

**Autores | Authors**

**André Luis Fachini de Souza\***

andre.fachini@ifc.edu.br

**Marlise Pompeo Claus\*\***

marlise.claus@ifc.edu.br

**Viviane Milczewski \*\*\***

viviane.milczewski@ifc.edu.br

**MODELAGEM MOLECULAR DA ENZIMA URIDINA 5'-DIFOSFO-GLICURONIL-TRANSFERASE (E.C.2.4.1.17) HUMANA: UMA PROPOSTA DIDÁTICA**

**MOLECULAR MODELING OF THE HUMAN URIDINE 5'-DIPHOSPHOGLUCURONYLTRANSFERASE ENZYME (E.C.2.4.1.17): A DIDACTIC PROPOSAL**

**Resumo:** A enzima uridina 5'-difosfato-glicuronil-transferase (UGT1A1) é responsável pela conversão da bilirrubina em metabólitos solúveis e excretáveis. Mutações na região promotora ou no gene UGT1A1 ocasionam uma elevação sérica da bilirrubina não conjugada, estando relacionada com algumas doenças humanas. A estrutura da UGT1A1 ainda não foi resolvida, sendo determinada apenas por modelagem estrutural. Possui domínios bastante conservados entre diferentes espécies e outros bastante variáveis. Assim, o objetivo deste trabalho foi modelar a estrutura da proteína UGT1A1 pelo método da homologia, utilizando softwares de acesso livre, como proposta de uma metodologia de ensino para o estudo de proteínas. Nesse sentido, a estrutura da UGT1A1 foi modelada a partir da sequência de aminoácidos e os parâmetros de qualidade do modelo analisado. Foram então identificados tópicos importantes para o ensino de estrutura de proteínas, viabilizando uma metodologia ativa para a aprendizagem do estudante.

**Palavras-chave:** ensino-aprendizagem, estrutura de proteína, metodologia ativa.

**Abstract:** The enzyme uridine 5'-diphosphate-glucuronyltransferase (UGT1A1) is responsible for the conversion of bilirubin into soluble and excretable metabolites. Mutations in the promoter region or the UGT1A1 gene cause a serum elevation of unconjugated bilirubin related to some human diseases. UGT1A1 structure has not yet been solved, is only defined by structural modeling. There are well-conserved domains among different species, and others are quite variable. Thus, the objective of this work was to model the structure of the UGT1A1 protein utilizing the homology method, using open-source software as a teaching methodology to study proteins. Thus, the UGT1A1 structure was modeled from the amino acid sequence, and the quality parameters were analyzed in the model. Important topics were identified for teaching protein structure, making an active methodology possible for student learning

**Keywords:** active methodology, protein structure, teaching-learning.

### Introdução

A uridina 5'-difosfato-glicuronil-transferase é uma enzima envolvida no metabolismo de bilirrubina, sendo essencial para sua excreção biliar. A isoforma UGT1A1 catalisa a reação de glicuronidação da bilirrubina, sendo expressa a partir do *locus* UGT1A1 que expressa outras oito isoformas, porém todas sem muita significância para a catálise desta reação (RITTER et al., 1992; WANG, CHOWDHURY e CHOWDHURY, 2006).

A bilirrubina é o produto da quebra do grupo prostético heme da hemoglobina e outras hemeoproteínas (citocromo, mioglobina, entre ou-

tras). (WANG, CHOWDHURY e CHOWDHURY, 2006). Na primeira etapa da degradação da hemoglobina, a enzima heme-oxigenase catalisa uma quebra entre os anéis pirrol I e II do grupo heme, com consumo de três moléculas de oxigênio e NADPH como agente redutor (Figura 1). Os produtos formados nesta reação são  $\text{Fe}^{3+}$ , CO e biliverdina, que na sequência é transformada em bilirrubina pela ação da biliverdina redutase. A bilirrubina formada é então carregada até o fígado pela albumina para ser conjugada e excretada (RODWELL et al., 2017).

A bilirrubina é capturada pelos hepatócitos por difusão facilitada, ligada em glutathiona-S-transferases e conjugada a ácido glicurônico pela uridina difosfato-glicuronil-transferase (UGT1A1) microsossomal. Uma vez ligadas a glicuronosídeos, são transportadas para os canalículos biliares (WANG, CHOWDHURY e CHOWDHURY, 2006). Esse processo de secreção de bilirrubina conjugada (direta) é bastante eficiente, mantendo baixos os níveis de bilirrubina não conjugada (indireta). Doenças que acarretam no aumento das taxas de formação de bilirrubina reduzem a incorporação de bilirrubina ou reduzem a taxa de conjugação (síndrome de Gilbert) e resultam em níveis elevados de bilirrubina indireta (VANWAGNER e GREEN, 2015).

As proteínas UGT humanas são bem conhecidas com relação à atividade, porém existe pouca informação estrutural, devido, principalmente, a problemas no processo de purificação (LAAKKONEN e FINEL, 2010).

A enzima UGT1A1 é composta por quatro diferentes domínios: N- e C-terminal (domínios maiores) e um envelope de hélices e o segmento transmembrana que inclui uma cauda citoplasmática (domínios menores) (LAAKKONEN e FINEL, 2010).

São duas as regiões bem conservadas entre diferentes espécies, o núcleo hidrofóbico do domínio N-terminal e dois envelopes de hélices que conectam os domínios maiores. Entretanto, a região N-terminal, relacionada à ligação do substrato possui uma estrutura geral bastante variável (LAAKKONEN e FINEL, 2010).

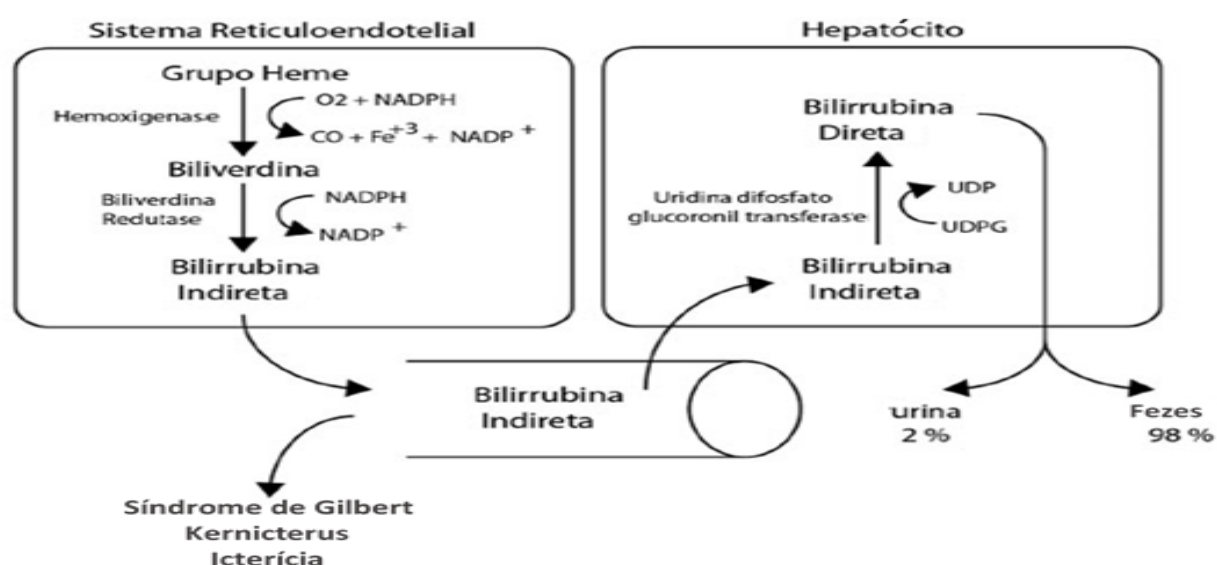
A porção C-terminal, mais conservada, é composta por três domínios estruturais. Nesta porção, uma região  $\alpha/\beta/\alpha$  contém o sítio de ligação de uridina-ácido glicurônico (UDP-GlcUA) (LAAKKONEN e FINEL, 2010).

