

Estudos de modelagem molecular da proteína: Ácido siálico sintase de *Homo sapiens*

Molecular modelling studies of the protein: Sialic acid synthase of *Homo sapiens*

RESUMO

Joyce Araújo de Oliveira
joycearau@outlook.com
Universidade Tecnológica Federal do Paraná. Ponta Grossa, Paraná, Brasil.

Marcio Silva
marcios@utfpr.edu.br
Universidade Tecnológica Federal do Paraná. Ponta Grossa, Paraná, Brasil.

O gene NANS de *Homo sapiens*, disponível no Genbank NM_018946.4, codifica a enzima ácido siálico sintase (E.C: 2.5.1.56) que participa da via biosintética dos ácidos siálicos. Pesquisas apontam que mutações neste gene afetam a atividade catalítica da enzima, o que leva a um quadro de nanismo e retardo mental. No banco de dados PDB há o depósito de parte da estrutura (PDB: 1WVO), correspondente à apenas 18% da proteína, revelando a necessidade de estudos estruturais mais completos. Este trabalho tem por objetivo a construção da estrutura molecular da ácido siálico sintase por homologia. Tratou-se de dois modelos homólogos construídos com o programa MODELLER, onde para um modelo utilizou-se múltiplas proteínas moldes e no outro apenas uma. Os aminoácidos encontrados em regiões desfavoráveis no gráfico de Ramachandran foram ajustados por minimização de energia. Submetidas a *softwares* com parâmetros analíticos estruturais, obteve-se que os modelos se enquadraram como ótimos. Assim, acredita-se que os homólogos agregaram mais conhecimento sobre a estrutura completa da proteína NANS sintase, sendo esta essencial para propor estratégias de combate a disfuncionalidade da enzima.

PALAVRAS-CHAVE: Enzima. Bioinformática. Estrutura molecular.

Recebido: 19 ago. 2020.

Aprovado: 01 out. 2020.

Direito autoral: Este trabalho está licenciado sob os termos da Licença Creative Commons-Atribuição 4.0 Internacional.



ABSTRACT

The NANS gene of the *Homo sapiens*, available in Genbank NM_018946.4, encodes the enzyme sialic acid synthase (E.C: 2.5.1.56), that participates in the biosynthetic pathway of sialic acids. Research points out that mutations in this gene affect the catalytic activity of the enzyme leading to dwarfism and mental retardation. In the PDB database there's the deposit of part of the structure (PDB: 1WVO), corresponding to only 18% of the protein, revealing the need for more complete structural studies. So, this work has as objective the construction of the molecular structure of the sialic acid synthase by homology. It was the building of two homologous models with the MODELLER program, where for one model multiple protein molds were used and in the other only one. The amino acids found in unfavorable regions in the Ramachandran graph were adjusted by energy minimization. Submitted to software with structural analytical parameters, it was obtained that these were framed as excellent models. So, it's believed that the homologues added more knowledge about the complete structure of the NANS synthase protein being this essential to propose strategies to combat the dysfunctionality of the enzyme.

KEYWORDS: Enzyme. Bioinformatics. Molecular structure.



INTRODUÇÃO

O gene *N*-acetilneuraminato sintase (NANS) de *Homo sapiens*, do qual informações estão disponíveis no *National Center for Biotechnology Information* (NCBI), sob o código HGNC: 19237 codifica para uma transferase. Essa enzima (E.C: 2.5.1.56), cuja sequência de aminoácidos está disponível no NCBI: NP_061819.2, atua nas vias biossintéticas dos ácidos siálicos e produz formas fosforiladas de ácido *N*-acetilneuramínico (Neu5Ac) e 2-ceto-3-desoxi-D-glicero-Dgalacto-nonônico ácido (KDN); sendo a biossíntese de Neu5Ac a principal reação catalisada por essa enzima, conforme descrito por Lawrence et al.(2000). Van Karnebeek et al.,2017 apontam que o nanismo presente em pessoas com mutações no gene NANS, leva a crer que a síntese mediada pela enzima ácido *N*-acetilneuramínico sintase ou, também chamada, ácido siálico sintase exerce um papel fundamental no desenvolvimento do esqueleto estando relacionada com a cartilagem na placa de crescimento. O cérebro possui altas concentrações de ácido siálico, presente nos glicoesfingolípídios, encontrados em maior concentração no córtex, como é apresentado nos estudos de Wang e Brand-Miller (2003), contudo ainda sem total clareza de funções metabólicas.

