



INSTITUTO DE PESQUISAS ENERGÉTICAS E NUCLEARES
Autarquia Associada à Universidade de São Paulo

Bioquímica estrutural da tireotrofina de Arapaima gigas e sua expressão em células de mamífero

RENAN PASSOS FREIRE

Dissertação apresentada como parte dos requisitos para obtenção do Grau de Mestre em Ciências na Área de Tecnologia Nuclear - Aplicações

Orientador:
Prof. Dr. Carlos Roberto Jorge Soares

São Paulo
2022

INSTITUTO DE PESQUISAS ENERGÉTICAS E NUCLEARES
Autarquia Associada à Universidade de São Paulo

Bioquímica estrutural da tireotrofina de *Arapaima gigas* e sua expressão em células de mamífero

RENAN PASSOS FREIRE

Dissertação apresentada como parte dos requisitos para obtenção do Grau de Mestre em Ciências na Área de Tecnologia Nuclear - Aplicações

Orientador: Dr. Carlos Roberto Jorge Soares

Versão corrigida
Versão original disponível no IPEN

São Paulo
2022

Auxílio CAPES: 88887.506371/2020-00

Autorizo a reprodução e divulgação total ou parcial deste trabalho, para fins de estudo e pesquisa, desde que citada a fonte.

FREIRE, R. P. Bioquímica estrutural da tireotrofina de *Arapaima gigas* e sua expressão em células de mamífero. 2022. 133 f. Dissertação (Mestrado em Tecnologia Nuclear) Instituto de Pesquisas Energéticas e Nucleares – IPEN-CNEN/SP. São Paulo.

(Dados Internacionais de Catalogação na Publicação – CIP)
Catalogação na fonte

Freire, Renan Passos

F866b

Bioquímica estrutural da tireotrofina de *Arapaima gigas* e sua expressão em células de mamífero. – São Paulo, SP: IPEN, 2022.
133 f.

Orientador: Prof. Dr. Carlos Roberto Jorge Soares.

Dissertação (Mestrado) – Programa de Pós-Graduação em Tecnologia Nuclear (Aplicações) – Instituto de Pesquisas Energética e Nucleares, São Paulo, 2022.

1. Hormônio estimulante da tireoide (TSH). 2. Pirarucu.
3. AlphaFold. 4. HEK293
I. Soares, orient., Carlos Roberto Jorge. II. Título.

CDD 621.48



Universidade de São Paulo
Instituto de Pesquisas Energéticas e Nucleares

CERTIFICADO DE DEFESA

CERTIFICO, para os devidos fins, que Sr(a). Renan Passos Freire, número USP 9895781, defendeu no dia 19 de setembro de 2022, no Programa de Tecnologia Nuclear do(a) Instituto de Pesquisas Energéticas e Nucleares da Universidade de São Paulo, a Dissertação de Mestrado intitulada:

"Bioquímica estrutural da tireotrofina de *Arapaima gigas* e sua expressão em células de mamífero"

CERTIFICO, ainda, que o(a) aluno(a) faz jus ao título de Mestre em Ciências - Área: Tecnologia Nuclear - Aplicações, tendo sido a ata da Comissão Julgadora homologado pela Comissão de Pós-Graduação, em 26 de setembro de 2022.

São Paulo, 07 de outubro de 2022.

Dedicatória

Mi amor

in memoriam a Maria Helena Ferraz Passos, minha avó, da qual carrego a minha criação e eterno agradecimento por ter feito parte da minha vida.

*na minha infância,
muita das coisas eu não entendia seu significado
curiosamente já falava “mi amor”:
representava inteiramente o seu abraço*

*dos almoços e cuidado na criação
a dor, inevitável das quedas, não era um problema
apenas precisava olhar na janela
que Você estava à disposição,*

*eu entendo que na vida a gente cresce,
as pessoas veem, vão, desaparecem
o sentimento, é diferente
recupera, intensifica, resplandece!*

*e como uma estrela,
seu brilho é eterno
carrego sua luz e espalho por onde for
seus ensinamentos e seu amor.*



*Sorrisos de uma tarde ensolarada.
Fotografia registrada em meados de 2006.*

Agradecimentos

Aos meus pais, **Renato Freire** e **Roseli Aparecida de Passos Freire**, pelo apoio em todos os âmbitos e incentivo a continuar no ambiente acadêmico. São minha base e tenho eterno agradecimento;

A minha irmã **Raquel Passos Freire** pelo companheirismo, conversas e momentos de distração e lazer para aliviar as tensões naturais do dia a dia, sendo fundamental no desenvolvimento do trabalho;

A meu tio **Marcos Paulo de Passos** por me incentivar a buscar o ambiente acadêmico e me orientar na tomada de decisões frente aos caminhos que são impostos, sem você, este trabalho não seria desenvolvido;

Aos meus avôs, **Benedito Aparecido de Passos †**, **Oswaldo Freire †** e **Antônia Teresa Freire** pelo auxílio na criação da maneira que lhes cabiam;

Aos parentes **Oswaldo Luiz Freire**, **Benedito Sérgio de Passos**, **Ana Lucia Renesto**, **Bruna Renesto**, **Ildenor Rosa Silva** e **Luci Pereira da Silva** presentes antes, durante e após a realização do trabalho;

A meu orientador **Carlos Roberto Jorge Soares** pela oportunidade, autonomia e acolhimento dentro do ambiente acadêmico desde minha graduação, tornando a caminhada mais tranquila;

Aos pesquisadores que influenciaram no desenvolvimento deste trabalho, **Paolo Bartolini**, **Cibele Nunes Peroni**, **Patrick Jack Spencer**, **João Ezequiel de Oliveira**, **Miriam Fussae Suzuki**, **Lucas Torati**, **Jorge Hernández**, **Moisés Zorzetto**. Os diálogos, apoio técnico, científico e coorientações foram imprescindíveis para conclusão do trabalho;

A **Amanda Rodrigues Takahara** pelo apoio e estar presente nos momentos de intimidade, diluindo a intensidade dos desafios do dia a dia. Você foi e será parte de toda caminhada;

Aos ex-alunos(as) e amigos(as) que fizeram parte dessa caminhada, **Felipe Douglas Silva**, **Thais Cristina dos Anjos Sevilhano**, **Thais Feitosa Caramori**, **Alissandra Moura Gomes**. Vocês sempre serão parte do processo de conclusão deste trabalho e sou grato por todas as experiências juntas;

As amigas e companheirismo desses que fazem parte do centro de Biotecnologia, **Vanessa Luna Yosidaki**, **Eliana Rosa Lima Filha**, **Larissa Andrade Almeida**, **Stephanie Ângelo Pomim**, **Filipe Menezes Bezerra**, **Amanda Palermo Nunes**, **Valentina Andreatta Dutton**, **Mayelle Maria Paz Dias**, **Ana Cristina Gomes Nascimento**, **Giovanna Dias**, **Leonardo William**, **Daniel Perez Vieira**, **Rafael da Silva Costa**, **Antônio Carlos Junqueira**, **Marlene Dias Lima**, **Maria Aparecida Oliveira da Silva** e **Joana Ferreira da Silva**;

Aos meus amigos mais íntimos, **Leonardo Baggio**, **Lucas Baggio**, **André Campos Machado**, **Gabriel Lencione**, **João Gabriel Morato**, **Levy Amback** pelo companheirismo de sempre!

Epígrafe

*Na tribo dos Uaiás, no sudoeste da Amazônia, nasceu um índio chamado **Pirarucu**, que mais tarde se tornaria um grande e forte guerreiro. **Pirarucu** era filho de Pindarô, chefe dos Uaiás e conhecido por sua bondade.*

***Pirarucu**, porém, não herdou do pai o bondoso coração. O rapaz era cheio de vaidade e egoísmo; vivia se vangloriando de sua força e poder; e ainda reclamava dos deuses.*

*Certa vez, enquanto seu pai visitava outras tribos, **Pirarucu** capturou índios de sua própria tribo e os matou sem motivo algum.*

*Essa maldade foi a gota d'água para Tupã, o mais poderoso dos deuses, que então enviou fortes torrentes de chuva sobre **Pirarucu**, quando este pescava com outros índios. O jovem, porém, não se abalou com as águas que Tupã enviava, e ainda zombou dos deuses com palavras de desprezo.*

*Por isso, Tupã chamou Xandoré, divindade que odeia os homens, e ordenou-lhe que atirasse raios e trovões sobre **Pirarucu**. Ao ver o perigo que corria, o guerreiro tentou escapar, mas foi acertado por um raio bem no coração. Mesmo assim, **Pirarucu** não se arrependeu nem pediu perdão.*

*Foi então que os outros índios viram **Pirarucu**, ainda vivo, ser transformado num gigante e escuro peixe e levado para as profundezas do rio; de onde nunca mais voltou*

Resumo

FREIRE, R. P. **Bioquímica estrutural da tireotrofina de *Arapaima gigas* e sua expressão em células de mamífero.** 2022. 133 f. Dissertação (Mestrado em Tecnologia Nuclear) Instituto de Pesquisas Energéticas e Nucleares – IPEN-CNEN/SP. São Paulo.