A modelagem de proteínas por homologia é o método mais acurado para a modelagem da estrutura tridimensional de proteínas, consistindo de quatro passos: seleção de uma estrutura molde, alinhamento com o molde, construção e avaliação do modelo (SANTOS FILHO e ALENCASTRO, 2003).

Um estudo do uso da modelagem de proteínas como proposta de ensino de estrutura de proteínas em disciplina de Bioquímica mostrou ser apropriada como uma alternativa complementar e que contribui para uma maior motivação e compreensão por parte dos estudantes (SABINO et al., 2009).

De acordo com as recomendações das Diretrizes Curriculares Nacionais (DCNs) para cursos na área da saúde, metodologias ativas devem ser incorporadas na prática docente em disciplinas e nos projetos pedagógicos de curso de ensino superior para o fortalecimento de uma educação problematizadora e significativa (MACEDO et al., 2018).

**Figura 1** - Metabolismo de bilirrubina.



Fonte: Adaptado de BESA, CALVO e HARRIS (2014).

A mediação de uma metodologia ativa como a modelagem molecular no ensino ainda é bastante insipiente havendo algumas inserções em estudos isolados em Química ou em Biologia, entretanto sem explorar seu caráter interdisciplinar (RAMOS e SERRANO, 2013).

Nesse sentido, o objetivo deste trabalho foi modelar a estrutura terciária da enzima uridina 5'-difosfato-glicuronil-transferase (UGT1A1) humana a partir da sua sequência de aminoácidos, utilizando o método da homologia como uma proposta de metodologia ativa para o ensino de estrutura de proteínas em disciplina de Bioquímica do ensino superior.

## Material e Métodos

A modelagem proposta neste trabalho como uma metodologia ativa de ensino destina-se ao estudo de estrutura tridimensional de proteínas, principalmente em disciplina de Bioquímica Geral, presente no núcleo básico de cursos superiores de diversas áreas. A escolha da proteína UGT1A1 como modelo para esta proposta aplica-se melhor em cursos da área da saúde, uma vez que mutações na região promotora ou no gene que codifica para esta proteína estão relacionadas à algumas doenças humanas.

## Seleção da Proteína de Estudo

A ferramenta *online Genetics Home Reference* (<https://ghr.nlm.nih.gov/>) foi utilizada para a obtenção de informações a respeito da síndrome de Gilbert, como genes e cromossomos associados (U.S. NATIONAL LIBRARY OF MEDICINE, 2003). Neste sítio, em *Health Conditions* foi efetuada uma busca por diversas categorias de condições, doenças ou síndromes.

No caso da síndrome de Gilbert, o gene relacionado é o *UGT1A1 (UDP glucuronosyltransferase family 1 member A1)*. Acessando o *link*, o usuário é redirecionado para informações do gene como função, doenças relacionadas, localização cromossomal, entre outras características.

Uma vez selecionado o gene envolvido na condição genética em estudo, utilizou-se o banco de dados de proteínas UniProt (<http://www.uniprot.org/>) para obtenção da sequência de aminoácidos. Inseriu-se o nome da proteína codificada pelo gene de escolha (UGT1A1) e selecionou-se a proteína de humanos (*Homo sapiens*). Uma vez selecionada a proteína da espécie de interesse, várias informações são disponíveis como número de aminoácidos, localização cromossomal e subcelular, referências de trabalhos científicos relacionados, entre outras. Também, a sequência de aminoácidos pode ser

acessada no formato FASTA. Neste caso, foi selecionada a sequência de aminoácidos da isoforma 1 da proteína nativa.

A análise da topologia, com visualização da sequência de aminoácidos, foi efetuada por meio da ferramenta *online Protter* versão 1.0 (<http://wlab.ethz.ch/protter/start/>) (OMASITS et al., 2014), utilizando o número de acesso Uniprot da proteína UGT1A1 ([P22309](https://www.uniprot.org/)) (<https://www.uniprot.org/>) (THE UNIPROT CONSORTIUM, 2017).

Para análise da estrutura secundária foi utilizado o programa NPS@: *Network Protein Sequence Analysis* (COMBET et al., 2000). Este programa gera uma predição consenso de estrutura secundária a partir do alinhamento das predições efetuadas pelos métodos DPM (DELEAGE e ROUX, 1987), DSC (KING e STENBERG, 1996), GOR1 (GARNIER, OSGUTHORPE e ROBSON, 1978), GOR3 (GIBRAT, GARNIER e ROBSON, 1987), HNN (GUERMEUR, 1997), MLRC (GUERMEUR, OSGUTHORPE e ROBSON, 1978), PHD (ROST e SANDER, 1993), Predador (FRISHMAN e ARGOS, 1996) e SOPM (GEOURJON e DELEAGE, 1994).

A identificação de domínios ou sítios funcionais foi realizada com a utilização do programa Prosite (<https://prosite.expasy.org/>) (SIGRIST et al., 2002).

## Modelagem estrutural

Para a modelagem estrutural da proteína UGT1A1 utilizou-se o espaço de trabalho *online* do servidor SWISS-MODEL (<https://swissmodel.expasy.org/>) (BIASINI et al., 2014; KIEFER et al., 2009; GUEX, PEITSCH e SCHWEDE, 2009; ARNOLD et al., 2006), o qual modela estruturas proteicas por homologia.

A modelagem de proteínas pelo método da homologia é baseada na concepção de que a semelhança de sequências homólogas de estruturas tridimensionais conhecidas (proteínas-molde) implica na similaridade da estrutura dessas proteínas (CHOTHIA e LESK, 1986; SANTOS FILHO e ALENCASTRO, 2003).

Na página do resultado, em *Summary* é possível ter-se a informação da quantidade de modelos (*templates*) que coincidem com a sequência alvo. Na opção *Templates* visualiza-se a lista completa dos modelos utilizados e na opção *Models* o modelo construído.

Para a predição da estrutura tridimensional de uma proteína é importante uma etapa de estimativa da qualidade da estrutura. São fornecidas informações como *escore GMQE (Global Model Quality Estimation)*, um número entre 0 e 1 que reflete a acurácia esperada no modelo construído baseado em propriedades do alinhamento entre a sequência alvo e o modelo (*template*).

Quanto maior o valor, maior a confiabilidade no modelo. O QMEAN (*Qualitative Model Energy Analysis*) (BENKERT, TOSATTO e SCHOMBURG, 2008) é uma função composta de pontuação que descreve a maioria dos aspectos geométricos da estrutura de proteínas.

A estimativa da qualidade dos modelos construídos foi efetuada na ferramenta QMEAN (<https://swissmodel.expasy.org/qmean/>) (BENKERT, KÜNZLI e SCHWEDE, 2009) e para o refinamento da estrutura construída foi utilizado o servidor 3D<sup>refine</sup> (<http://sysbio.rnet.missouri.edu/3Drefine/>) (BHATTACHARYA et al., 2016; BHATTACHARYA E CHENG, 2013; BHATTACHARYA E CHENG, 2012).

