessa maneira eliminam a fonte sérica de asparagina das células tumorais, o que causa a morte dessas células, uma apoptose seletiva das células cancerígenas. Esse efeito é possível pela presença da enzima asparaginase nos medicamentos, um importante agente antineoplásico (LACERDA, 2017). Os principais medicamentos utilizados fazem uso de enzimas asparaginases derivadas dos

microorganismos *Escherichia coli* e *Erwinia chrysanthemi*. Alguns pacientes apresentam hipersensibilidade ao medicamento com a enzima de *E. coli*, que pode causar problemas ao sistema imune, desencadeando a síntese de anticorpos anti-asparaginases, e levando à inativação enzimática e outros efeitos colaterais imunogênicos. Segundo Casas *et al.* (2018), entre os principais efeitos colaterais do uso da asparaginase de *E. coli*, estão a hipersensibilidade (10-30%), pancreatite (5-10%), tromboembolismo venoso (3,2%), distúrbios do sistema nervoso central (33%), hiperglicemia (11-19%), disfunção hepática (87%), infecções (19-29%), mielossupressão (63-92%) e segundas neoplasias (1,1-5,4%). Por essa razão, se busca uma proteína de outro organismo para atuar como fármaco no tratamento anticancerígeno. Muitos pesquisadores utilizam a modelagem por homologia para construir modelos proteicos para atuar como esses fármacos (CUSTODIO, 2018). A modelagem de proteínas por homologia, também conhecida como modelagem comparativa, se apresenta no cenário atual como uma alternativa de baixo custo e bem-sucedida para a elucidação de estruturas tridimensionais de proteínas. Essa estratégia consiste em modelar uma proteína em sua

estrutura tridimensional (proteína-problema) com base em uma proteína homóloga (proteína-molde), que se apresenta como sequência primária de aminoácidos (sequência fasta). A razão de se usar estruturas primárias como molde é que a conformação estrutural de uma proteína é mais bem conservada em sua sequência de aminoácidos durante o processo evolutivo e, assim, há uma similaridade entre as sequências primárias da proteína-problema e a proteína-molde implicando correspondência entre as estruturas tridimensionais delas (SANTOS FILHO e ALENCASTRO, 2003; CUSTODIO, 2018). A prospecção e o design de novos fármacos vêm ganhando impulso através do uso de ferramentas computacionais, devido a reduções de custo, tempo e tornando conveniente o do ferramental de bioinformática. O objetivo desta investigação foi realizar análise *in silico* do potencial biotecnológico de asparaginase de *Erwinia billingiae* e *Erwinia raphontici*. A partir da sequência fasta da L asparaginase de *Erwinia billingiae* (Eb661), que foi obtida do genbank (ID: CAX58976.1), foi feita uma pesquisa no site do Protein Data Bank (PDB) para obtenção de proteínas de outros organismos, as quais apresentaram sequências primárias semelhantes à da L

asparaginase do organismo-alvo. Com os resultados obtidos dessa pesquisa, foi montada uma tabela no programa excel, contendo os valores de cada proteína para resolução, comprimento, valor de E, pontuação, identidade, similaridade e espaços. Esses valores foram usados como critérios para a escolha do melhor modelo proteico. Com o melhor modelo proteico dentre os quatro pré-selecionados do PDB, foi realizado um alinhamento entre as sequências fastas do modelo molde escolhido e da L asparaginase de *Erwinia billingiae*. Esse processo foi feito no programa modeller 9v8. Nessa etapa de alinhamento são comparadas as duas sequências fasta das proteínas baseado nas características de similaridade entre os resíduos. Para que o alinhamento ocorresse, foram necessários três arquivos de input: alvo.ali (contendo o fasta da proteína do organismo alvo); 4O0E (fasta da proteína molde escolhida); Align2d.py (script indicando em que cadeia o alinhamento seria realizado). Ao final do alinhamento, outros três arquivos foram produzidos, sendo o alvo-1bdmA.ali o mais importante, pois seria usado na próxima etapa de construção dos modelos. Nessa etapa foram construídos cinco modelos proteicos com base nas características da proteína molde