Introdução: O Pirarucu (*Arapaima gigas*) é uma espécie de peixe nativa da bacia amazônica que pode alcançar até 200 kg e dois metros de comprimento. Alvo de pesca predatória e modificações ambientais, especialmente político-econômicas, já se nota um relativo declínio da população local e, segundo órgãos fiscalizadores, a espécie está em possível risco de extinção. Há ainda uma dificuldade no manejo em cativeiro, visto sua gama de necessidades nutricionais e comportamentos que ainda são investigados. A tireotrofina (TSH) atua no controle da reprodução em diferentes espécies de peixes. E do ponto de vista ecológico tem sua expressão modificada quando em contato com substâncias poluentes, como metais pesados e agrotóxicos. O TSH é uma glicoproteína de peso molecular variando de 24-28 kDa formada pela ligação não-covalente de duas subunidades alfa e beta, sendo a alfa comum para outros hormônios glicoproteicos, como o FSH, LH e CG. Sua função é bem estabelecida, participando diretamente na síntese dos hormônios tireoidianos (HTs) que atuam no desenvolvimento embrionário, modulação da atividade cerebral, gonadal, hepática, entre outros. **Objetivos:** Obter a subunidade beta do TSH de *Arapaima gigas*, caracterizar sua estrutura *in silico* e expressar em células de mamífero. **Metodologia:** Foram utilizadas metodologias de identificação genômica da subunidade beta *in silico* e construção da estrutura do heterodímero com técnicas de inteligência artificial. Aplicou-se duas ferramentas de predição: gênica (Gnomon) e estrutural (AlphaFold), como também se realizou *in silico* estudos sobre distância evolutiva, dinâmica molecular e propriedades físico-químicas. As informações recuperadas das predições foram em parte confirmadas com PCR de hipófises retiradas de Pirarucus e a expressão de agTSH em células de mamífero HEK293F. **Resultados:** A similaridade da cadeia beta variou de 40,10-81,60% entre 39 espécies alinhadas com a sequência obtida. São 146 aminoácidos com 12 cisteínas (que formam seis pontes dissulfeto) e um único sítio de N-glicosilação conservados. A árvore filogenética construída indica maior grau de parentesco com a ordem elopiformes e clupeiformes. A estrutura do heterodímero é majoritariamente polar, com predominância de folhas-beta na sua estrutura secundária, sendo uma citocina com nó de cistina (CATH 2.10.90.10). Os estudos de dinâmica molecular da interação do TSH com receptor do seu ancestral comum, Aruanã-dourado (*S. formosus*), ambos resolvidos por *AlphaFold*, apontam valores comparáveis ao modelo humano, com diferença de valores relacionados a energia de ligação (ΔG) entre 1,50-8,64%, indicando um caminho para estudos *in vitro* e *in vivo*. Sua expressão, feita de maneira transiente, obteve rendimento de agTSH 25 mg/L, similar a trabalho anterior com agFSH 28 mg/L, quantificados através de técnicas cromatográficas no meio condicionado. **Conclusões:** A região codificante e 3' não traduzida da subunidade beta foi identificada e utilizada juntamente com a subunidade alfa para a produção de agTSH em células HEK293. A metodologia utilizada poderá ser aplicada para que novas proteínas da espécie sejam descritas. Muito embora ainda requeria otimização para aumento da expressão, este trabalho contribuirá para que futuros estudos explorem sua atividade biológica, investigando a eficiência quanto à reprodução e melhoria da capacidade reprodutiva em cativeiro do pirarucu.

Palavras-chave: Hormônio estimulante da tireoide (TSH); Pirarucu; AlphaFold; HEK293

Abstract

FREIRE, R. P. **Structural biochemistry of *Arapaima gigas* thyrotropin and its expression in mammalian cells.** 2022. 133 p. Dissertação (Mestrado em Tecnologia Nuclear) Instituto de Pesquisas Energéticas e Nucleares – IPEN-CNEN/SP. São Paulo.

Introduction: The Pirarucu (*Arapaima gigas*) is a specie of fish native to the Amazon basin that can reach up to 200 kg and two meters in length. Targeted by predatory fishing and environmental changes, especially political-economic, it's already noted a relative decline of the local population and, according to enforcement agencies, the species is at possible risk of extinction. Also remain a difficulty in the management in captivity, given its range of nutritional needs and behaviors that are still being studied. Thyrotropin (TSH) has activity on the in different fish species. From the ecology point of view, your expression changes when in contact with contaminating substances, such as heavy metals and pesticides. TSH is a glycoprotein of molecular weight ranging from 24-28 kDa formed by the non-covalent bonding of two subunits, alpha and beta, once alpha being common to other glycoprotein hormones, such as FSH, LH and CG. Your function is well established, participating directly in the synthesis of thyroid hormones (THs) that act, among other functions, in embryonic development, modulation of brain, gonadal and liver activity. **Objectives:** Obtain the beta subunit of TSH from *Arapaima gigas*, characterize the structure *in silico* and express in mammalian cells. **Methodology:** We used methodologies of *in silico* genomic identification of the beta subunit and construction of the heterodimer structure with artificial intelligence techniques. Two prediction tools were applied: genomic (Gnomon) and structural (AlphaFold), as well as explored *in silico* with evolutionary distance, molecular dynamics and physicochemical properties studies. The information recovered from the predictions was partly confirmed with PCR reaction of pituitary glands taken from Pirarucus and the expression of agTSH in HEK293F mammalian cells. **Results:** The beta chain similarity ranged from 40.10-81.60% among 39 species aligned with the obtained sequence. There are 146 amino acids with 12 cysteines (which form six disulfide bridges) and a single N-glycosylation site conserved. The constructed phylogenetic tree indicates may degree of kinship with the order Elopiformes and Clupeiformes. The structure of the heterodimer is mostly polar, with a predominance of beta-sheet in the secondary structure, classified as a Cystine-knot cytokines (CATH 2.10.90.10). Molecular dynamics studies of the interaction of TSH with receptor of its common ancestor, Asian arowana (*S. formosus*), both solved by AlphaFold, point to values comparable to the human model, with a difference in binding energy (ΔG) related values between 1.50-8.64%, indicating a pathway for *in vitro* and *in vivo* studies. The transient expression in mammalian cells obtained a yield of agTSH 25 mg/L, similar to previous work with agFSH 28 mg/L), quantified by chromatographic techniques in the conditioned medium. **Conclusions:** Beta subunit coding sequence and untranslated 3' region was identified and used in combination with alpha subunit to HEK293F protein expression. The methodology used could be applied in order to new proteins of the specie being described. Although requires optimization to enhance the expression, this work contributed to future studies to explore your biological activity, looking for the efficiency on the reproduction and improvement of reproduction capacity in Pirarucu captivity

Keywords: Thyroid Stimulating Hormone (TSH); AlphaFold; Pirarucu; HEK293

Lista de tabelas

	Página
Tabela 1 - Localização dos trabalhos com o termo <i>Arapaima gigas</i> em diferentes áreas de conhecimento	22
Tabela 2 - Ações dos hormônios tireoidianos em três famílias diferentes	31
Tabela 3 - Lista dos primers utilizados.....	59
Tabela 4 - Similaridade de aminoácidos dos hormônios glicoproteicos de <i>A. gigas</i>	77
Tabela 5 - Varredura teórica dos valores para <i>foldi</i> ng do agTSH	89
Tabela 6 - Valores obtidos da dinâmica molecular para agTSH	90
Tabela 7 - Informações recuperadas com base na sequência primária de agTSH.....	92
Tabela 8- Concentração de agTSH estimados em HPSEC após coleta em RP-HPLC.....	99

Lista de figuras

	Página
Figura 1 - <i>Arapaima gigas</i> apoiado em colchão	20
Figura 2 - Localização das principais glândulas e tecidos presentes no sistema endócrino de teleósteos	29
Figura 3 - Hipótese apresentada sobre a influência dos hormônios tireoidianos (e suas glândulas) no sistema reprodutivo.....	33
Figura 4 - Representação esquemática da atuação dos hormônios tireoidianos na reprodução de teleósteos	34
Figura 5 - Esquema de construção do DNA recombinante a partir de duas fontes de DNA diferentes	36
Figura 6 - Síntese e endereçamento proteico	48
Figura 7 - Os cinco diferentes tipos de glicosilação em proteínas	49
Figura 8 - Diferentes tipos de N glicanos	50
Figura 9 - Histórico da utilização das células HEK293	55
Figura 10 - Fluxograma das etapas do método de predição Gnomon	64
Figura 11- Mapa do vetor pcDNA3.4 com suas respectivas regiões que conferem características ao plasmídeo	68
Figura 12 - Sequência de nucleotídeos e aminoácidos correspondente a subunidade beta do TSH de <i>Arapaima gigas</i>	75
Figura 13 - Alinhamento múltiplo de sequências (Clustal Ω) envolvendo a sequência de ag β TSH (33) com mais 39 espécies.....	79
Figura 14 - Árvore filogenética construída com sequências de β TSH	80
Figura 15 - Modelo tridimensional <i>in silico</i> do agTSH a partir da metodologia <i>AlphaFold</i>	82
Figura 16 - Alinhamento do modelo de agTSH (magenta) com proteínas humanas cristalizadas.....	83
Figura 17 - Gráfico de ramachandran para agTSH	84
Figura 18 - Diagrama da topologia para as cadeias α e β	85

Figura 19 -	<i>Seat-belt</i> da subunidade β de <i>Arapaima gigas</i> com similaridade entre espécies.....	87
Figura 20 -	Representação do agTSH com superfície colorida por potencial eletrostático Poisson Boltzman.....	89
Figura 21 -	Perfis dos resultados obtidos por dinâmica molecular e <i>umbrella sampling</i>	91
Figura 22 -	Cromatograma da análise de 1 μ L do mRNA extraído da hipófise de <i>Arapaima gigas</i>	93
Figura 23 -	Eletroforese em gel de agarose para 3' RACE.....	93
Figura 24 -	Eletroferograma e alinhamento do sequenciamento do RACE correspondente a região 3' do ag- β TSH com a predita <i>in silico</i>	94
Figura 25 -	Cromatograma da análise de 1 μ L do plasmídeo pcDNA 3.4 + ag β TSH amplificado <i>Escherichia coli</i> e purificado	95
Figura 26 -	Eletroforese em gel de poliacrilamida 15% referente a produção piloto de agTSH em células HEK293	96
Figura 27 -	Comparação entre os três cromatogramas obtidos em RP-HPLC ...	97
Figura 28 -	Cromatogramas da análise em HPSEC e SDS-PAGE dos pools #P1 e #P2 coletados da primeira etapa de purificação em RP-HPLC.....	98