Sendo assim, é apontado que o mau funcionamento enzimático, leva a um quadro de displasia, doença observada ainda na infância e que afeta o desenvolvimento do cérebro e esqueleto gerando um quadro de nanismo e retardo mental. Hamada et al.,2006 elucidou 18% da estrutura da proteína ácido siálico sintase (*Sialic acid synthase-SAS*) por ressonância magnética nuclear (RMN), depositada no *Protein Data Bank* (PDB) sob o código PDB: 1WVO, de forma que até o momento não há a elucidação completa da estrutura da enzima. O presente trabalho tem por objetivo a construção da estrutura molecular da proteína, por homologia, respaldando-se em dados encontrados na literatura sobre o gene NANS e seu produto *N*-acetilneuraminato sintase. Acredita-se que a modelagem por homologia é um método eficiente, de baixo custo financeiro, e que contribui para o entendimento sobre a estrutura completa da proteína NANS sintase, sendo possível serem feitas inferências estruturais que auxiliem para propor estratégias de combate a disfuncionalidade da enzima.

MATERIAL E MÉTODOS

A proteína codificada pelo gene NANS foi utilizada como objeto de estudo deste trabalho *in silico*. Na construção dos modelos por homologia utilizou-se o programa MODELLER. Para tal seguindo as seguintes etapas: obtenção da sequência de aminoácidos da proteína NANS através do site da GenScript®; Seleção das proteínas molde com a ferramenta *Basic Local Alignment Search Tool* (BLAST), utilizando o PDB como parâmetro de busca de proteínas homólogas. A modelagem foi realizada de duas formas: com a utilização de apenas uma proteína molde, sendo denominado neste trabalho como modelo único; com a utilização de múltiplas proteínas, sendo denominada como modelo múltiplo.

As análises de qualidade e confiabilidade dos modelos foram feitas por meio dos softwares PROCHECK, WHATIF e *Pictorial database of 3D structures in the Protein Data Bank* (PDBsum), que possuem parâmetros analíticos estruturais que

os enquadram como: ruim, bom ou ótimo. Os ajustes dos ângulos ϕ e ψ dos resíduos de aminoácidos encontrados em regiões desfavoráveis no gráfico de Ramachandran, fornecido através do programa *PROCHECK*, foram realizados no programa *Swiss-PDB Viewer* (SPDB) através da ferramenta minimização de energia.

RESULTADOS E DISCUSSÃO

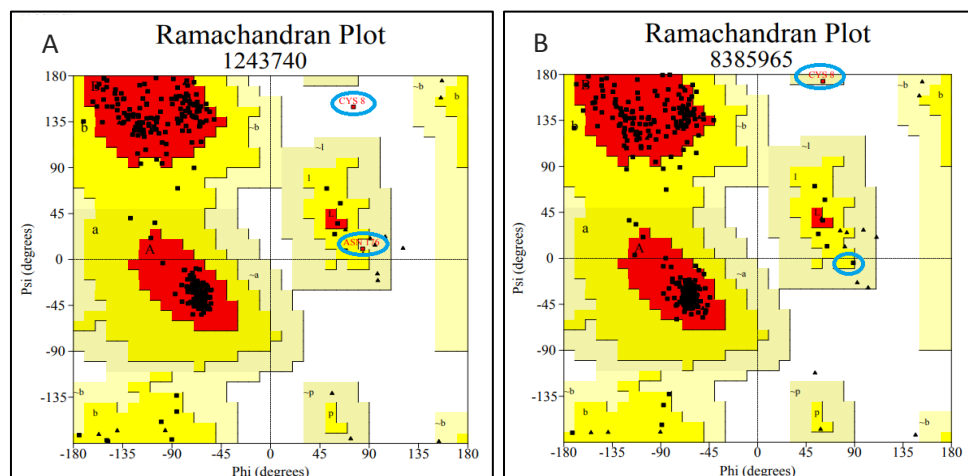
Na construção do modelo estrutural da E.C: 2.5.1.57 por homologia as proteínas moldes escolhidas foram: ácido siálico sintase (Neub) de *Neisseria meningitidis*, cuja a estrutura foi resolvida por Gunawan et al. (2005) pelo método de difração de raio-X, que indicou 31,15% de identidade sob o código PDB: 1XUU. E a ácido siálico sintase, parcial, de *Homo sapiens* com 100% de identidade sob o código PDB: 1WVO. Para quesito de comparação foi feita outra modelagem utilizando apenas a ácido siálico sintase (Neub).