A avaliação dos resíduos de aminoácidos dos modelos refinados foi efetuada por meio da análise do gráfico de Ramachandran com o auxílio do programa RAMPAGE (<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>) (LOVELL et al., 2002).

## Resultados e Discussão

Uma vez escolhida alguma condição genética relacionada com vias metabólicas a serem abordadas no estudo de bioquímica, seleciona-se uma proteína relacionada à essa condição. Neste trabalho optou-se pela modelagem da estrutura tridimensional da enzima UGT1A1, relacionada à síndrome de Gilbert, como modelo didático, pois esta enzima ainda não possui sua estrutura tridimensional determinada experimentalmente. A partir da sequência de aminoácidos da proteína UGT1A1 foi possível acessar uma série de informações estruturais por meio da utilização de ferramentas de bioinformática. A sequência foi submetida à ferramenta Protter (OMASITS et al., 2014) para análise de topologia e determinação de estruturas transmembrana e peptídeo sinal.

Os resultados mostraram que os resíduos 1 a 25 correspondem a uma sequência de peptídeo-sinal e os resíduos 491 a 507 correspondem a uma hélice transmembrana.

Pedagogicamente, por meio desta análise poderão ser abordados tópicos relacionados à função do peptídeo sinal como marcador de proteínas para exportação para determinados locais celulares onde exercerão sua função. Também, é possível a exploração das características da estrutura de uma  $\alpha$ -hélice transmembrana, destacando a predominância de resíduos de aminoácidos com cadeia lateral hidrofóbica (alanina, valina, isoleucina e fenilalanina) e as suas interações com a bicamada lipídica por interações hidrofóbicas e forças de van de Waals.

A topologia predita da proteína UGT1A1 está de

acordo com o fato das uridinas difosfato-glicuronil-transferases estarem localizadas no retículo endoplasmático, ancoradas à membrana. O domínio transmembrana está localizado próximo à extremidade C-terminal, sendo que apenas uma pequena porção da proteína está voltada para o citosol. A maior parte da proteína está localizada no lúmen do retículo endoplasmático, incluindo o domínio de ligação aos substratos e o sítio catalítico (MEECH e MACKENZIE, 1997).

Para a análise da estrutura secundária foram alinhadas as predições de estruturas secundárias de vários métodos, dando origem à uma estrutura consenso contendo 37,2% da sequência (173 aminoácidos) em  $\alpha$ -hélice, 17% (79 aminoácidos) em cadeia estendida (conformação  $\beta$ ), 43,2% (201 aminoácidos) de estrutura aleatória e 2,6% (12 aminoácidos) de estrutura ambígua (Figura 2). Para o alinhamento, as sequências referentes ao peptídeo sinal (resíduos 1 a 25) e a hélice transmembrana e porção C-terminal (resíduos 491 a 533) foram removidas.

No ensino de estrutura de proteínas, a predição da estrutura secundária a partir da sequência de aminoácidos contribui para o entendimento do estudante acerca do significado do termo estrutura secundária, descrita como o arranjo espacial local dos átomos da cadeia principal, sem considerar a conformação de suas cadeias laterais (NELSON e COX, 2008). Nesse sentido, é possível explorar os tipos de estrutura secundária, em especial  $\alpha$ -hélice e conformação  $\beta$  e os tipos de interações intermoleculares que as mantêm.

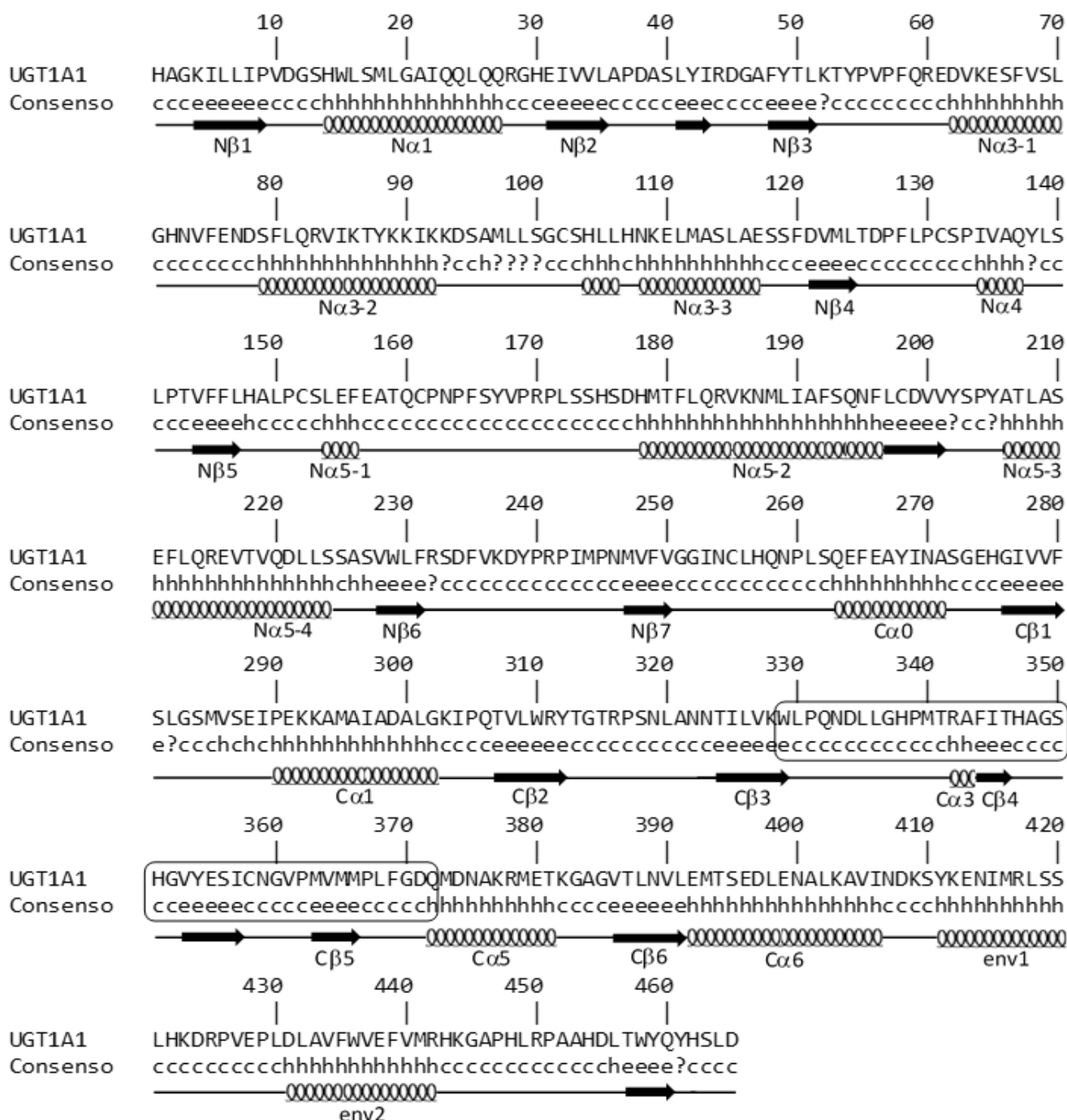
Outro nível hierárquico de estrutura de proteínas é a terciária. O conhecimento da estrutura terciária representa uma informação importante para a determinação de sua função. Atualmente, a estrutura terciária é determinada com técnicas experimentais como a ressonância magnética nuclear (RMN) e cristalografia por difração de raio-X. Entretanto, essas metodologias requerem um certo tempo e são de difícil execução, além de serem limitadas para várias proteínas, as quais são difíceis de serem purificadas ou cristalizadas.