Sumário

INTRODUÇÃO E OBJETIVOS	14
2. A TIREOTROFINA COMO OBJETO DE ESTUDO NO CENÁRIO DA CONSERVAÇÃO E REPRODUÇÃO DE ARAPAIMA GIGAS.....	19
2.1. ARAPAIMA GIGAS, O GIGANTE PEIXE AMAZÔNICO	19
2.1.1. Origem e características fisiológicas e estruturais	19
2.1.2. O Arapaima gigas como alvo de estudo científico	21
2.1.3. O Arapaima gigas no cenário de conservação	23
2.1.4. O Arapaima gigas e sua importância cultural	25
2.1.5. O Arapaima gigas no Centro de Biotecnologia do Instituto de Pesquisas Energéticas e Nucleares.....	27
2.2. HORMÔNIO ESTIMULANTE DA TIREOIDE.....	28
2.2.1. Conceito de hormônio e disposição do sistema endócrino em teleósteos.....	28
2.2.2. Hormônio Estimulante da Tireoide (TSH).....	29
2.3. APLICAÇÕES DA BIOLOGIA MOLECULAR, BIOINFORMÁTICA E BIOQUÍMICA ESTRUTURAL COMPLEMENTA ESTUDOS SOBRE O ARAPAIMA GIGAS.....	36
2.3.1. Contexto histórico da clonagem molecular.....	36
2.3.2. Tecnologia do DNA recombinante para Arapaima gigas.....	37
2.3.3. Bioinformática e aplicações.....	41
2.3.4. Bioquímica estrutural de proteínas.....	42
2.4. EXPRESSÃO DE PROTEÍNAS RECOMBINANTES UTILIZANDO CÉLULAS DE MAMÍFEROS	46
2.4.1. Uso de organismos vivos para produção de proteínas e regulamentação das agências sanitárias.....	46
2.4.2. Processo de expressão proteica em eucariontes	47
2.4.3. Componentes do meio de cultura sem soro (Serum-Free) para células de mamífero.....	51
2.4.4. Células de ovário de Hamster Chinês (CHO).....	52
2.4.5. Células de rim de embrião humano 293 (HEK293)	54
3. MATERIAIS E MÉTODOS.....	57
3.1. HIPÓFISES DE ARAPAIMA GIGAS	57
3.2. EXTRAÇÃO DE RNA DA HIPÓFISE COLETADA DE ARAPAIMA GIGAS.....	58
3.3. AMPLIFICAÇÃO RÁPIDA DAS EXTREMIDADES DE CADEIA (5'/3' RACE)	58
3.3.1. REAÇÃO EM CADEIA DA POLIMERASE (PCR) E ELETROFORESE EM GEL DE AGAROSE.....	60
3.4. SEQUENCIAMENTO DE SANGER	61
3.5. ANÁLISE FILOGENÉTICA POR MÁXIMA VEROSSIMILHANÇA (ML)	62
3.6. OBTENÇÃO DA REGIÃO CODIFICANTE <i>IN SILICO</i>	62
3.6.1. Alinhamento de sequências cDNA β TSH de espécies contra o genoma do Arapaima gigas.....	63
3.6.2. Identificação da região codificante pela metodologia GNOMON.....	63
3.7. CONSTRUÇÃO DE MODELO TRIDIMENSIONAL DO AGTSH.....	66
3.8. ANÁLISE DOS MODELOS TRIDIMENSIONAIS E SEQUÊNCIA DE AMINOÁCIDOS DO AGTSH	66
3.9. CONSTRUÇÃO DOS COMPLEXOS.....	66
3.10. CÁLCULO DE ENERGIA LIVRE DE LIGAÇÃO (MM/GBSA) ENTRE AGTSH + SFTSHR.....	67
3.11. AVALIAÇÃO DA FORÇA DE LIGAÇÃO POR UMBRELLA SAMPLING.....	67
3.12. TRANSFORMAÇÃO DE BACTÉRIAS <i>ESCHERICHIA COLI</i> DH5 α COM PLASMÍDEO pCDNA 3.4 + AG-bTSH E SUA PURIFICAÇÃO	68
3.13. CULTURA DE CÉLULAS DE MAMÍFERO E TRANSFECCÃO TRANSIENTE COM LIPÍDEO CATIÔNICO.....	69
3.13.1. Transfecção transiente de células HEK293 - ensaio inicial	70
3.14. ANÁLISE QUALITATIVA DA PRODUÇÃO EM SDS-PAGE.....	70

3.15.	ANÁLISE DO MEIO CONDICIONADO EM RP-HPLC.....	71
3.16.	ANÁLISE QUANTITATIVA DOS <i>POOLS</i> EM HPSEC.....	71
4.	RESULTADOS	73
4.1.	ESTUDOS <i>IN SILICO</i> PARA AG-BTSH E AGTSH: IDENTIFICAÇÃO GENÔMICA.....	73
4.1.1.	<i>Confirmação do scaffold com outras sequências correspondente ao gene</i>	74
4.1.2.	IDENTIFICAÇÃO DA REGIÃO CODIFICANTE CONTRASTANDO COM <i>S. FORMOSUS</i> (SPLIGN & SPLIGNPRO)	74
4.1.3.	<i>Evidência para proposta in silico: porcentagem de identidade dos aminoácidos.....</i>	75
4.1.4.	<i>Evidência para proposta in silico: assinatura da subunidade beta.....</i>	76
4.1.5.	<i>Evidência para proposta in silico: conservação dos aminoácidos e domínios</i>	78
4.1.6.	ÁRVORE FILOGENÉTICA.....	80
4.1.7.	CONSTRUÇÃO DO MODELO TRIDIMENSIONAL PARA AGTSH E CÁLCULO DE RMSD COM PROTEÍNAS CRISTALIZADAS	81
4.1.8.	PROPRIEDADES FÍSICO-QUÍMICAS E ESPACIAIS PARA AGTSH	83
4.1.9.	<i>Pontos de Ramachandran para agTSH</i>	83
4.1.10.	<i>Topologia proteica para agTSH</i>	84
4.1.11.	<i>Valores teóricos dos parâmetros eletrostáticos de agTSH.....</i>	88
4.1.12.	<i>Dinâmica molecular e cálculos de energia livre de MM-GBSA e umbrella sampling para agTSH com diferentes receptores</i>	90
4.1.13.	TABELA DE INFORMAÇÕES RECUPERADAS COM EXPASY	91
4.2.	SÍNTESE DO HORMÔNIO RECOMBINANTE AGTSH: ATIVIDADES PRÁTICAS.....	92
4.2.1	<i>Quantificação da extração de RNA total a partir da hipófise de Arapaima gigas.</i>	92
4.2.2	<i>Amplificação das extremidades de cadeia (RACE).....</i>	93
4.2.3	<i>Sequenciamento de SANGER e comparação com proposta in silico</i>	94
4.2.4.	<i>Quantificação do plasmídeo pcDNA 3.4 + ag-βTSH amplificado em Escherichia coli</i>	94
4.2.5.	<i>Avaliação por SDS-PAGE da produção de agTSH em células HEK293 cultivadas em suspensão</i>	95
4.2.6.	<i>Quantificação por RP-HPLC do agTSH expresso em célula HEK293.....</i>	97
4.2.7.	<i>Análise dos pools em HPSEC do agTSH</i>	98
4.2.8.	<i>Rendimento estimado da produção</i>	99
5.	ANÁLISE E DISCUSSÃO DOS RESULTADOS	99
5.1.	SEQUÊNCIA PRIMÁRIA DO AG-BTSH.....	100
5.2.	ESTRUTURA QUATERNÁRIA DO AGTSH	101
5.3.	EXPRESSÃO, CARACTERIZAÇÃO E PURIFICAÇÃO DO AGTSH.....	102
	CONCLUSÕES	104
	REFERÊNCIAS BIBLIOGRÁFICAS.....	105
	APÊNDICE A - RANQUEAMENTO DOS 56 MODELOS DE SUBSTITUIÇÃO DE AMINOÁCIDOS PARA CONSTRUÇÃO DA ÁRVORE DE MÁXIMA VEROSSIMILHANÇA	113
	APÊNDICE B - ORDEM DOS COMANDOS EM PYTHON UTILIZADOS NO ALPHAFOLD	115
	APÊNDICE C - RELATÓRIO DOS VALORES DE PKA DOS AMINOÁCIDOS PARA <i>FOLDING</i>	128
	ANEXO 1 – DOCUMENTO REFERENTE A DOAÇÃO DE HIPÓFISES	132
	ANEXO 2 - CONFIRMAÇÃO DA SEQUÊNCIA DEPOSITADA NO GENBANK.....	133

Introdução e objetivos

Difícilmente, seja no cenário político, acadêmico, social ou comercial, hoje um trabalho não discute os impactos no clima e ambiente. Majoritariamente, ações sociopolíticas que modificam o meio ambiente são pautas que estão ganhando cada vez mais espaço, visto um acelerado crescimento populacional e tecnológico - estima-se que em 2050 o mundo tenha mais de 8,5 bilhões de pessoas (LI *et al.*, 2022). Tal aceleração impacta na manutenção do ecossistema, especialmente considerando ações que o afetam diretamente tais como a extração constante de matéria-prima com modificação/destruição local, a construção de barragens e áreas urbanas, o aumento da poluição, uso extensivo de agrotóxicos, e indiretamente como a variação no regime fluvial e migração de espécies nativas. Nesse sentido, potenciais problemas como o desaparecimento de espécies, o aumento da temperatura, entre outros fatores (CHEN *et al.*, 2020) se tornam objeto de atenção cujo prejuízo vem sendo alertado por especialistas e pesquisadores que atuam diretamente sob os impactos e escassez desses recursos naturais, acentuando ainda mais um desequilíbrio biológico, gerando uma discussão acerca do fornecimento de alimentos e energia.