Foram construídos 20 modelos com o programa MODELLER. Com o *software* PROCHEK foi construído o diagrama de Ramachandran. Considerando as informações do diagrama, foram selecionados os modelos com mais aminoácidos dentro das regiões inteiramente permitidas e regiões permitidas com conformações em limites extremos de contatos atômicos desfavoráveis. Sendo escolhidos os que apresentaram os melhores índices no *Coarse Packin Quality Control*. O modelo único que apresentou o índice de -0,394, já o modelo múltiplo que apresentou o índice de -0,558.

No programa SPDB os ângulos ϕ e ψ dos aminoácidos Cisteína 8 (Cys) e Asparagina 176 (Asn), do modelo único foram alterados através da ferramenta de minimização de energia, que segundo De Alencar (2010, p.27) “[...] é feita uma série iterativa de pequenas mudanças nas posições dos átomos da proteína, visando obter uma estrutura de mínimo de energia local.”

Desse modo, os resíduos que estavam em regiões consideradas como proibidas dentro do gráfico Ramachandran (Figura 1) foram ajustados para, as regiões favoráveis ou generosamente permitidas. Após os ajustes, o resíduo de que ocupa a posição número 8, ainda permaneceu fora da angulação. Após estas alterações nos ângulos, o índice *Coarse Packin Quality Control* normalizado apresentou-se igual -0,467, este considerado como um ótimo modelo. No Ramachandran foi indicado que 93,2% (289) dos resíduos estão nas regiões mais favoráveis, 6,5% (20) em regiões adicionalmente permitidas e 0,3% (1) de resíduos nas regiões generosamente permitidas.

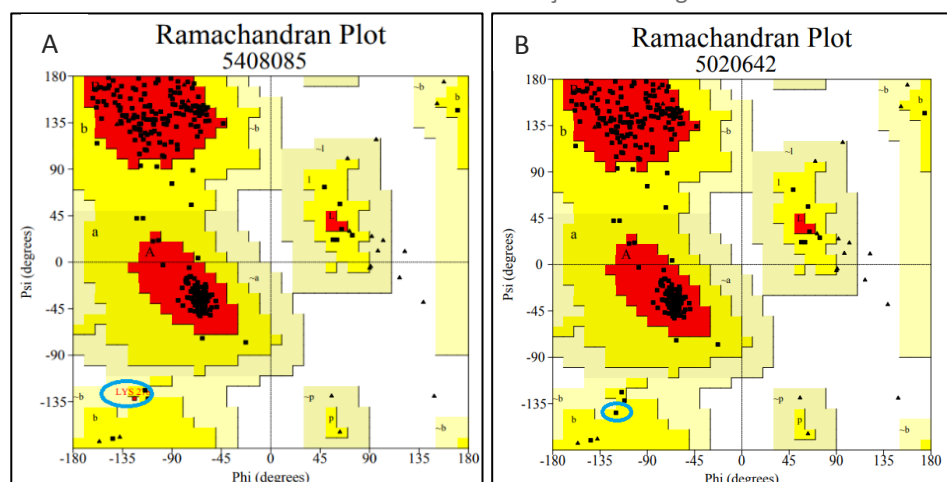
Figura 1 – A- Gráfico Ramachandran do modelo único, antes dos ajustes nos ângulos ϕ e ψ . B- Gráfico Ramachandran após os ajustes nos ângulos ϕ e ψ através da ferramenta de minimização de energia.