A modelagem molecular é um método alternativo e não experimental que permite prever a conformação de proteínas a partir de suas sequências de aminoácidos (estrutura primária). Esse método se baseia em conhecimentos da estereoquímica de aminoácidos e nas informações das estruturas terciárias já resolvidas e depositadas em banco público de estruturas (PROSDOCIMI et al., 2012).

A predição da estrutura tridimensional de proteínas a partir de sua estrutura primária é determinada pelo método de modelagem conhecido como *ab initio*, no qual é realizado o cálculo da energia envolvida no

processo de enovelamento até encontrar-se a estrutura com menor energia livre. Este princípio é baseado na hipótese termodinâmica na qual a estrutura nativa de uma proteína é aquela com menor energia livre global (RICHARD e DAVID, 2001).

**Figura 2** - Predição da estrutura secundária da proteína UGT1A1 a partir do alinhamento de diferentes métodos.

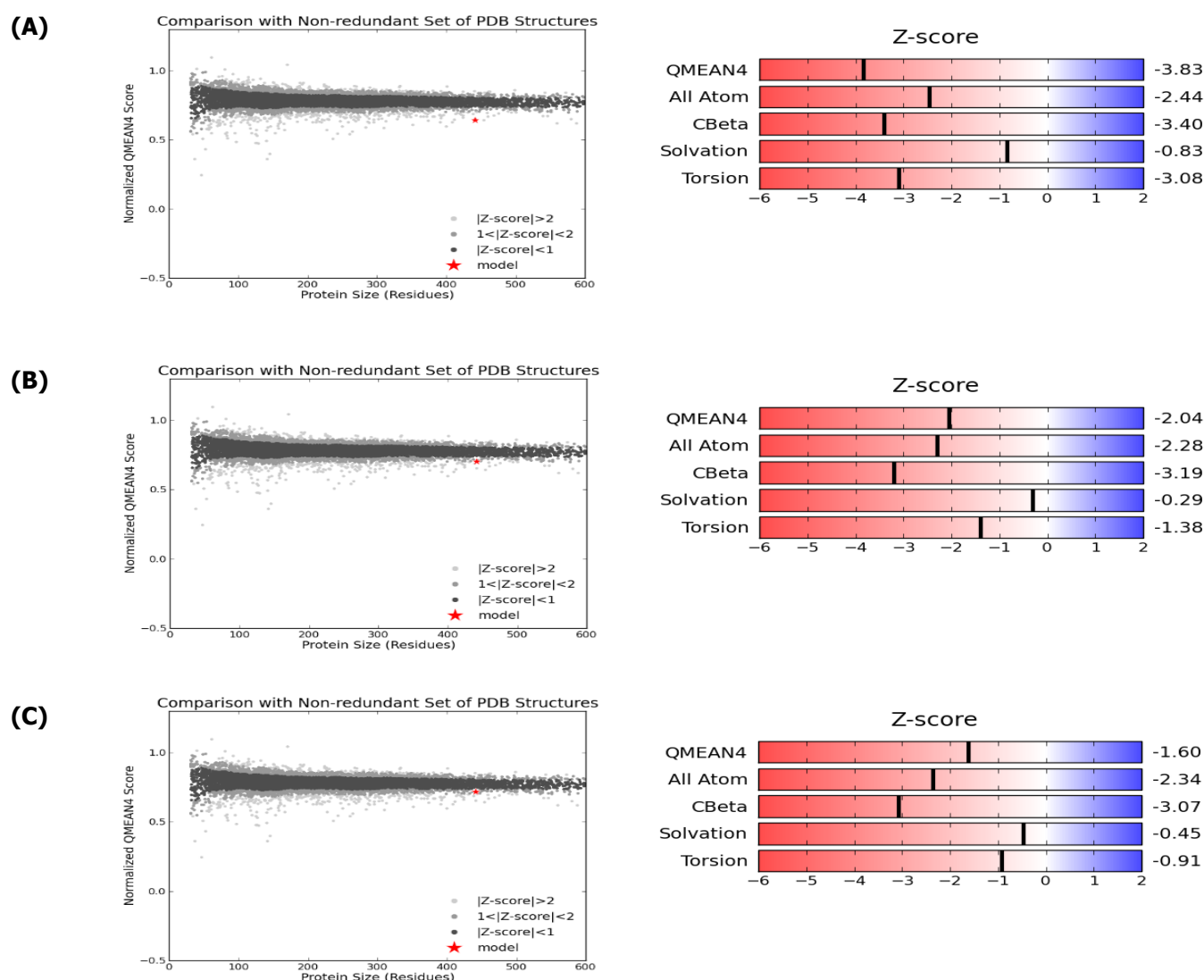


h, α-hélice; e, cadeia estendida (conformação β); c, estrutura aleatória; ?, estrutura ambígua. A sequência na caixa indica a assinatura UDP-glicosil-transferase (Prosite). Abaixo das estruturas secundárias são identificadas as α-hélices e conformações β pertencentes às subestruturas C-terminal (C), N-terminal (N) e envelope de hélices (env) coincidentes com aquelas descritas por LAAKKONEN e FINEL (2010).  
Fonte: Servidor NPS@ (COMBET et al., 2000).

A proteína estudada foi identificada como pertencente à família UDP-glicuronil/UDP-glicuronil-transferase (IPR002213). UDP-glicuronil-transferase (UGT) é uma superfamília de enzimas que catalisa a adição de grupos glicosil a partir de um UTP-açúcar em uma molécula hidrofóbica.

Para a construção do modelo foi utilizada a sequência de aminoácidos da proteína nativa, com remoção dos segmentos 1-25 (peptídeo sinal) e 491 a 533 (hélice transmembrana e extremidade C-terminal). Para esta sequência resultante de 465 aminoácidos foram encontradas 294 estruturas molde. Foi selecionado um modelo que utilizou como molde a estrutura da proteína oleandomicina glicosiltransferase de *Streptomyces antibioticus* (2iya.1.A) (BOLAM et al., 2007). Esta proteína é um monômero, capaz de ligar uridina-5'-difosfato (UDP) a oleandomicina, e teve a sua estrutura determinada por difração de raio-X 1,70 Å.

**Figura 3** - Estimativa da qualidade dos modelos bruto e refinado da proteína UGT1A1 isoforma 1 humana.



(A) UGT1A1 original; (B) Modelo refinado 1 e (C), Modelo refinado 5. Medida da qualidade absoluta dos modelos – desvio apresentado pelos Z-scores dos modelos em relação à população de estruturas resolvidas. Quanto mais próximo da região preta, melhor o modelo. A estrela vermelha representa a posição dos modelos preditos; (B), Z-score individual para cada termo presente na função QMEAN, referentes às estruturas preditas. Estruturas que apresentam valores fortemente negativos são consideradas não favoráveis. A média de Z-score de estruturas cristalográficas de alta resolução são próximas de zero (BENKERT, TOSATTO e SCHOMBURG, 2008; BENKERT, BIASINI e SCHWEDE, 2011).