Levando em consideração a preocupação apontada, hoje no Brasil, as políticas governamentais a respeito da bacia hidrográfica amazônica estão direcionadas mais para uma visão comercial do que preservacionista. A partir de um levantamento realizado em 2017, 140 barragens foram construídas e mais de 428 estão sob análise e algumas já em processo de construção em regiões que compõem a bacia (LATRUBESSE *et al.*, 2017). Essas barragens para geração de energia são apenas uma das alternativas, dentre possibilidades - menos impactantes - que o território brasileiro possa oferecer na geração de energia (eólica e solar, no caso). O desvio do curso natural dos rios para construção de barragens e de usinas hidrelétricas, além de provocar alagamentos, impulsiona a economia local que resulta na migração da população, com a desocupação de um território antes preservado, ao passo que modifica processos e dinâmicas naturais dali cujas consequências colocam em risco a fauna/flora local. Em vista disso e considerando que o Pirarucu é uma espécie nativa da bacia amazônica, não podíamos deixar de apontar esses detalhes que interferem diretamente na sua preservação.

O cenário descrito está interferindo no ecossistema local de forma irreversível e, portanto, diferentes espécies sofrem impactos dessas intervenções. O Pirarucu (*Arapaima Gigas*), peixe teleósteo nativo da bacia Amazônica, que podem atingir até três metros de

comprimento e pesar mais de 200 quilogramas é uma das espécies afetadas. Suas características morfológicas provocam um apelo comercial - prevalentemente alimentício - dentro da comunidade ribeirinha e da região. Além do sabor de sua carne (popularmente conhecido como "Bacalhau da Amazônia"), diversas partes podem ser aproveitadas quando é abatido. Dentre estas, podemos citar a sua língua óssea, que pode ser comercializada e utilizada como lixa ou adornos, e sua pele pode virar artigo de luxo, como revestimento de bolsas, tênis, blusas, etc. Além disso, há um grande interesse no ramo da piscicultura, uma vez que sua criação em cativeiro estimula a economia local. Contudo, alguns aspectos identificados no seu comportamento e na sua reprodução limitam esta prática (LIMA, 2018).

Nas últimas décadas houve um crescimento das colaborações entre instituições de pesquisa e empresas privadas, com avanços na determinação do sexo da espécie e melhor compreensão de sua fisiologia e respostas ao ambiente que afetam diretamente seu comportamento. Mas, quando tratamos da questão reprodutiva, um desses problemas acaba sendo a sazonalidade da reprodução (FERREIRA, MARCOVITCH, VAL, 2020). Em outras palavras, a ninhada ocorre apenas em determinadas condições, como em tempos chuvosos (caso ambos sexos estejam aptos na mesma época). Tentativas de mudanças ambientais (temperatura, umidade, espaço, alimentação, luminosidade) já foram realizadas em pisciculturas e nenhuma obteve um sucesso direto. Acreditamos que uma das possibilidades possa ser uma mudança a nível fisiológico dos hormônios que estão diretamente ligados com a reprodução (NUNES *et al.*, 2011).

Os trabalhos com o *Arapaima gigas* realizados em nosso grupo de pesquisa buscaram dar uma importância maior na resposta dos hormônios pituitários (originados da hipófise, uma glândula localizada em uma estrutura óssea - sela turca - na região cerebral) frente aos problemas reprodutivos. Os hormônios da adenohipófise (ou região anterior da hipófise) são: o hormônio de crescimento (GH), prolactina (PRL), hormônio adrenocorticotrófico (ACTH), hormônio estimulante de melanócito, hormônio estimulante da tireóide (TSH), hormônio luteinizante (LH) e hormônio folículo estimulante (FSH). Quando trabalhamos especificamente com TSH e seus respectivos hormônios tireoidianos, algumas de suas funções podem ser exploradas a fim de esclarecer melhor seu papel na reprodução.

A tireotrofina - ou TSH - é um hormônio heterodimérico de aproximadamente 27-30 kDa formado por duas subunidades: alfa e beta. A subunidade alfa é comum entre os hormônios glicoproteicos (GTH), dos quais fazem parte o FSH, LH e gonadotrofina

coriônica (CG). O TSH tem como função regular a liberação dos hormônios tireoidianos (HTs): a tiroxina (T4) e triiodotironina (T3), por meio de uma resposta de *feedback* negativo. Quando na circulação há maior quantidade de HTs, há uma resposta para glândula tireoidiana e hipofisária na diminuição da produção e secreção de TSH e TRH (Hormônio liberador de tireotrofina), respectivamente. (SZKUDLINSKI *et al.*, 2002).

Comumente, a ação dos HTs se deve a forma ativa destes (T3), derivada da ação de enzimas deiodinases sobre a forma inativa (T4). Essas enzimas participam na sua ativação ou inativação removendo ou adicionando um iodo à molécula. Dos efeitos biológicos mais comuns resultantes da ação da forma ativa, em diferentes tecidos (nervoso, muscular e conjuntivo), listamos a proliferação, diferenciação e migração celular, além de influenciar no comportamento e na expressão gênica. (ACCORRONI; CHIELLINI; ORIGLIA, 2017)

Quando especializamos e direcionamos o estudo para o sistema reprodutor - majoritariamente formado por tecido conjuntivo - e considerando a influência dos HTs, localizamos estudos que demonstram o aumento ou retardamento da síntese dos esteroides relacionados ao sexo masculino (como a testosterona) e também femininos (como estradiol) (SILVA, OCARINO, SERAKIDES, 2018) Esse controle pode resultar tanto no desenvolvimento e/ou diferenciação de características fenotípicas relacionadas ao sexo, quanto também no aumento da incidência de células somáticas, resultante da maturação final dos tecidos reprodutivos (MERONI *et al.*, 2019). Sucintamente, também serão apresentados durante o referencial teórico desta dissertação, estudos relacionados a modulação hormonal com base em agentes externos - focalizando os agrotóxicos, correlacionando sua influência no sistema tireoidiano.

Considerando os estudos com espécies de peixes teleósteos (TOVO-NETO *et al.*, 2018), comprova-se a importância em investigar o papel da tireoide e os hormônios relacionados na participação do sistema reprodutivo juntamente com outros hormônios hipofisários de *Arapaima gigas*, como o folículo estimulante e luteinizante (agFSH e agLH, respectivamente) que também sendo estudados pelo nosso grupo de pesquisa e expressos em células de mamíferos.

O desenvolvimento da parte prática e computacional – com exceção das dinâmicas moleculares, feita por colaboradores da dissertação mencionado a seguir- foi realizada nas instalações do Centro de Biotecnologia (CEBIO), dentro do Instituto de Pesquisas Energéticas e Nucleares (IPEN) na Universidade de São Paulo, que possui infraestrutura

completa no seu desenvolvimento, com certificação em qualidade de biossegurança (CQB 0067/98¹).

O grupo de pesquisa possui uma vasta experiência em expressão de proteínas recombinantes, com trabalhos sendo publicados periodicamente desde 1996, onde foi possível expressar a prolactina modificada com histidina (tag-hPRL) utilizando a tecnologia do DNA recombinante (MORGANTI *et al.*, 1996).

Nesse cenário e com avanço das técnicas ligadas a biotecnologia, o grupo foi se especializando na expressão de proteínas recombinantes em *Escherichia coli* e em células de mamífero (inicialmente com CHO e, recentemente, HEK293) e na sua caracterização por diferentes técnicas físico-químicas, como cromatografia líquida de alta eficiência (HPLC). Também, com estrutura para purificação desses produtos recombinantes, o laboratório tem experiência com colunas de alta pressão por diferentes tipos de cromatografias (afinidade, hidrofobicidade e exclusão molecular) sendo possível realizar todas as etapas de síntese, caracterização e purificação do produto biológico. Um dos hormônios mais estudados pelo grupo foi o hormônio estimulante da tireoide humano (hTSH), com 6 trabalhos publicados em periódicos internacionais: *Applied Microbiology and Biotechnology*, *International Journal of Molecular Sciences*, *Environmental and Molecular Mutagenesis*, *Radiation on Environmental Biophysics* e *Analytical Methods* (SANT'ANA *et al.*, 2018; RIBELA *et al.*, 2017; VALGODE *et al.*, 2017; SILVA *et al.*, 2016; ALMEIDA *et al.*, 2014; DAMIANI *et al.*, 2013).

Vale ressaltar que nossa pesquisa obteve colaborações de cunho prático em dois segmentos: no âmbito acadêmico, a colaboração de estudos realizados por pesquisadores da Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP), em São José do Rio Preto, no Instituto de Biociências, Letras e Ciências Exatas, que tem vasta experiência na cristalização de proteínas e ensaios *in vitro* e *in silico* (GONZALEZ *et al.*, 2022; GONZALEZ *et al.*, 2021; GONZALES *et al.*, 2021) contribuiu para nossa abordagem da dinâmica molecular do agTSH sustentando nossa hipótese acerca do objeto em análise (ver Materiais e métodos, tópico 3.6; 3.7 e 3.8); no âmbito de prosseguirmos com metodologias analíticas, as visitas técnicas das instalações da piscicultura Raça (Canabrava do Norte –

¹ Disponível em: <http://ctnbio.mctic.gov.br/documents/566529/665172/Ata+081+PLEN%C3%81RIA-SETEMBRO+2004/240048e5-3890-4b3b-af45-a5795989475c;jsessionid=8F5C66D7C694BA00F87AC8E6152998BB.columba?version=1.3>. Acesso em: 17 de mai. 2021

Mato Grosso, MT) e EMBRAPA Pesca e Aquicultura (Palmas – Tocantins, TO), realizadas em fevereiro de 2020 junto ao grupo de pesquisa do Centro de Biotecnologia (CEBIO) do IPEN, nos permitiu conhecer e ampliar a nossa percepção acerca da criação em cativeiro do *Arapaima gigas* e das dinâmicas presentes fossem elas comerciais, conservacionista, manejo e de cunho científico, realizadas pelos profissionais que ali atuam.