escolhida no PDB e no alinhamento desta com a sequência da proteína-alvo. Esse processo foi realizado também dentro do programa modeller, utilizando quatro arquivos de *input*: alvo-1bdmA.ali (arquivo de alinhamento da etapa anterior); alvo.ali (sequência alvo); 4O0E (fasta da proteína molde escolhida); Model-single.py (script onde é indicado a quantidade de modelos a serem gerados). No fim do processo, foram gerados arquivos contendo os modelos construídos nomeados alvo1.pdb, alvo2.pdb, alvo3.pdb, alvo4.pdb e alvo5.pdb; além de um arquivo com um resumo da construção dos modelos chamado model-single.log. Os modelos gerados foram visualizados com o auxílio de um programa para visualização 3D, sendo utilizado um chamado VMD (Visual Molecular Dynamics) e outro o Pymoll, sendo feita a descrição estrutural do modelo escolhido com base nessas imagens em 3D. Os cinco modelos de proteína gerados passaram por um processo de validação para ser escolhido o melhor com base em critérios de avaliação estereoquímica, de energia livre do sistema entre outros. Esse processo foi realizado nos sites: SWISSMODEL; PROSA; PROQ. No site SWISSMODEL o primeiro critério de avaliação é o desempenho no diagrama de Ramachandran, o qual

permite a observação de aminoácidos em que não há colisão entre os átomos e nos em que pode haver esse choque, pela determinação da rotação dos ângulos de torção anteriores (ϕ) e posteriores (ψ) ao carbono central do aminoácido (carbono alfa). O Ramachandran fornece uma pontuação para quantificar os aminoácidos que estão na zona pintada, que é considerado como bom se for quanto mais acima de 90%. Outro critério levado em conta no SWISSMODEL foi o QMEAN, o qual é um importante estimador tanto da qualidade para toda a estrutura da proteína, quanto da qualidade por resíduos. Esse critério constitui-se de quatro parâmetros o C beta, All Atom, Solvation e Torsion. O gráfico chamado Comparison With Non-Redundant Set Of PDB Structure avalia a proteína em comparação com outras estruturas proteicas, sendo melhor a proteína (representada pela estrela vermelha) está o mais próximo possível da parte mais pigmentada, pois essa área representa o desempenho de outras estruturas proteicas segundo as variáveis tamanho da proteína (horizontal, eixo x) e valor normalizado de QMEAN4 (vertical, eixo y). Prosa. No critério chamado Z-score é fornecida uma pontuação que é considerada boa se for um número negativo. Esse posicionamento avalia a

energia de cada aminoácido, sendo as energias de valores mais negativos mais favoráveis. O Preditor de Qualidade Proteica (ProQ) avalia a qualidade da estrutura proteica, por meio da análise de rede neurais, e realiza a mensuração por meio de duas medidas, o LGscore e o MaxSub. Na sua avaliação, o ProQ observa características estruturais, como a frequência de contato átomo-átomo. Segundo o ProQ, um modelo bom tem LGscore $> 1,5$ e Maxsub $> 0,1$; um modelo muito bom LGscore $> 2,5$ e Maxsub $> 0,5$; um modelo extremamente bom LGscore > 4 e Maxsub $> 0,8$. Na observação do sítio ativo, é realizado um alinhamento no programa Pymoll entre o melhor modelo gerado e a proteína 4O0E para que seja possível observar onde estão localizados os sítios ativos, descritos no artigo da proteína molde, dentro do modelo gerado. No PDB, foi feita a busca por meio da sequência de aminoácidos da proteína-alvo, e o resultado foi de 63 proteínas homólogas à *Erwinia billingiae*. Essas proteínas foram avaliadas segundo os quesitos: resolução, comprimento, valor de E, pontuação, identidade, similaridade e espaços, sendo que, no critério de resolução mensurado em angstrom, e, quanto menor o número, melhor será o modelo, indicando uma menor incerteza na

posição de átomos e um maior nível de detalhes disponíveis nos dados experimentais. O comprimento deve um número mais próximo possível do tamanho da proteína-alvo, que é de 320 aminoácidos, indicando maior similaridade entre as sequências. O valor de E, quanto mais próximo de zero, é melhor, significando que as duas sequências alinhadas têm grande semelhança. A pontuação deve ser a maior possível, sinalizando um melhor alinhamento. Identidade e similaridade, expressas em porcentagem, devem ser as maiores possíveis, já no critério de lacuna o número ideal é 0%, quanto menos lacunas houver, menos alças irão ser formadas na estrutura tridimensional, e, assim, aumenta a estabilidade. Dessa maneira, foi escolhida a proteína 3C17 em primeiro, 4O0E em segundo, 4O0C em terceiro e 4PVR em quarto. A proteína identificada com o código 3C17 foi utilizada para rodar no programa *modeller* 9v8, gerando cinco modelos de proteínas, porém, quando esses modelos foram submetidos aos métodos de validação das plataformas SWISMODEL, ProsA e ProQ, os resultados não foram satisfatórios, o que levou a testagem da proteína molde 4O0E no programa. Os cinco modelos obtidos a partir desse segundo molde se saíram melhores na validação