Fonte: Swiss Model (BIASINI et al., 2014; KIEFER et al., 2009; GUEX, PEITSCH e SCHWEDE, 2009; ARNOLD et al., 2006)

A identidade entre as sequências foi de 18,64%, com uma cobertura variando entre os aminoácidos 3 e 443. Os valores de GMQE para o modelo construído foi de 0,51 e QMEAN de -3,81. Esta estrutura molde foi apontada pelo programa como a melhor para a construção de um modelo por homologia.

Os modelos precisam ser avaliados e validados quanto à sua qualidade estereoquímica, energia livre do sistema, acessibilidade do solvente, entre outros parâmetros. O GMQE (*Global Model Quality Estimation*) é uma estimativa da acurácia da estrutura terciária do modelo construído e varia de 0 a 1, onde quanto maior o valor maior a confiabilidade no modelo (BIASINI et al., 2014).

O parâmetro QMEAN (*Qualitative Model Energy Analysis*) é uma função que descreve os principais aspectos geométricos da estrutura proteica. É composta por cinco aspectos: i) geometria local, que é analisada pelo ângulo de torção ao longo de três aminoácidos consecutivos, ii) interações de longo alcance, iii) potencial de solvatação que descreve o estado de enterramento dos resíduos, iv) o acordo entre a estrutura secundária predita e a calculada, e v) acessibilidade de solvente (BENKERT, TOSATTO e SCHOMBURG, 2008).

O parâmetro Z-score compara o modelo construído com outras estruturas de mesmo tamanho depositadas no banco público de estruturas de proteínas PDB (*Protein Data Bank*), sendo ideal que seja um valor negativo. No gráfico de estimativa da qualidade absoluta (Figura 3A, B e C, lado esquerdo), os pontos coloridos em escala de cinza representam escores do QMEAN para outras estruturas de tamanho similares depositadas no PDB, com os quais o escore do modelo (estrela vermelha) é comparado. O Z-score do QMEAN indica quantos desvios padrões o escore difere dos valores esperados de estruturas experimentais. Nesse caso, quanto mais próximo da parte preta melhor (menor Z-score, ou seja, menor desvio padrão). Também são mostrados o Z-score de termos individuais, permitindo a identificação das características geométricas responsáveis pelos baixos valores (Figura 3A, B e C, lado direito).

A função global de pontuação QMEAN4 (BENKERT et al., 2008) (Figura 3) é uma combinação de quatro características estruturais utilizando potenciais estatísticos: a geometria local é analisada por meio de um potencial de ângulo de torção a cada três aminoácidos consecutivos. Dois potenciais de interação dependentes da distância são usados para avaliar as interações de longo alcance, sendo a primeira a nível de resíduos com base apenas em átomos de C-beta e a segunda um potencial de todos os átomos que é capaz de capturar mais detalhes do modelo. Um potencial de solvatação investiga o estado de enterramento dos resíduos. QMEAN4 é uma pontuação de confiabilidade para todo o modelo que pode ser usada para comparar e classificar modelos alternativos da mesma sequência. A estimativa de qualidade varia entre 0 e 1 com valores mais elevados para melhores modelos.

Após o refinamento do modelo construído, o servidor 3D<sup>refine</sup> originou cinco modelos, dos quais os modelos Refinamento 1 e 5 apresentaram os melhores parâmetros e foram analisados por meio do gráfico de Ramachandran (Tabela 1).

**Tabela 1** - Parâmetros de análise dos modelos refinados.

Modelo	Gráfico de Ramachandran				3D <sup>refine</sup>					
	Regiões favoráveis (%)	Regiões permitidas (%)	Regiões periféricas (%)	Aminoácidos periféricos	3D <sup>refine</sup> score <sup>1</sup>	GDT-TS <sup>2</sup>	GDT-HA <sup>3</sup>	RMS D (Å) <sup>4</sup>	MolProbability <sup>5</sup>	RWPlus <sup>6</sup>
UGT1A1	90,7	6,6	2,7	12	-	-	-	-	-	-
Refinamento 1	93,4	5,2	1,4	6	3103 7,7	1,00 00	1,000 0	0,14 3	2,241	-96775,747139
Refinamento 2	94,1	4,6	1,4	6	2566 6,8	1,00 00	0,999 4	0,19 3	2,297	-96948,329457
Refinamento 3	93,8	4,8	1,4	6	2496 6,6	1,00 00	0,998 3	0,22 6	2,280	-97055,484094
Refinamento 4	93,4	5,2	1,4	6	2459 5,9	1,00 00	0,996 0	0,25 0	2,346	-97239,025580
Refinamento 5	93,3	5,5	1,4	6	2439 2,8	1,00 00	0,994 9	0,27 0	2,280	-97306,072325

<sup>1</sup>  $3D^{\text{refine}}$  score – representa a energia potencial do modelo refinado. Valores menores indicam modelos de qualidade melhor;

<sup>2,3</sup> GDT-TS (*Global Distance Test\_TotalScore*) e GDT-HA (*GDT-High Accuracy*) – são escores de similaridade do modelo refinado em relação ao modelo inicial. Valores maiores indicam refinamento conservativo;

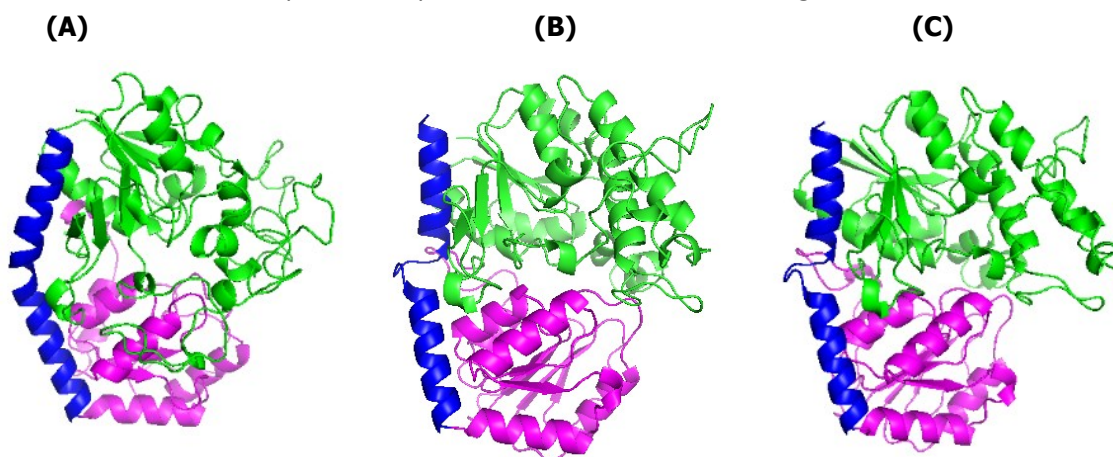
<sup>4</sup> RMSD (Å) – é o escore de desvio (em Å) do CA-RMS (*Global Distance Calculation for side chains*) do modelo refinado em relação ao modelo original. Valores altos indicam um refinamento agressivo;

<sup>5</sup> MolProbity – escore de realismo físico do modelo refinado. Valores menores indicam modelos mais realísticos;

<sup>6</sup> RWPlus – é a energia potencial do modelo refinado. Valores menores geralmente indicam modelos de melhor qualidade.

Fonte: Autores (2018).