Esse trabalho tem por proposta contribuir em dois âmbitos: o da preservação ambiental e o da melhoria da capacidade reprodutiva em cativeiro. Durante o seu desenvolvimento essas questões sempre estiveram presentes, promovendo a revisão da literatura e diálogos com grupos de pesquisa, tal como condução dos trabalhos em laboratório.

A pesquisa tem como objetivo principal avaliando e considerando as colocações apresentadas, que se articulam a estudos realizados pelo grupo de pesquisa, identificar a cadeia de nucleotídeos correspondente a subunidade beta do hormônio estimulante da tireoide de *Arapaima gigas* (ag- β TSH) e construir a estrutura tridimensional do agTSH a partir da sequência de aminoácidos *in silico*, com as mais recentes metodologias de bioinformática. Como objetivos específicos visamos:

- Explorar as propriedades bioquímicas do agTSH com base na sua estrutura quaternária;
- Estudos de modelagem molecular e interação com receptor de TSH de *Scleropages formosus* comparando com modelo humano utilizando metodologias *in silico*;
- Transfectar, de forma transiente, células de mamíferos para expressão da proteína recombinante;
- Caracterizar e quantificar, através de métodos físico-químicos, a proteína expressa no meio condicionado.

Esta pesquisa é continuidade de estudos anteriores iniciados com projetos de iniciação científica financiados pelas agências de fomento do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq 01/01/2017-31/07/2017, 101573/2017-0) e Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, 01/01/2019-31/12/2019, 2018/10697-2). Nesse sentido, buscamos investigar, apontar e complementar lacunas existentes sobre o tema, ampliar e discutir, com maior profundidade os aspectos que

circundam a problemática apontada, do ponto de vista das ciências biológicas e, a partir desse estudo inédito para espécie, desdobram-se as possibilidades de expandir as abordagens dos estudos relacionados aos hormônios hipofisários de *Arapaima gigas*.

2. A tireotrofina como objeto de estudo no cenário da conservação e reprodução de *Arapaima gigas*.

Esse capítulo tem como direcionamento contextualizar o trabalho no âmbito acadêmico. Através das principais bases de dados, três grandes capítulos foram organizados com publicações importantes na compreensão dos assuntos que norteiam o macro e micro referentes a pesquisa.

Durante a leitura, serão abordados os temas referentes ao *Arapaima gigas* e sua importância ambiental, cultural e científica e o hormônio estimulante da tireoide, como potencial alvo de estudo quando se trata dos problemas apresentados com o peixe. Os capítulos que fazem menção as técnicas de biologia molecular, bioinformática e bioquímica estrutural, assim como o de expressão de proteínas recombinantes e células de mamíferos, buscaram contextualizar a parte prática do trabalho, envolvendo os conceitos mais utilizados e novos aplicados a dissertação.

Por fim, o leitor deve compreender o trabalho nos capítulos seguintes, envolvendo as metodologias e os resultados obtidos durante a realização desse projeto.

2.1. *Arapaima gigas*, o gigante peixe amazônico

2.1.1. Origem e características fisiológicas e estruturais

O *Arapaima gigas* (popularmente conhecido como Pirarucu, no Brasil; Bodecos, na região Amazônica e Paiche, no Peru e Equador) é considerado um dos maiores peixes de água doce do mundo. A espécie é nativa da bacia hidrográfica do rio Amazonas, com sua extensão compreendendo além do Brasil, outros países da América Latina: Peru, Equador, Bolívia e parte da Colômbia. Seu nome, Pirarucu, é de origem tupi, onde *Pira* significa “peixe” e *urucu* “vermelho”, cuja referência etimológica vem da coloração avermelhada do fruto Urucu (*Bixa orellana* L.). A coloração está presente em ambos os sexos da espécie, predominando na orla posterior das escamas e em algumas regiões do corpo (Figura 1). Na imagem, podemos visualizar sua cor avermelhada e de comprimento aproximado de 170 cm. (VENTURIERI, BERNARDINO, 1999). Podendo alcançar, nas exceções, cerca de 3 metros de comprimento e 200 kg, o Pirarucu torna-se um interessante alvo de pesca para consumo

e comercialização. Geralmente, o encontramos com cerca de 45-90 kg, o que já apresenta vantagem expressiva quando comparada com outras espécies pescadas na região amazônica, como Pirapitinga – 14 kg (*Piaractus brachypomus*), Mapará – 3 kg (*Hypophthalmus edentatus*) e Matrinchã – 5 kg (*Brycon amazonicus*).

Figura 1: *Arapaima gigas* apoiado em colchão.



Fonte: Fotografia cedida por Moisés Zorzeto, em Canabrava do Norte, MT, Brasil.

Pertencente à família Arapaimidae, o *Arapaima gigas* está inserido na ordem dos osteoglossiformes, das quais as espécies apresentam em comum, a língua óssea. Sua classe, Actinopterygii, remete a classificação de peixes que possuem nadadeiras com sustentação de lepidotríquias (escamas calcificadas que articulam as barbatanas) (BEMVENUTI, FISCHER, 2010). A maioria dos peixes catalogados fazem parte da classe mencionada anteriormente. Essas espécies apresentam também, uma única nadadeira dorsal, uma estrutura conhecida como acrodinia em seus dentes (uma capa de tecido mineralizado), com seu corpo coberto de escamas ganóides².

Em relação a sua fisiologia reprodutiva, a espécie se desenvolve em regiões de clima equatorial, com temperaturas de água entre 24 – 26 °C, principalmente em épocas de chuvas (dezembro-junho). Provavelmente essa informação está relacionada com a variação do nível

²Ganóides: Estrutura em deposição lamelar com revestida de ganoína, uma substância semelhante à encontrada no esmalte dos dentes.

da água em regiões dos rios, onde acontece a desova e fertilização (em regiões de fundo de lagos, onde a água é mais calma). A maturidade sexual ocorre entre o terceiro e quinto ano de vida e, apesar de já ter sido noticiado a mudança de coloração da espécie no período reprodutivo, isso não segue um padrão (FARIAS *et al.*,2015). Sua reprodução é bem conhecida e a produção científica colabora para melhor desenvolvimento da espécie em cativeiro. Porém, vale ressaltar que a maior dificuldade da reprodução da espécie é na criação dos alevinos (MONTEIRO *et al.*,2010), uma vez que a reprodução do Pirarucu acontece (ou não) em uma determinada época e, caso tenha sucesso, também evidencia uma alta mortalidade de alevinos, dificultando seu manejo.

Sua estrutura anatômica e fisiológica desperta interesse em vários segmentos econômicos, trazendo um reflexo considerável na pesquisa, visto a quantidade de trabalhos realizados no âmbito nacional e internacional, abordados na Tabela 1 localizada no tópico 2.1.2. Sua importância, além dos eixos mencionados, também reflete como simbologia da cultura nacional, na literatura, teatro, pintura e até mesmo na música.

2.1.2. O *Arapaima gigas* como alvo de estudo científico

Como mencionado no tópico anterior, a espécie percorre desde o Brasil até os países vizinhos que compõem a bacia hidrográfica amazônica. Dessa vantagem, temos a participação de diversos núcleos de pesquisas com a espécie em âmbito nacional e internacional. Apesar de ser endêmica, o conhecimento sobre a espécie transpassa a América do Sul, com publicações na América do Norte e Europa. Em maio de 2022, foi realizada uma busca nas principais bases de dados (*Web of Science e Scielo*) e dissertações que foram publicadas no Brasil (Catálogo de Teses e Dissertações – CAPES) com o termo "Arapaima gigas". As informações obtidas foram organizadas com o número de trabalhos aproximado.

Tabela 1: Localização (em ordem decrescente) dos trabalhos com o termo *Arapaima gigas* em diferentes áreas de conhecimento.

Área de conhecimento	Número de publicações
Dissertações (CAPES)	141
Pesca (WoS)	86
Teses (CAPES)	61
Zoologia (WoS)	67
Biologia marinha de água doce (WoS)	48
Ciências veterinária (WoS)	29
Biologia molecular e bioquímica (WoS)	27
Ecologia (WoS)	24
Ciências multidisciplinares (WoS)	20
Ciências biológicas (Scielo)	20
Ciências Agrárias (Scielo)	19
Ciências dos materiais (WoS)	19
Parasitologia (WoS)	19
Fisiologia (WoS)	19
Multidisciplinar (Scielo)	19
Ciências da saúde (Scielo)	13
Ciências exatas e da terra (Scielo)	5
Engenharias (Scielo)	1
Ciências humanas (Scielo)	1

Abreviações: **WoS**: Web of Science.

Fonte: Autor da dissertação.

Há um número expressivo de publicações no Brasil, assim como o número de dissertações e teses defendidas. Um resultado esperado visto que grande parte da bacia amazônica compreende o território nacional e, como simbologia forte de cultura local, a espécie é visada no âmbito acadêmico

De todos os trabalhos mencionados em periódicos científicos, o grupo de Hormônios Hipofisários do Centro de Biotecnologia (IPEN/CNEN, Universidade de São Paulo, São

Paulo, SP, Brasil) teve a participação com dois trabalhos publicados com o Pirarucu (FARIA *et al.*, 2013; SEVILHANO *et al.*, 2017) demonstrando a participação do grupo nas pesquisas realizadas com a espécie.