que os anteriores, sendo o modelo 1 o melhor entre os cinco. O programa modeller forneceu cinco modelos terciários proteicos já citados anteriormente como “alvo1.pdb”, “alvo2.pdb”, “alvo3.pdb”, “alvo4.pdb” e “alvo5.pdb”, sendo considerado como o melhor modelo o “alvo1.pdb” segundo as plataformas utilizadas para validação. A proteína “alvo1.pdb” contém 320 aminoácidos (Aa) formando sua estrutura, sendo onze betas folhas (β), oito alfas hélices (α) e vinte alças. A proteína “alvo1.pdb” foi criada a partir de uma proteína homóloga a L asparaginase de *Erwinia billingiae*, sendo esse molde a 4O0E (uma L-Asparaginase humana – hASNase 3), portanto, essas duas devem apresentar aspectos em comum. A 4O0E é uma proteína que contém suas cadeias (cadeia A e cadeia B) idênticas de 308 aminoácidos, e cada uma das cadeias tendo como metal o íon de sódio (Na); sendo em cada cadeia encontrado treze betas folhas, oito alfas hélices e vinte e duas alças. Esses números são bem similares aos descritos na proteína “alvo1.pdb” com onze beta folhas, oito-alfa hélices e vinte alças. Também é observado que as distribuições espaciais das estruturas são bem parecida, estando as betas folhas localizadas mais ao centro enquanto as alfas hélices ficam mais

externas ao redor. Essa distribuição das estruturas pode ser justificada ao comparar os aminoácidos presentes nas betas folhas com os das alfas hélices e, assim, perceber que nas betas o número de aminoácidos polares é 18, um número bem inferior ao presente nas alfas que é de 38 na “alvo1.pdb”, e na 4O0E ocorre o mesmo fenômeno com 26 polares nas betas folhas e 53 polares nas alfas hélices. Percebeu-se que a parte mais apolar da proteína (beta folha) estava mais ao centro sem contato com o meio externo, e a parte mais polar (alfa hélice) estava mais externa e em contato com o meio, e, tendo em vista que a polaridade determina a afinidade com água, pôde-se, por assim, determinar que beta folhas são hidrofóbicas e alfa hélices são hidrofílicas. No artigo correspondente a 4O0E, é descrita como sítio-ativo dessa proteína uma tríade de serina presente nos aminoácidos de número 168, 186 e 219, sendo este trio responsável pela hidrólise do substrato; a 4O0E, a qual é uma asparaginase do tipo planta referida como hASNase 3 é membro da família hidrolase de nucleófilos N-terminais (Ntn), sendo as enzimas desse grupo produzidas sem atividade enzimática, assim, necessitando ocorrer uma clivagem autoproteolítica a fim de formar duas subunidades alfa e

beta, para atingir um estado enzimaticamente ativo. Essa autoclivagem ocorre entre os resíduos Gly167 e Thr168. O grupo amino N-terminal da subunidade beta da treonina catalítica funciona como ativador do grupo hidroxila, com o intuito de diminuir a necessidade de enzimas Ntn sofrerem autoclivagem para que possam se tornar enzimaticamente ativas. O artigo também cita que, sobre a presença de um aminoácido com características diferentes, deve preceder cada uma das treoninas da tríade, para exercer um mecanismo de compensação. A 4O0E é a hASNase 3 com a mutação T186V, onde se percebeu que a treonina da posição 186 não é essencial na etapa de clivagem, pois com a valina em seu lugar foi observada a clivagem parcial quando a proteína foi submetida a uma incubação com glicina (fator de aceleração da autoclivagem) com concentração de 2M. Com esses experimentos com glicina, foi também observado que Thr168 é essencial para autoclivagem por meio do uso da enzima com a mutação T168S; já Thr219 não é considerado essencial devido aos resultados obtidos com a mutação T219A/V. Por meio da plataforma PDBsum, foi observado na proteína 4O0E que no local do aminoácido 168 se encontrava uma treonina (polar) localizada em uma região de alça e antecedida por