Fonte: Autoria própria (2020).

O modelo múltiplo apresentou apenas o resíduo Lisina 278 (Lys) na região considerada como proibida dentro do diagrama de Ramachandran (Figura 2). Os ângulos ϕ e ψ deste resíduo também foram alterados por de minimização de energia, no SPDB, de forma que este foi para a região favorável e com o índice *Coarse Packin Quality Control* normalizado permanecendo igual -0,588, sendo considerado como um ótimo modelo. O diagrama de Ramachandran indicou que 93,5% (290) dos resíduos estão nas regiões mais favoráveis, 6,1% (20) dos resíduos em regiões adicionalmente permitidas.

Figura 2 – B- Diagrama de Ramachandran do modelo múltiplo, antes dos ajustes nos ângulos ϕ e ψ . B- Diagrama Ramachandran após os ajustes nos ângulos ϕ e ψ com da ferramenta de minimização de energia.

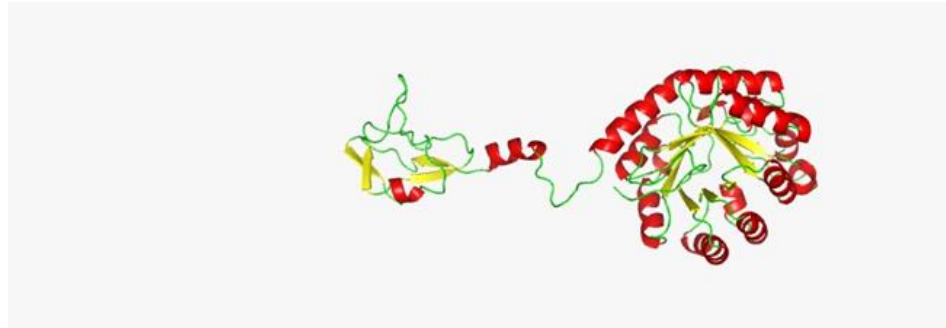


Fonte: Autoria própria (2020).

A proteína E.C: 2.5.1.56, metabolicamente ativa é um homodímero, estrutura que contém duas cópias da mesma cadeia polipeptídica, e apresenta dois domínios, algo que já havia sido discutido por Hamada et al. (2006), mas que não há nada disponível no PDB sobre a estrutura completa. Tanto no modelo único

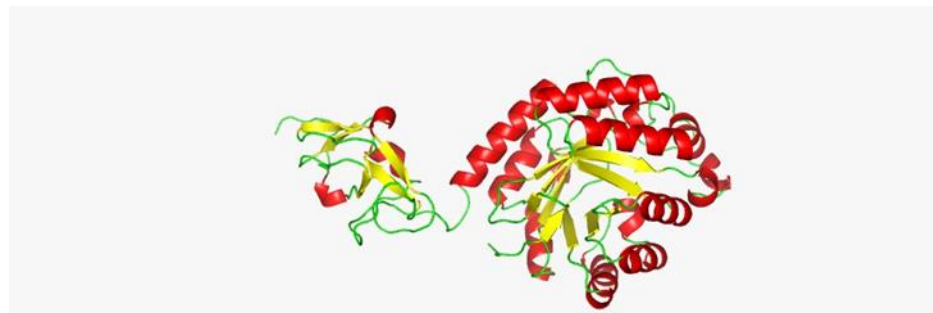
(Figura 3), quanto no modelo múltiplo (Figura 4) são mostrados apenas a estrutura terciária de um dos monômeros.

Figura 3 –Estrutura terciária homóloga da proteína NANS, modelo único.



Fonte: Autoria própria (2020).

Figura 4 – Estrutura terciária homóloga da enzima NANS, modelo múltiplo.



Fonte: Autoria própria (2020).