2.1.3. O *Arapaima gigas* no cenário de conservação

Após discutir sobre as características do *Arapaima gigas* assim como sua importância na área acadêmica, discutiremos a seguir os aspectos que caracterizam o *Arapaima gigas* como uma espécie claramente ameaçada de extinção. Apesar de ser classificada pela *International Union of Conservation Nature* (IUCN) como uma espécie Deficiente de Dados (DD), publicações recentes demonstram a escassez dos estoques naturais da espécie, resultado de uma intensa pesca predatória e ilegal que acontece na região. Em contrapartida, a *Convention on International Trade in Endangered Species of Wild Fauna and Flora* (CITES) aloca o Pirarucu no Apêndice II, que não necessariamente está em extinção, porém devido ao seu comércio exacerbado, podem vir a entrar em risco de extinção.

Em suma, será apresentado em poucos detalhes dos órgãos que articulam parâmetros relativos à densidade de espécies que, por sua vez, se não forem idênticos, corroboram no acúmulo de informações para adensar detalhes sobre sua informação na natureza, avaliando o seu possível risco de extinção. No caso do Pirarucu, há uma lacuna de dados que está sendo preenchida aos longos dos anos, fruto do trabalho excepcional que vem sendo realizado por conservacionistas e pesquisadores de diferentes áreas, assim como instituições que trabalham fortemente na conservação da espécie.

A fim de não prolongar a discussão e posicionar a espécie nos dois órgãos, o tópico **3.1.3.1.** abordará os critérios utilizados para os dois órgãos e, o tópico **3.1.3.2.** as publicações referentes ao declínio *in situ* do *Arapaima gigas*.

2.1.3.1. Organizações para conservação de espécies

A IUCN é uma organização fundada em 1948 na França, cujo principal objetivo é informar e catalogar espécies que podem ou não estar ameaçadas de extinção, seguindo critérios e avaliações definidas e previamente atualizadas em 2001. Os critérios são divididos em 4 classificações (A à D), avaliando principalmente o tamanho da população e sua distribuição geográfica. Cada critério pode determinar se a espécie está Criticamente em Perigo (CR), Em Perigo (EN) e vulnerável (VU). Caso não tenha dados suficientes para classificação de VU, a espécie pode ser Quase Ameaçada (NT) ou Menor Preocupação (LC);

se classifica como deficiente de dados (DD) aquela que não possui informações suficientes nos quatro critérios definidos pela IUCN. Além dessas classificações, podemos citar com menores detalhes: Não Avaliada (NE), da qual não foi analisada pelos responsáveis da organização; extinto na Natureza (EW), onde a espécie só existe em situações de preservação *ex situ*; Extinto (EX) que remete ao desaparecimento completo da espécie.

Avaliando outro órgão e seus critérios de posicionamento em categorias a fim de promover a conservação e evitar a extinção de táxons, a CITES, fundada em 1963 nos Estados Unidos após um acordo com membros da IUCN, trabalha com acordos entre governos de 183 países para divulgação e catalogação de diferentes espécies, assim como estabelece limites de comercialização em suas classificações. De maneira simplificada, são divididas em três apêndices (I, II e III) dos quais: I: remete as espécies que são ameaçadas de extinção e sua comercialização só pode acontecer em específicas circunstâncias; II: são espécies que não necessariamente estão em extinção, mas sua comercialização precisa ser controlada (de acordo com regimentos pré-estabelecidos) para não tornar inviável sua sobrevivência; III: todas aquelas que são protegidas em seu próprio país de origem que usaram o CITES como recurso de proteção e controle do comércio.

2.1.3.2. Cenário *in situ* do *Arapaima gigas* com base em publicações em periódicos

Como mencionado anteriormente, esse subtópico irá tratar as referências atualizadas sobre a predominância da espécie *in situ*. Foi realizada então uma pesquisa das principais publicações que trabalham o assunto de conservação da espécie e dessa forma, a síntese sobre essas informações, buscando reafirmar que a espécie *Arapaima gigas* está sobre explorada em seu habitat natural.

Sabe-se que a genética associada à conservação é uma potente ferramenta que permite, aos pesquisadores, identificar e avaliar de uma maneira mais precisa os dados referentes a determinado objeto de estudo. A presença de diversos marcadores moleculares que podem ser utilizados amplia o horizonte de estudo, diversificando as técnicas e ferramentas presentes nesse campo (SAMARAI & AL-KAZAZ, 2015). Nogueira e colaboradores (2020) avaliaram através da técnica de microssatélites do DNA mitocondrial (mtDNA) a diversidade genética de espécies de *Arapaima gigas* na reserva biológica do lago Piratuba (Amapá, Brasil), região que faz divisa em duas bacias hidrográficas – Amazonas e Araguaia-Tocantins. Como resultado, 3 populações (*clusters*) distintas foram identificadas

na região, influenciada principalmente pela dispersão causada em áreas de baixo curso do rio na bacia do rio Araguaia-Tocantins, assim como a introdução da espécie, em outras regiões que naturalmente não ocorre. Das três populações analisadas, duas estão com baixos níveis de diversidade genética, evidenciado, pelos autores, como causa a pesca exploratória e introdução da pecuária de Búfalo (com o aumento de corpos estranhos na água, que aumentam a mortalidade de peixes).

Por ser um peixe de alto valor econômico e cultural, o *Arapaima gigas* acaba se tornando o peixe de várzea mais estudado dentre os existentes. Evidencia-se de fato que a captura de Pirarucu imaturo em seu habitat natural afeta na reprodução de novas espécies, uma vez que altera completamente o ciclo biológico da espécie, induzindo migrações não necessárias (GURDAK *et al.*, 2019). Além disso, reitera-se o detalhe da degradação de habitat, como o corte de madeira local e instalação do ser humano em novos habitats, na contribuição para uma deficiência de espécies *Arapaima gigas*. Endurecer as medidas restritivas, como propõe os autores (como limite de pesca da espécie pelo tamanho, uma vez que peixes maiores são muitas vezes reprodutores), são fatores cruciais para sua conservação.

Por fim, estudos conduzidos em duas reservas protegidas do rio Juruá (cuja nascente fica no Peru e tem seu curso entre os estados do Acre e Amazonas) monitoram 13 espécimes de Pirarucu em diferentes épocas, avaliando sua capacidade migratória e adaptativa a ambientes que sofreram exploração e acabaram sendo deterioradas (CAMPOS-SILVA, HAWES, PERES, 2019). O trabalho enfatiza a importância da comunidade local sobre a importância de ter acordos de pesca e trabalhar concomitante com a pesquisa científica, uma vez que determinando e observando a capacidade migratória do peixe, pode-se traçar objetivos de recolonização em áreas em que a presença é extremamente reduzida, aumentando a diversidade genética e contribuindo na sua preservação.

2.1.4. O *Arapaima gigas* e sua importância cultural

Além do alto valor econômico já mencionado anteriormente, o Pirarucu também tem seu reconhecimento no campo das artes. Esse tópico pretende apresentar brevemente alguns relatos da imagem da espécie em composições, pinturas, literatura e no teatro. O compilado de informações que será descrito tem como finalidade a complementação de informações que podemos obter acerca do *Arapaima gigas*, entendendo que os fins de estratégias de

conservação não são discutidos apenas na esfera monetária e ambiental, mas também em outros campos de atuação.

Em 1886, José Veríssimo Dias Matos (1857-1916), escritor nascido no Pará e idealizador da Academia Brasileira de Letras, publica um livro de ficção “Cenas da Vida Amazônica” contendo uma variedade de contos e crônicas que trazem um traçado de características etnográficas da Amazônia paraense como um ensaio social, apresentando dois discursos da época: a influência das doutrinas científicas e a visão heterogênea da mestiçagem na composição da cultura e identidade brasileira (DA SILVA, 2016). Nesse livro, o *Arapaima gigas* aparece em vários pontos: preparação de pratos típicos [p.42,43], comportamental [p.44], meios de comércio [p.174], reiterando a presença do peixe no cotidiano de uma população local (VERÍSSIMO, 1886). Outra abordagem feita na literatura foi realizada por Márcio Souza, jornalista, que publicou em 1979 o livro “Tem Piranha no Pirarucu”, como uma simbologia da espécie na cultura indígena.

Samuel Rawet (1929-1984) foi um contista, romancista, dramaturgo e ensaísta nativo da Polônia, que teve sua imigração para o Brasil em 1936 devido às perseguições nazistas. No seu estilo de teatro (comumente chamado de transgressão), aborda em uma das suas peças “A farsa da pesca do Pirarucu”: é a aparição no teatro brasileiro, no formato do fastiche crítico, como objeto da cultura brasileira, tal como outros elementos utilizados, seja o tupi, folclore, MPB etc. (TONUS, 2004).

Há poucos registros na música brasileira com a espécie. Rita Lee, cantora, compositora, atriz, escritora e mais conhecida como “Rainha do Rock Brasileiro” compõe uma música intitulada “Pirarucu”, fazendo parte das 12 músicas que compõem o álbum Bombom (1983), lançado juntamente com Roberto de Carvalho. José Clementino Bispo dos Santos (mais conhecido como Jamelão) (1913-2008), intérprete dos sambas-enredo da escola de samba Mangueira, também lança em formato de samba a composição “Pirarucu”. Mais interessante frisar que ambas as canções apresentam, na sua letra, pronúncias e palavras regionais, como “dotô, nós é” e outros elementos que compõem a fauna Brasileira, como “uirapuru”, espécie de ave nativa da floresta amazônica.

A pintura (predominantemente nos estados que cercam a aparição da espécie – bacia hidrográfica do rio Amazonas) traz o pirarucu também como símbolo da comunicação direta com consumidor/cliente em restaurantes, peixarias (BERND, 2011). Ari Paiva, artista

nacional, utiliza também a figura da espécie em seus trabalhos em muros e telas (AQUINO, 2012).