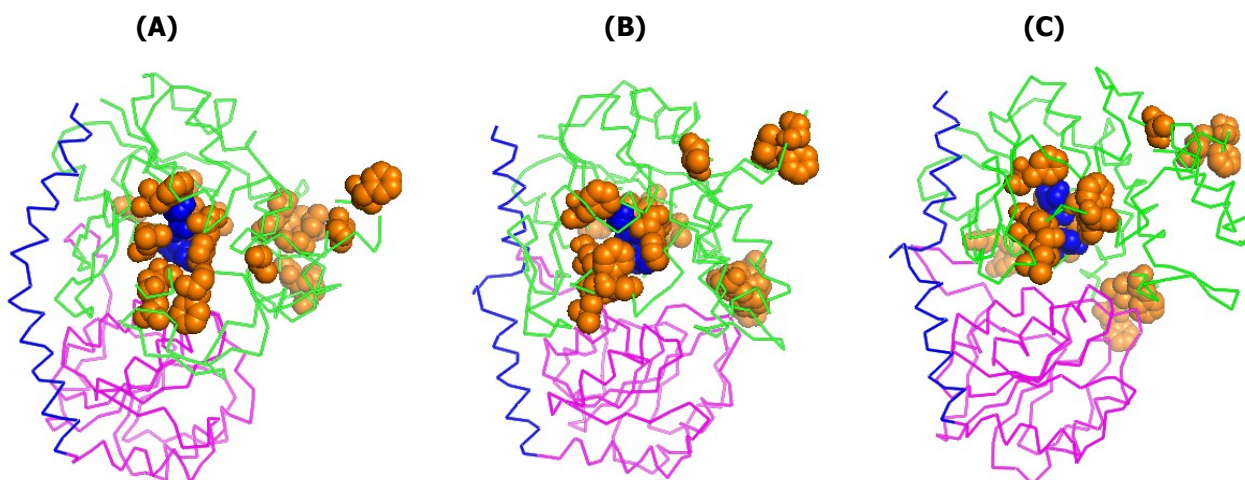
**Figura 4** - Estrutura tridimensional predita da proteína UGT1A1 isoforma 1 segmento 26-490.



A, estrutura preliminar construída pelo servidor SWISS-MODEL; B e C, estruturas após refinamento efetuado pelo servidor  $3D^{\text{refine}}$  (Refinamentos 1 e 5, respectivamente). Os domínios estão destacados em diferentes cores: N-terminal, verde; C-terminal, magenta; envelope de hélices, azul.

Fonte: SWISS-MODEL e  $3D^{\text{refine}}$

**Figura 5** - Estrutura tridimensional predita da proteína UGT1A1 isoforma 1 segmento 26-490.



A, estrutura preliminar construída pelo servidor SWISS-MODEL; B e C, estruturas após refinamento efetuado pelo servidor  $3D^{\text{refine}}$  (Refinamentos 1 e 5, respectivamente). Os domínios estão destacados em diferentes cores: N-terminal, verde; C-terminal, magenta; envelope de hélices, azul. Em laranja estão representados pelo modelo de bolas os aminoácidos hidrofóbicos do bolso de acomodação do substrato e em azul os resíduos de aminoácidos hidrofílicos (H39 e D151) que supostamente interagem com o substrato.

Fonte: RasMol versão 2.7.5.2.



Em termos didáticos, a análise das informações provenientes dos gráficos de Ramachandran permitem o estudo da conformação do peptídeo pelos valores dos ângulos  $\phi$  e  $\psi$ , onde são apresentadas as conformações possíveis em função de impedimentos estéricos, com base em cálculos usando raios de van der Waals e ângulos diedros (NELSON e COX, 2008).

A bilirrubina existe em duas configurações em solução aquosa, (Z,E)-bilirrubina e (Z,Z)-bilirrubina (NOGALES e LIGHTNER, 1995). Para a sua glicuronidação, a UGT1A1 humana catalisa a transferência de ácido glicurônico do UDP-ácido glicurônico para os grupos propionato C8 e C12. De acordo com o modelo proposto por Li e Wu (2007), a bilirrubina é encaixada em um bolso hidrofóbico com o anel porfirina I em uma extremidade e o anel porfirina IV na outra extremidade do acceptor N-terminal do bolso. As cadeias laterais propionato estão localizadas na região central do bolso, próximas ao ácido glicurônico da molécula doadora e próximos a um resíduo de histidina (H39), altamente conservado. O bolso hidrofóbico é formado pelos resíduos P34, V35, A64, L66, Y67, G71, F92, V93, G96, V99, F100, F153, F170, L172, A174, L175, F181, F221, V225 e A231.

Assim, nos modelos construídos, os resíduos hidrofóbicos do bolso de acomodação do substrato foram destacados, assim como aqueles envolvidos na glicuronidação da bilirrubina (Figura 5).

Estes modelos mostram a importância da conformação de uma proteína para a constituição do sítio ativo, bem como a contribuição dos resíduos de aminoácidos diretamente envolvidos na catálise, por meio de importantes interações intermoleculares.

Existem várias limitações e possíveis fontes de erros na modelagem de estrutura baseada no método da homologia. Dentre eles, destacam-se erros de modelagem das cadeias laterais, quando há baixo grau de identidade entre a estrutura em análise e a estrutura molde. Além disso, erros em regiões sem equivalência na sequência molde, erros de alinhamento, principalmente quando o grau de identidade entre as sequências é inferior a 30% e moldes incorretos, mais comum quando o grau de identidade é menor que 25%, podem causar distorções entre as estruturas real e modelada (SANTOS FILHO e ALENCASTRO, 2003).

Na modelagem da estrutura da proteína UGT1A1 deste trabalho utilizou-se uma proteína molde com identidade na sequência de aminoácidos menor que 20%, além de se tratar de uma estrutura-molde determinada por difração de raio-X 1,70 Å, uma vez que são preferíveis modelos com resolução superior ou igual a 2,0 Å, pois a qualidade do modelo depende da estrutura-molde utilizada (SANTOS FILHO e ALENCASTRO, 2003).

Os parâmetros de qualidade do modelo construído corroboram essas informações e sugerem que o modelo pode apresentar alguns erros, limitado pela existência de estruturas-molde com maior homologia.

Entretanto, segundo Torres e Galembeck (2009), no ensino de estrutura de proteínas exige-se que o estudante faça uma representação mental tridimensional de um conteúdo apresentado em livros ou projetado em lousa em duas dimensões.

Desta maneira, a modelagem de estruturas com o uso de ferramentas de bioinformática como proposta pedagógica para o estudo de estrutura de proteínas pode contribuir de forma decisiva para o processo ensino-aprendizagem, uma vez que exige um comportamento ativo de alunos e professores, permitindo uma aproximação maior do estudante com a temática e estimulando a construção de conhecimentos mais sólidos acerca da estrutura de proteínas.

O uso adequado de estratégias de ensino implica na seleção intencional e diretiva das metodologias que melhor funcionam como mediadoras e facilitadoras do processo de ensino-aprendizagem. A escolha de metodologias que despertem o interesse e a curiosidade dos alunos pelos processos bioquímicos pode fazer uma grande diferença na relação professor-aluno e no processo de aprendizagem (TORRES e GALEMBECK, 2009).