Em 2006 foi lançada no PDB a estrutura do *antifreeze-like* (AFL) domínio anticongelante do terminal C, parcial de *Homo sapiens*, resolvida por RMN, sob o código PDB: 1WVO. Os autores responsáveis pela elucidação comparam a SAS de bactérias a de mamíferos “[...] ambas as enzimas são compostas por dois domínios distintos que são unidos por uma região de ligação estendida.” (HAMANDA *et al*, 2006, p. 1010-1011, tradução nossa¹). Sendo essa conformação também observadas nos modelos homólogos construídos.

Com base na estrutura SAS da *Neisseria meningitidis* Hamada *et al.* (2006, p. 2011, tradução nossa²) propuseram que “o domínio AFL parece desempenhar um papel importante na ligação do substrato pelo SAS e ser a chave para caracterizar as vias metabólicas do ácido siálico”. De forma, que este PDB: 1WVO é proposto que está faça parte do sítio catalítico da enzima ácido siálico humana.

As construções foram submetidas ao programa PDBsum, que fornece um aspecto geral de cada estrutura macromolecular. O modelo único apresentou 30,4% de estruturas hélices- α e 2 folhas- β paralelas e 1 folhas- β antiparalelas. Já o modelo múltiplo evidenciou 32,6% de estruturas hélices- α e, também, 2 folhas- β paralelas e 1 folhas- β antiparalelas.

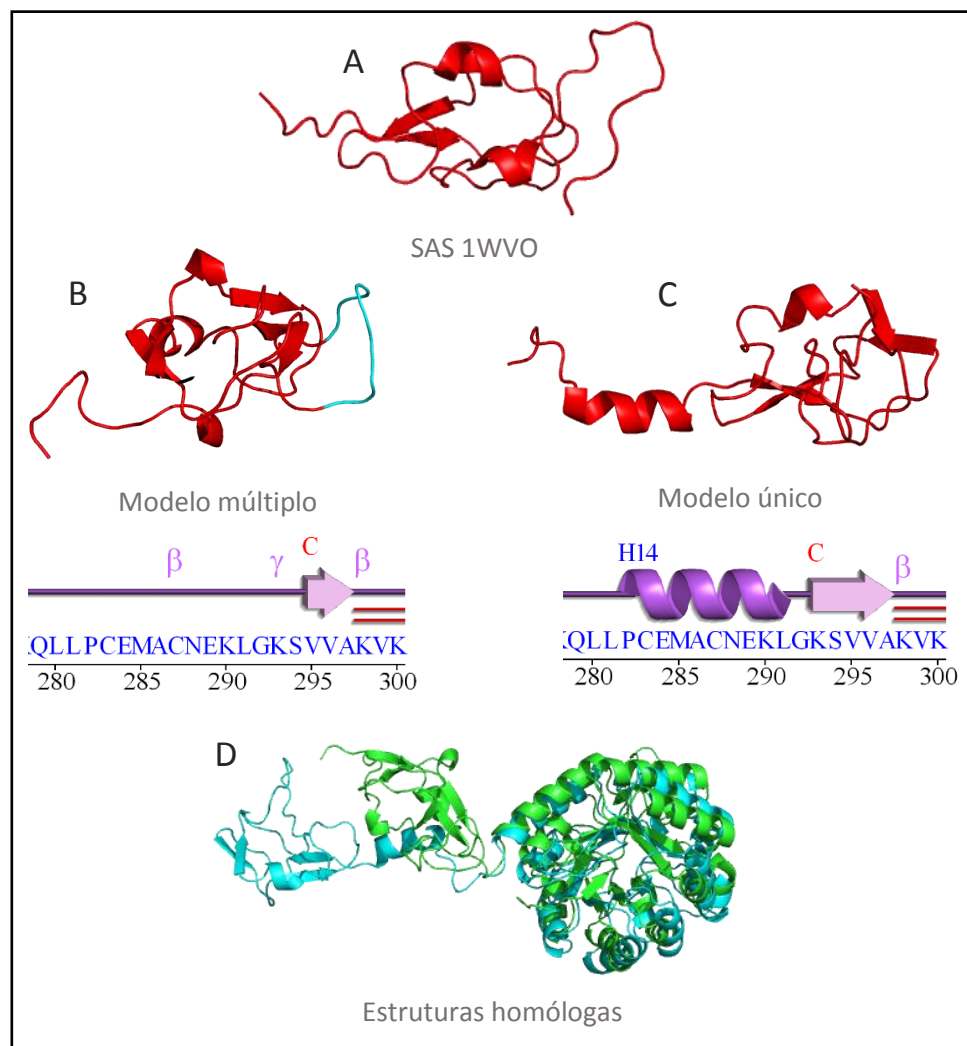
¹ “[...] both enzymes are composed of two distinct domains that are joined by an extended linker region.”.