2.1.5. O *Arapaima gigas* no Centro de Biotecnologia do Instituto de Pesquisas Energéticas e Nucleares

A expertise na expressão, caracterização e purificação de hormônios hipofisários do Centro de Biotecnologia do IPEN/CNEN amplia a partir de 2013 os estudos focados no Pirarucu. Introduzir o estudo da espécie no grupo de pesquisa tiveram alguns pontos principais:

- Não havia publicações referentes a área de estudo com o peixe, apresentando uma lacuna de conhecimento importante dado que, os hormônios hipofisários (do pirarucu) são proteínas que estão diretamente ligadas às etapas de crescimento e reprodução da espécie;
- A oportunidade de colaborar com a equipe de pesquisa da empresa pública EMBRAPA – Amazônia Oriental (Belém, PA) nos estudos de identificação de proteínas. Até então, nunca havia sido utilizado técnicas de biologia molecular no laboratório, incentivando a capacitação de profissionais e alunos nessa área do conhecimento;
- O *Arapaima gigas* como espécie ameaçada constantemente de entrar em risco de extinção.

Os nossos primeiros trabalhos publicados nessa área foram através da aplicação das técnicas já utilizadas em proteínas humanas para as do *Arapaima gigas*, partindo da coleta *in loco* de hipófises e tecidos. Somente com a extração do tecido foi possível identificar, clonar, expressar e caracterizar o hormônio folículo-estimulante, luteinizante e, nesse trabalho, o hormônio estimulante da tireoide e, concluindo paralelamente, o hormônio de crescimento. Acompanhando as metodologias mais recentes em biologia molecular, bioinformática e bioquímica estrutural de proteínas, houve uma maior compilação dos dados que contribuem para estudos já em andamento e direcionam o estudo para aprimorar a conservação da espécie.

O próximo capítulo de referencial teórico que contribui no desenvolvimento desta dissertação será descrever com detalhes as informações referentes ao hormônio estimulante da tireoide, objeto de estudo nesse projeto. Iremos compreender principalmente sua

importância no aspecto reprodutivo e como pode ser utilizado como estratégia para melhoria da reprodução em cativeiro.

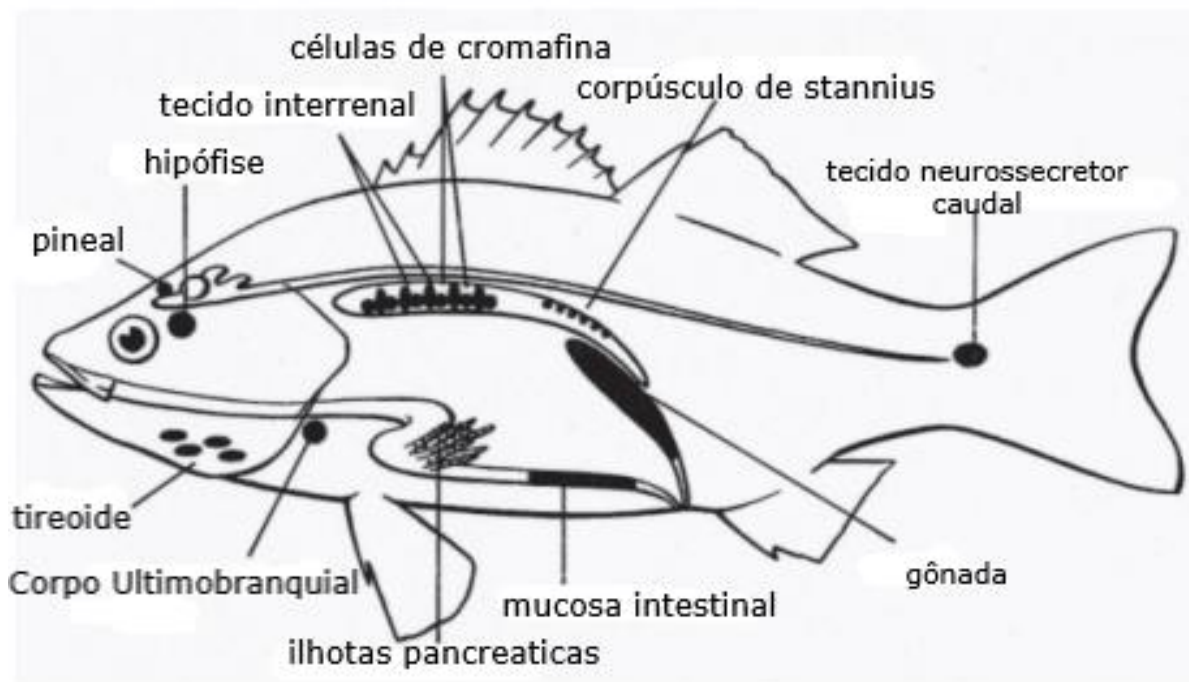
2.2. Hormônio Estimulante da Tireoide

2.2.1. Conceito de hormônio e disposição do sistema endócrino em teleósteos

Antes de descrever as funções biológicas, síntese e características físico-químicas do hormônio em questão, precisamos definir o conceito de hormônio: são pequenas moléculas lipídicas ou proteicas, sintetizadas em um tecido específico que podem ou não ser liberadas na circulação sanguínea e, atuam em outro órgão/tecido específico por meio de receptores, determinando atividades metabólicas que refletem no desenvolvimento e comportamento (NELSON & COX, 2018). Essa síntese pode ser feita de forma endócrina (circulação sanguínea); sinalização neural, por impulsos nervosos que transmitem uma mensagem para o órgão-alvo; autócrina, onde o alvo é o mesmo tipo celular e parácrina, quando o tipo celular alvo está próximo e difere da célula que sintetiza o hormônio.

A disposição do sistema endócrino em mamíferos difere entre espécies. Como apresentado na figura abaixo (Figura 2), nota-se que, o sistema endócrino em teleósteos, possui semelhanças quando comparado a espécie *Homo sapiens*. Alguns de seus órgãos/glândulas são observados em ambas as espécies, como a glândula pineal, hipófise e tireóide, por exemplo. As estruturas presentes que diferem da espécie humana, são elas: o corpúsculo de stannius: glândulas encontradas apenas em teleósteos que são responsáveis pela secreção do hormônio hipocalcemiante SCT-1, regulador da captação de cálcio em peixes (GREENWOOD *et al.*, 2009); a estrutura externa denominada corpo ultimobranquial – responsável pelo armazenamento da célula parafolicular na estrutura (responsável pela síntese de calcitonina, hormônio regulador do Ca^{2+}). A localização da glândula tireóide em peixes ósseos está em torno da artéria e em cartilagosos abaixo do músculo da língua.

Figura 2 – Localização das principais glândulas e tecidos presentes no sistema endócrino de teleósteos.



Fonte: Adaptado e traduzido de Pait e Nelson, 2002.

2.2.2. Hormônio Estimulante da Tireóide (TSH)

O hormônio estimulante da tireoide – ou tireotrofina - (TSH) é um hormônio pertencente à família dos “hormônios glicoproteicos”³. Sintetizado pela porção anterior da hipófise, a proteína depende da estimulação do hormônio liberador da tireotrofina (TRH), cujo responsável pela síntese e liberação é o hipotálamo. Os dois hormônios (TSH e TRH) tem a secreção e produção controladas através do mecanismo de *feedback* (auto regulação) negativo: à medida que um hormônio atinge uma alta concentração na circulação sanguínea, o outro hormônio em questão regula a sua liberação, e vice-versa (SZKUDLINSKI *et al.*, 2002). O TSH por sua vez, estimula as células da tireóide a secretar em maior quantidade a tiroxina (T4), que por sua vez irá sofrer o processo de desiodação – por enzimas conhecidas como desiodases - e, manter em maior quantidade na circulação sanguínea, a triiodotironina (T3) (NUNES, 2003). Já os hormônios tireoidianos (HTs) são conhecidos pela sua captação

³ Os hormônios glicoproteicos possuem oligossacarídeos ligados em sua estrutura proteica, em sítios específicos de glicosilação. Todos são heterodímeros (formados por duas subunidades proteicas) com a mesma sequência de aminoácidos na subunidade alfa e, na sua subunidade beta, difere, caracterizando a especificidade de cada hormônio. São pertencentes ao grupo: o hormônio estimulante da tireoide (TSH), gonadotrofina coriônica (CG), hormônio folículo estimulante (FSH) e o hormônio luteinizante (LH) (CAHOREAU; KLETT; COMBARNOUS, 2015)

de iodo, adquirido através da ingestão de alimentos. Contudo, participam de uma maneira bastante explorada no metabolismo intracelular, regulação da expressão gênica e até neuroprotetoras. Notamos que os hormônios tireoidianos participam diretamente no controle de mecanismos que estão ligados à reprodução. Por mais que outros hormônios hipofisários sejam estritamente ligados à reprodução como, por exemplo, o hormônio luteinizante (LH) e folículo estimulante (FSH), alguns mecanismos desempenhados pelo T3 e T4 mencionam sua influência direta no desenvolvimento, maturação e manutenção de órgãos como fígado, cérebro, gônadas e, suas respectivas sinalizações intracelulares, controlando, assim, tais mecanismos reprodutivos (CYR, EALES, 1996; HABIBI, NELSON, ALLAN, 2012; DEAL, VOLKOFF, 2020)

Em humanos, a proteína é formada por duas subunidades proteica, alfa e beta, ligadas não-covalentemente, com peso molecular variando entre 28-30 kDa. Possui três sítios de N-glicosilação em sua estrutura, duas na subunidade alfa e uma na subunidade beta, com uma alta porcentagem de conservação na transição da evolução entre peixes e mamíferos (SZKUDLINSKI *et al.*, 1996). Estudos filogenéticos relacionado ao receptor de proteína ricos em leucina contendo a proteína-G (LGR), estritamente relacionada à ligação dos hormônios glicoproteicos, comprovam a similaridade entre espécimes (PARK *et al.*, 2005), além de outros trabalhos que também buscam comparar a genética envolvida ao longo dos anos (POWER *et al.*, 2001; FORTIN *et al.*, 2009).