uma glicina, que é apolar (característica oposta à da treonina); na posição 186, encontra-se uma valina (apolar) situada na região da β -5 e antecedida por uma serina de características contrárias a valina, sendo polar; na posição 219, observa-se a presença de uma treonina na região da β -8, antecedida por uma outra treonina. Na proteína “alvo1.pdb”, o sítio equivalente ao T168 da tríade catalítica é o T185, o qual é antecedido por uma glicina com polaridade contrária a treonina. O segundo sítio dessa tríade, que, na 4O0E, é o T186V na proteína gerada T203, antecedido por uma serina com as mesmas características da treonina. No último sítio a posição equivalente à T219 é à T236 antecedida de uma cisteína com características iguais a treonina. **A** proteína que melhor se enquadrou nos critérios de seleção foi a 2P2D, cuja diferença da 2P2N está na resolução, porém esta proteína foi reagrupada para o segundo lugar no ranking e deu-se preferência para a 2P2N devido a sua quantidade de 4 ligantes contra 1 da 2P2D. A quantidade de ligantes favorece a interação da estrutura com outras moléculas, fato esse de extrema relevância no processo de docagem. Como resultado desse alinhamento, foram gerados outros 3 arquivos: align2d.log, que apresentam um resumo do alinhamento;

TvLDH-1bdmA.pap, que é um arquivo demonstrativo que faz a comparação do alinhamento entre as proteínas alvo e molde; e TvLDH- 1bdmA, que é um arquivo de alinhamento entre alvo e molde. Esse último arquivo foi utilizado na etapa seguinte do processo. Da etapa descrita acima, foram gerados cinco modelos: TvDH.B99990001, TvDH.B99990002, TvDH.B99990003, TvDH.B99990004 e TvDH.B99990005, O gráfico de Ramachandran mostrou os resíduos que se encontram nas regiões energeticamente mais favoráveis e desfavoráveis, indicando em quais áreas se encontram os aminoácidos com menor risco de colisão entre os átomos, e as áreas em que pode ocorrer esse choque. Essa estimativa é dada a partir da rotação dos ângulos de torção anteriores ϕ (phi) e posteriores ψ (psi) ao carbono central do aminoácido (carbono alfa). A estrutura proteica possui três ângulos principais, Ω (ômega), ϕ e ψ . Enquanto o ângulo Ω é fixo, os ângulos ϕ e ψ são flexíveis fazendo com que haja variação conformacional na cadeia principal. Para que o resultado dos modelos gerados seja considerado positivos, é necessário que no mínimo 90% dos ângulos ϕ e ψ da cadeia principal estejam na região mais favorável do gráfico de Ramachandran. Logo, as áreas mais verdes

representam as regiões mais favoráveis, a área de verde claro representa a área favorável, a área cinza representa a região pouco favorável, e área branca representa a região desfavorável (SWISS-MODEL, 2019). O QMEAN faz a estimativa de qualidade absoluta global (para toda a estrutura) e local (por resíduo) dos modelos gerados, com base em diferentes propriedades geométricas a partir de um único modelo (SWISS-MODEL, 2019). O escore Z indica se a pontuação QMEAN do modelo gerado é compatível ao tamanho de estruturas experimentais semelhantes. Esta pontuação, quanto mais perto de zero, é melhor, pois mostra que há uma boa concordância entre a estrutura do modelo e as estruturas experimentais de tamanho semelhante. Após a validação dos modelos gerados, foi feita a comparação entre eles para que o melhor fosse escolhido para dar sequência ao trabalho em desenvolvimento. Com base na análise do resultado da validação dos cinco modelos gerados, o escolhido para o estudo foi o modelo 3, por estar mais dentro dos critérios de seleção exigidos pelo SWISS-MODEL. A maioria dos aminoácidos (aa) das proteínas moldes se conservaram nos modelos gerados, havendo apenas algumas trocas, em sua maioria, por aa de mesmo grupo físico químico. Os

resultados foram promissores e garantem a continuidade dos estudos como a realização das etapas de docagem e dinâmica molecular. Essas etapas trarão refinamento para os resultados obtidos neste estudo e podem levar a um patamar experimental.