² “the AFL domain seems to play an important role in substrate binding by the SAS and to hold the key toward characterizing the sialic acid metabolic pathways”.

Ao verificar o gráfico Ramachandran do modelo múltiplo notou-se que os ângulos ϕ e ψ dos resíduos de aminoácidos 283 ao 291 não estão nas angulações favoráveis para assumir a conformação hélice- α . Diferentemente do diagrama Ramachandran do modelo único, que indicou angulações propícias para esses resíduos assumirem esta forma. Isso explica a diferença na porcentagem desta configuração quando os modelos são comparados.

Tanto o modelo único quanto o modelo múltiplo construídos, quando comparados ao PDB: 1WVO (Figura 5), pode-se observar a similaridade quanta a presença de folhas- β e também estruturas hélices- α nessa região catalítica. Contudo, o sítio ativo pode sofrer alterações estruturais quando considerada a proteína completa.

Figura 5 – Comparação das estruturas A, B e C. Resíduos 283 ao 291 das imagens B e C assumem conformações diferentes, evidenciado também pelo quadro da estrutura secundária abaixo de cada imagem, sendo destacado em azul na imagem B uma região de alça o que na imagem C é uma estrutura hélices- α . A imagem D é a sobreposição das estruturas preditas, em azul o modelo múltiplo em verde o modelo único, sendo observado que ambas assumem conformações similares.



Fonte: Autoria própria (2020).

Vale ressaltar que os resíduos 283 ao 291 do modelo múltiplo e único estão, respectivamente, dentro das regiões favoráveis e regiões adicionalmente favoráveis no gráfico Ramachandran e por isso foram considerados bons modelos. Segundo Rodwell et al. (2017, p. 39) “Para muitas enzimas, as alças que fazem ligação entre os domínios responsáveis por ligar substratos, frequentemente contêm resíduos aminoácil que participam da catálise”. A estrutura 1WVO não tem equivalência para esses aminoácidos, também não foi encontrada na literatura que esses fazem parte da região catalítica, o que deve ser posteriormente analisado.

CONCLUSÕES

Os modelos construídos para a enzima ácido siálico sintase de *Homo sapiens*, permitiram ampliar os conhecimentos estruturais da enzima completa. As conformações das estruturas secundárias e terciárias observadas poderão ser exploradas quanto a suas relações e interações com o sitio catalítico já relatado na literatura. Cabe ressaltar que as informações obtidas por modelagem, apesar de importantes, não suprem a necessidade de estudos estruturais da enzima completa por métodos experimentais. Nessa perspectiva, há a intenção de realizar testes de expressão do gene NANS que visam a produção da proteína para posterior elucidação estrutural por cristalografia.

REFERÊNCIAS

BENSON, Dennis A. et al. GenBank. **Nucleic acids research**, v. 45, n. D1, p. D37-D42, 2017. Disponível em: https://www.ncbi.nlm.nih.gov/nuccore/NM_018946.4. Acesso: 3 ago. 2020.

Bioquímica ilustrada de Harper [recurso eletrônico] / Victor W. Rodwell ... [et al.]; [tradução: Luís Fernando Marques Dorvillé, Maria Elisabete Costa Moreira, Simone Kobe de Oliveira; revisão técnica: Guilhian Leipnitz]. – 30. ed. – Porto Alegre: AMGH, 2017.

DE ALENCAR, S. A. (2010). **Utilização de ferramentas computacionais para o estudo do impacto funcional e estrutural de nsSNPs em genes codificadores de proteínas**. (Doctoral dissertation, Universidade Federal de Minas Gerais). Disponível em: <http://www.pgbioinfo.icb.ufmg.br/defesas/52D.PDF>. Acesso: 20 ago. 2020.