A importância da utilização da modelagem da estrutura tridimensional de proteínas a partir da sequência de aminoácidos no ensino reside muito mais na compreensão das características das estruturas primária, secundária, terciária e quaternária, no conceito e visualização do sítio ativo, na identificação de resíduos importantes e na compreensão das interações intermoleculares e dos ângulos de torção de uma cadeia polipeptídica, que na obtenção de modelos perfeitos e o entendimento dos parâmetros de qualidade de modelos construídos.

Considerando que o estudo da bioquímica exige conhecimentos prévios de química e biologia e o evidente déficit em conhecimentos básicos acumulado pelos estudantes que acessam o ensino superior, ferramentas facilitadoras e que aproximem o ensino à realidade do estudante podem contribuir consideravelmente para um incremento no processo ensino-aprendizagem.

### Considerações Finais

A modelagem de proteínas por meio de ferramentas de bioinformática representa uma técnica acessível, de baixo custo e não experimental que permite um estudo aprofundado de proteínas. Apesar de não

existir um *software* que apresente um resultado totalmente correto, muito se tem evoluído, principalmente para regiões sem moldes para comparação.

Como ferramenta pedagógica, a modelagem da estrutura de proteínas pode contribuir para o estudo dos diferentes níveis de estrutura, bem como na compreensão de conhecimentos básicos necessários para o estudo de proteínas. O modelo da proteína utilizada nesse trabalho (UGT1A1), apesar de apresentar parâmetros de qualidade que sugiram erros estruturais, representou um modelo pedagógico viável, uma vez que está inserida em um contexto bioquímico importante, relacionado a algumas patologias, além de apresentar peptídeo sinal, um domínio transmembrana, ser uma proteína globular com diferentes tipos de estrutura secundária e sítio ativo com o envolvimento de interações intermoleculares de diferentes resíduos de aminoácidos. Esse modelo permite a exploração dos principais tópicos do estudo de estrutura tridimensional de proteínas, abordados na disciplina de Bioquímica, uma componente curricular básica de diferentes cursos de graduação.

## Referências

- ARNOLD, K.; BORDOLI, L.; KOPP, J.; SCHWEDE, T. The SWISS-MODEL Workspace: A web-based environment for protein structure homology modeling. **Bioinformatics**, v.22, p.195-201, 2006.
- BENKERT, P.; BIASINI, M.; SCHWEDE, T. Toward the estimation of the absolute quality of individual protein structure models. **Bioinformatics**, v.27, n.3, p.343-50, 2011.
- BENKERT, P.; KÜNZLI, M.; SCHWEDE, T. QMEAN Server for Protein Model Quality Estimation. **Nucleic Acids Res.**, v.37(Web Server issue), p.W510-4, 2009.
- BENKERT, P.; TOSATTO, S.C.E.; SCHOMBURG, D. QMEAN: A comprehensive scoring function for model quality assessment. **Proteins: Structure, Function, and Bioinformatics**, v.71, n.1, p.261-277, 2008.
- BESA, S.; CALVO, C.I.; HARRIS, P.R. Evolución prolongada em síndrome de Crigler-Najjar tipo I. **Rev. Med. Chile**, v.142, n.1, p.109-113, 2014.
- BHATTACHARYA, D., CHENG, J. [3D<sup>refine</sup>: Consistent Protein Structure Refinement by Optimizing Hydrogen Bonding Network and Atomic Level Energy Minimization](#). **Proteins: Structure, Function, and Bioinformatics**, v.81, n.1, p.119-131, 2012.
- BHATTACHARYA, D., CHENG, J. [i3D<sup>refine</sup> software for protein 3D structure refinement and its assessment in CAS-P10](#). **PLOS ONE**, v.8, n.7, p.e69648, 2013.
- BHATTACHARYA, D., NOWOTNY, J., CAO, R., CHENG, J. [3D<sup>refine</sup>: an interactive web server for efficient protein structure refinement](#). **Nucleic Acids Research, Web Server Issue**, 2016.
- BIASINI, M.; BIENERT, S.; WATERHOUSE, A.; ARNOLD, K.; STUDER, G.; SCHMIDT, T.; KIEFER, F.; CASSARINO, T.G.; BERTONI, M.; BORDOLI, L.; SCHWEDE, T. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information **Nucleic Acids Research**, v.42(W1), p.W252-W258, 2014.
- BOLAM, D.N.; ROBERTS, S.; PROCTOR, M.R.; TURKENBURG, J.P.; DODSON, E.J.; MARTINEZ-FLEITES, C.; YANG, M.; DAVIS, B.G.; DAVIES, G.J.; GILBERT, H.J. The Crystal Structure of Two Macrolide Glycosyltransferases Provides a Blueprint for Host Cell Antibiotic Immunity. **Proc.Natl.Acad.Sci.**, v.104, n.13, p.5336-41, 2007.
- CHOTHIA, C.; LESK, A.M. The relation between the divergence of sequence and structure in proteins. **The EMBO Journal**, v.5, n.4, p.823-826, 1986.
- COMBET, C.; BLANCHET, C.; GEOURJON, C.; DELÉAGE, G. NPS@: Network Protein Sequence Analysis. **TIBS**, v.25, n.3, p.147-150, 2000.
- DELEAGE, G.; ROUX, B. An algorithm for protein secondary structure prediction based on class prediction. **Protein Eng.**, v.1, n.4, p.289-294, 1987.
- FRISHMAN, D.; ARGOS, P. Incorporation of non-local interactions in protein secondary structure prediction from the amino acid sequence. **Protein Eng.**, v.9, n.2, p.133-142, 1996.
- GARNIER, J.; OSGUTHORPE, D.J.; ROBSON, B. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. **J. Mol. Biol.**, v.120, n.1, p.97-120, 1978.
- GEOURJON, C.; DELEAGE, G. SOPM: a self-optimized method for protein secondary structure prediction. **Protein Eng.**, v.7, n.2, p.157-164, 1994.
- GIBRAT, J.F.; GARNIER, J.; ROBSON, B. Further developments of protein secondary structure prediction using information theory. New parameters and consideration of residue pairs. **J. Mol. Biol.**, v.198, n.3, p.425-443, 1987.
- GUERMEUR, J.; OSGUTHORPE, D.J.; ROBSON, B. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. **J. Mol. Biol.**, v.120, n.1, p.97-120, 1978.
- GUERMEUR, Y. PhD Thesis - Combinaison de classifieurs statistiques, Application a la prediction de structure secondaire des proteines, 1997.
- GUEx, N.; PEITSCH, M.C.; SCHWEDE, T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. **Electrophoresis**, v.30, n.S1, p.S162-S173, 2009.
- KIEFER, F.; ARNOLD, K.; KÜNZLI, M.; BORDOLI, L.