G. Vriend, C. Sander. **Quality control of protein models: Directional atomic contact analysis**. J.Appl.Cryst. 1993. Disponível em: <https://swift.cmbi.umcn.nl/whatif/>. Acesso: 3 ago. 2020.

GenScript. **NANS cDNA ORF clone, Homo sapiens (humano)**. 2008. Disponível em: <https://www.genscript.com/gene/homo-sapiens/54187/nans.html#orf>. Acesso em: 3 ago. 2020.

GUEx, N.; PEITSCH, M. C. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. **Electrophoresis**, v. 18, n. 15, p. 2714–2723, dez. 1997.

HAMADA, Toshiyuki et al. Solution structure of the antifreeze-like domain of human sialic acid synthase. **Protein science**, v. 15, n. 5, p. 1010-1016, 2006. Disponível em: <https://doi.org/10.1110/ps.051700406>. Acesso em: 10 ago. 2020.

LAWRENCE, Shawn M. et al. Cloning and expression of the human N-acetylneuraminic acid phosphate synthase gene with 2-Keto-3-deoxy-d-glycero-d-galacto-nononic acid biosynthetic ability. **Journal of Biological Chemistry**, v. 275, n. 23, p. 17869-17877, 2000. Disponível em: <https://www.jbc.org/content/275/23/17869.abstractN2>. Acesso em: 10 ago. 2020.

NCBI (Org). **N-acetylneuraminic acid synthase [*Homo sapiens* (human)]**. Disponível em: <https://www.ncbi.nlm.nih.gov/gene/54187>. Acesso em: 3 ago. 2020.

PDB ID: 1WVO (2006). HAMADA, T. et al. Solution structure of the antifreeze-like domain of human sialic acid synthase. **Protein science**, v. 15, n. 5, p. 1010-1016, 2006. Disponível em: <https://www.rcsb.org/structure/1WVO>. Acesso em: 10 ago. 2020.

PDB ID: 1XUU (2005). GUNAWAN, Jason et al. Structural and mechanistic analysis of sialic acid synthase NeuB from *Neisseria meningitidis* in complex with Mn²⁺, phosphoenolpyruvate, and N-acetylmannosaminitol. **Journal of Biological Chemistry**, v. 280, n. 5, p. 3555-3563, 2005. Disponível em: <https://www.rcsb.org/structure/1XUU>. Acesso em: 10 ago. 2020.

PROCHECK. [S.I.]. Disponível em: <https://servicesn.mbi.ucla.edu/PROCHECK/>. Acesso: 3 agosto 2020.

SCHEER, M. et al. BRENDA, the enzyme information system in 2011. **Nucleic acids research** vol. 39. Disponível em: <https://www.brenda-enzymes.org/enzyme.php?ecno=2.5.1.56>. Acesso: 10 ago. 2020.

SHEN, M.-Y.; SALI, A. Statistical potential for assessment and prediction of protein structures. **Protein science: a publication of the Protein Society**, v. 15, n. 11, p. 2507–2524, nov. 2006. Disponível em: <http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html>. Acesso: 3 ago. 2020.

VAN KARNEBEEK, C. D. M. et al. Corrigendum: NANS-mediated synthesis of sialic acid is required for brain and skeletal development. **Nature genetics**, v. 49, n. 6,

p. 969, 2017. Disponível em:

<https://doi.org.ez48.periodicos.capes.gov.br/10.1038/ng0617-969a>. Acesso em:
10 ago. 2020.

WANG, Bing; BRAND-MILLER, Janette. The role and potential of sialic acid in human nutrition. **European journal of clinical nutrition**, v. 57, n. 11, p. 1351-1369, 2003. Disponível em: <https://doi.org/10.1038/sj.ejcn.1601704>. Acesso em: 10 ago. 2020.

WEBB, B.; SALI, A. Comparative Protein Structure Modeling Using MODELLER. **Current protocols in bioinformatics**, v. 54, p. 5.6.1-5.6.37, jun. 2016.