- SCHWEDE, T. The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res.*, v.37, p.D387-D392, 2009.
- KING, R.D.; STERNBERG, M.J. Identification and application of the concepts important for accurate and reliable protein secondary structure prediction. *Protein Sci.*, v.5, n.11, p.2298-310, 1996.
- LAACKONEN, L.; FINEL, M. A molecular model of the human UDP-glucuronosyltransferase 1A1, its membrane orientation, and the interactions between different parts of the enzyme. *Mol. Pharmacol.*, v.77, n.6, p.931-9, 2010.
- LI, C.; WU, Q. Adaptive evolution of multiple-variable exons and structural diversity of drug-metabolizing enzymes. *BMC Evolutionary Biology*, v.7, n.69, p.1-20, 2007.
- LOVELL, S.C.; DAVIS, I.W.; ARENDALL III, W.B.; DE BAKKER, P.I.W.; WORD, J.M.; PRISANT, M.G.; RICHARDSON, J.S.; RICHARDSON, D.C. Structure validation by C $\alpha$  geometry: phi,psi and C $\beta$  deviation. *Proteins: Structure, Function & Genetics*, v.50, p.437-450, 2002.
- MACEDO, K.D.S.; ACOSTA, B.S.; SILVA, E.B.; SOUZA, N.S.; BECK, C.L.C.; SILVA, K.K.D. Metodologias ativas de aprendizagem: caminhos possíveis para inovação no ensino em saúde. *Esc. Anna. Nery*, v.22, n.3, p.e20170435, 2018.
- MEECH, R.; MACKENZIE, P.I. Structure and function of uridine diphosphate glucuronosyltransferases. *Clin. Exp. Pharmacol. Physiol.*, v.24, n.12, p.907-915, 1997.
- NELSON, D.L.; COX, M.M. Principles of Biochemistry. 5 ed. New York: W H Freeman & Co, 2008. 1158p.
- NOGALES, D.; LIGHTNER, D.A. On the structure of bilirubin in solution.  $^{13}\text{C}[1\text{H}]$  heteronuclear Overhauser effect NMR analyses in aqueous buffer and organic solvents. *J. Biol. Chem.*, v.270, p.73-77, 1995.
- OMASITS, U.; AHRENS, C.H.; MÜLLER, S.; WOLLSCHIED, B. Protter: interactive protein feature visualization and integration with experimental proteomic data. *Bioinformatics*, v.30, n.6, p.884-6, 2014.
- PROSDOCIMI, F.; CERQUEIRA, G.C.; BINNECK, E.; SILVA, A.F.; REIS, A.N.; JUNQUEIRA, A.C.M.; SANTOS, A.C.F.; NHANI JÚNIOR, A.; WUST, C.I.; CAMARGO FILHO, F.; KESSEDIAN, J.L.; PETRETSKI, J.H.; CAMARGO, L.P.; FERREIRA, R.G.M.; LIMA, R.P.; PEREIRA, R.M.; JARDIM, S.; SAMPAIO, V.S.; FOLGUERAS-FLATSCHART, A.V. Bioinformática: manual do usuário. *Biociência* Ciência & Desenvolvimento, v.29, p.12-25, 2012.
- RAMOS, A.F.; SERRANO, A. Modelagem Molecular no Ensino de Ciências: uma revisão de literatura no período de 2001-2011 acerca de sua aplicabilidade em atividades de ensino. *Actia Scientiae*, v.15, n.2, p.363-382, 2013.
- RICHARD, B.; DAVID, B. *Ab initio* protein structure prediction: progress and prospects. *Annual review of biophysical and biomolecular structures*, v.30, p.73-88, 2001.
- RITTER, J.K.; CHEN, F.; SHEEN, Y.Y.; TRAN, H.M.; KIMURA, S.; YEATMAN, M.T.; OWENS, I.S. A novel complex locus UGT1 encodes human bilirubin, phenol and other UDP-glucuronosyltransferase isozymes with identical carboxy termini. *J. Biol. Chem.*, v.267, p.3257-61, 1992.
- RODWELL, V.W.; BENDER, D.A.; BOTHAM, K.M.; KENNELLY, P.J.; WEIL, P.A. Bioquímica ilustrada de Harper. 30 ed. Porto Alegre: AMGH, 2017, 832p.
- ROST, B.; SANDER, C. Prediction of protein secondary structure at better than 70% accuracy. *J. Mol. Biol.*, v.232, n.2, p.584-99, 1993.
- SABINO, G.; AMARAL, F.C.; SABINO, C.V.S.; KATTAH, L.R. Proposta de uma metodologia para o ensino da estrutura e função das proteínas na disciplina de Bioquímica. *Rev. Bras. Ens. Bioq. Biol. Mol.*, n.1, p.X1-x19, 2009.
- SANTOS FILHO, O.A.; ALENCASTRO, R.B. Modelagem de proteínas por homologia. *Química Nova*, v.26, n.2, p.253-259, 2003.
- SIGRIST, C.J.A.; CERUTTI, L.; HULO, N.; GATTIKER, A.; FALQUET, L.; PAGNI, M.; BAIROCH, A.; BUCHER, P. PROSITE: a documented database using patterns and profiles as motif descriptors. *Brief. Bioinform.*, v.3, p.265-274, 2002.
- THE UNIPROT CONSORTIUM. UniProt: the universal protein knowledgebase. *Nucleic Acids Research*, v.45, p.D158-D169, 2017.
- TORRES, B.B.; GALEMBECK, E. Proposta de uma metodologia para o ensino da estrutura e função das proteínas na disciplina de bioquímica. *Revista Brasileira de Ensino de Bioquímica e Biologia Molecular*, v.1, p.X1-X19, 2009.
- U.S. NATIONAL LIBRARY OF MEDICINE (Estados Unidos) (Org.). **Fact Sheet:** Genetics Home Reference. 2003. Disponível em: <<https://www.nlm.nih.gov/pubs/factsheets/ghr.html>>. Acesso em: 09 jan. 2017.
- VANWAGNER, L.B.; GREEN, R.M. Evaluating elevated bilirubin levels in asymptomatic adults. *JAMA*, v.313, n.5, p.516-517, 2015.
- WANG, X.; CHOWDHURY, J.R.; CHOWDHURY, N.R. Bilirubin metabolism: Applied physiology. *Current Paediatrics*, v.16, p.70-74, 2006.

## CURRÍCULO

\* Professor do Instituto Federal Catarinense. Possui graduação em Química Industrial (2000), mestrado (2003) e doutorado (2007) em Ciências (Bioquímica) pela Univer-

sidade Federal do Paraná e Pós-Doutorado (2010) no Departamento de Química e Ciências Biológicas da Universidade de Nova York (NYU). Também atuou como professor colaborador da Universidade do Estado de Santa Catarina-UDESC. Tem experiência na área de Biotecnologia, com ênfase em Biologia Molecular e química de macromoléculas.

\*\* Possui graduação em Medicina Veterinária pela Universidade Federal do Rio Grande do Sul (1994), mestrado em Ciência Animal pela Universidade Estadual de Londrina (2002) e doutorado em Ciência Animal pela Universidade Estadual de Londrina (2008). Atualmente é professora do Instituto Federal Catarinense - Campus Araquari. Coordenadora do Programa de Extensão "MIA - Movimento pela interação Animal" que desenvolve Intervenções Assistidas por Animais (IAA). Atua na área de virologia animal e diagnóstico molecular.

\*\*\* Possui graduação em Medicina Veterinária pela Universidade Federal do Paraná (1992), mestrado em Ciências Veterinárias pela Universidade Federal do Paraná (1999) e doutorado em Processos Biotecnológicos pela Universidade Federal do Paraná (2008). Atualmente é Professor do Ensino Básico Técnico e Tecnológico do Instituto Federal Catarinense - campus Araquari. Tem experiência na área de Medicina Veterinária, com ênfase em Produção, Medicina e Reprodução de Pequenos Ruminantes e Doenças Parasitárias de Animais, atuando principalmente nos seguintes temas: doenças parasitárias dos animais domésticos, métodos alternativos de controle de verminose em ovinos e caprinos, doenças de pequenos ruminantes. Atualmente é responsável pelo Laboratório de Parasitologia do Laboratório de Ensino e Diagnóstico em Medicina Veterinária (LEDVET).