Em peixes, as funções metabólicas envolvidas na participação do TSH, T4, T3 e deiodinases, apresentam vias similares aos humanos. Defende-se na academia uma conservação evolutiva, devida à ação dos hormônios tireoidianos tanto em relação a metamorfose (fase larval para juvenil) e como no desenvolvimento de estruturas esqueléticas e cardíacas.

Abaixo na Tabela 2 estão apresentadas as principais funções dos hormônios tireoidianos (HTs) descritos na literatura, assim como o possível mecanismo de ação que atua no metabolismo de mamíferos e peixes. As manifestações fisiológicas/clínicas foram observadas com diferentes técnicas de detecção de genes, mencionadas nos trabalhos citados na coluna “Referência”. A síntese dessa tabela teve como base os trabalhos mencionados, buscando compreender, determinar e elucidar, a importância do hormônio TSH.

Tabela 2 – Ações dos hormônios tireoidianos em três famílias diferentes.

Família	Efeito	Manifestação fisiológica/clínica	Referência
<i>Hominidae</i>	Regulação da síntese de GH	Influência direta na transcrição do gene do GH*	NUNES, 2003
	Regulação da contração cardíaca e esquelética	Participação na transcrição dos genes α MHC*, MHC II e SERCA I*	
	Regulação de genes envolvidos no crescimento	Estimula a síntese de fatores de crescimento como NGF* e IGF-1*.	
	Ações neuroprotetoras	Aumento da captação de glutamato, resultando no aumento da atividade colinérgica	ACCORRONI; CHIELLINI; ORIGLIA, 2017
	Manifestação da doença de Alzheimer (AD)	Estudos mostraram que pacientes com AD possuem níveis reduzidos de T3 no córtex cerebral	
	Metabolismo da glicose cerebral	Possível aumento do transportador de glicose GluT4*	JAHAGIRDAR; MCNAY, 2012
	Hipertireoidismo	Aumento das chances de desenvolver infertilidade primária e secundária, menstruação irregular, atresia folicular e cistos ovarianos	SILVA; OCARINO; SERAKIDES, 2018
	Hipotireoidismo	Ocorrência de puberdade tardia, infertilidade, aborto espontâneo e anovulação	
<i>Teleostei</i>	Estimulação da glândula da tireoide	Desenvolvimento testicular e maturação sexual	TOVO-NETO <i>et al.</i> , 2018.
	Tratamento com tiroxina (T4) exógena	Diferenciação sexual reduzindo a expressão da <i>cyp19a1</i> * e <i>erb</i> *, além de aumentar a taxa de <i>amh</i> *	
<i>Cyprinidae</i>	Manifestações anormais na síntese dos HTs	Retardamento no desenvolvimento neural	WALTER <i>et al.</i> , 2019

***GH**: Hormônio de crescimento; **MHC**: Complexo principal de histocompatibilidade; **SERCA**: Ca^{2+} -ATPases; **NGF**: Fator de crescimento nervoso; **IGF**: Fator de crescimento insulínico; **GluT4**: Transportador de glicose-4; **cyp19a1**: Aromatase; **Erb**: Receptor de estrogênio beta; **Amh**: Hormônio antimülleriano.

Fonte: Autor da dissertação.

2.2.2.1. Estrutura tridimensional do hormônio estimulante da tireóide

Promover estudos de estruturas de proteínas é um campo fundamental na identificação de sítios de ligação da estrutura, sua disposição espacial, cargas expostas, ligações entre as estruturas (em caso de proteínas formadas por duas ou mais cadeias proteicas), etc. Com grande interesse farmacológico e fisiológico, o campo da bioinformática (discutido no tópico de Biologia Molecular, da dissertação) ganha um grande destaque com suas ferramentas e possibilidades analíticas. O uso dos recursos implantados vem em alternativa a metodologia tradicional da cristalografia, das quais inúmeras dificuldades são apresentadas quando tentamos criar um modelo (principalmente em encontrar uma condição adequada para formação do cristal e ordenação da proteína). Além disso, expressar grandes quantidades do hormônio recombinante em células de mamífero (discutido com mais detalhes no capítulo 4: Resultados) acaba sendo um processo muito longo e de alto custo.

Não há publicação de uma estrutura cristalográfica com a estrutura do TSH. O modelo mais próximo existente é a estrutura do receptor humano (RCSB PDB nº 3G04) com o anticorpo contra TSH humano (SANDERS *et al.*, 2007). Dessa forma, os modelos que são previstos para estrutura são realizados por homologia com outros hormônios já depositados. Uma importante vantagem é que hoje dispomos de ferramentas de inteligência artificial (AI) para criação de modelos tão precisos quanto aqueles baseados na cristalografia. A maior desvantagem é que o comportamento fisiológico da proteína ainda é desconhecido, considerando modificações pós traducionais e complexos com outras moléculas (ex. cofatores, para enzimas) de modo que esses modelos são criados com um padrão de comportamento fisiológico já estabelecido.

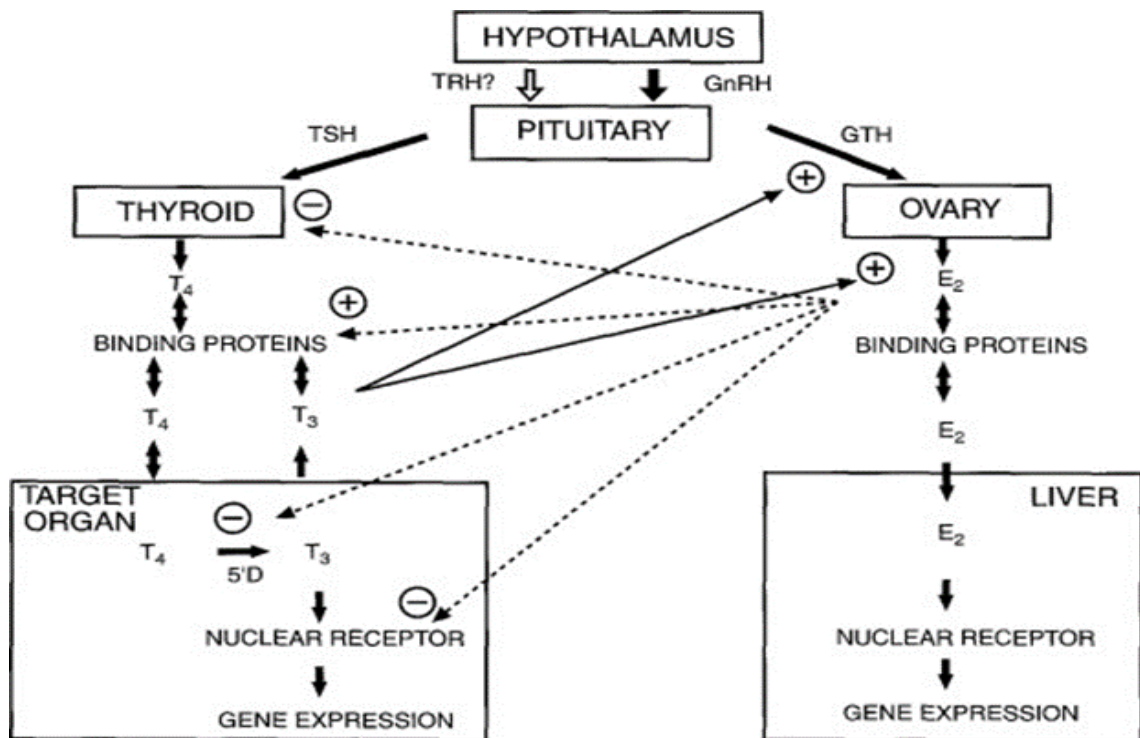
Entretanto, como apresentado nos resultados desta dissertação, os modelos tridimensionais e complexados com receptor do agTSH foram projetados com base nas metodologias mais recentes utilizadas para essa finalidade.

2.2.2.2. Participação dos hormônios tireoidianos e estimulante da tireoide em peixes

Através de ensaios de cultura de células de ovário (teca ou granulosa) de *Salvelinus fontinalis*, o T3 influenciou (em cerca de 30%) na conversão de testosterona em estradiol (E₂), quando incubado com hormônio gonadotrófico (GTH) (CYR & EALES, 1996). Esse resultado mostra que há um importante rota na vitelogênese da qual os hormônios tireoidianos atuam. Os autores também demonstram, em diferentes espécies de peixes, que

o T4 está diretamente relacionado com a captação de vitelogenina (VTG), importante proteína que atua como biomarcador, quando há expressão e ligação do E₂ à célula alvo (Figura 3). Nota-se que na figura, há o controle tanto a estimulação (+) ou retardamento (-) na liberação e síntese dos hormônios. Trabalho mais recente (MA *et al.*, 2020) também descreve a importância do T3 e T4 e sua correlação com o hormônio de crescimento (GH) e luteinizante (LH). Os autores, partindo de um tratamento direto com T3 em *Carassius auratus*, tiveram controle na expressão de GH e LH – redução dos níveis séricos de mRNA de LH; inibição dos níveis basais de GH na maturação sexual.

Figura 3 - Hipótese* apresentada sobre a influência dos hormônios tireoidianos (e suas glândulas) no sistema reprodutivo.



*Em inglês, temos a tradução: *Hypothalamus*: Hipotálamo; *Pituitary*: Hipófise; *Ovary*: Ovário; *Binding Proteins*: Proteínas de ligação; *Liver*: Fígado; *Nuclear Receptor*: Receptor nuclear; *Gene expression*: Expressão gênica; *Thyroid*: Tireóide; *Target Organ*: Órgão alvo. As abreviações são: TRH: Hormônio liberador da tireóide; GnRH: Hormônio liberador da gonadotrofina; GTH: Hormônio gonadotrófico; TSH: Hormônio estimulante da tireoide; E₂: Estradiol; T4 e T3: Hormônios tireoidianos; 5'D: Deiodinase

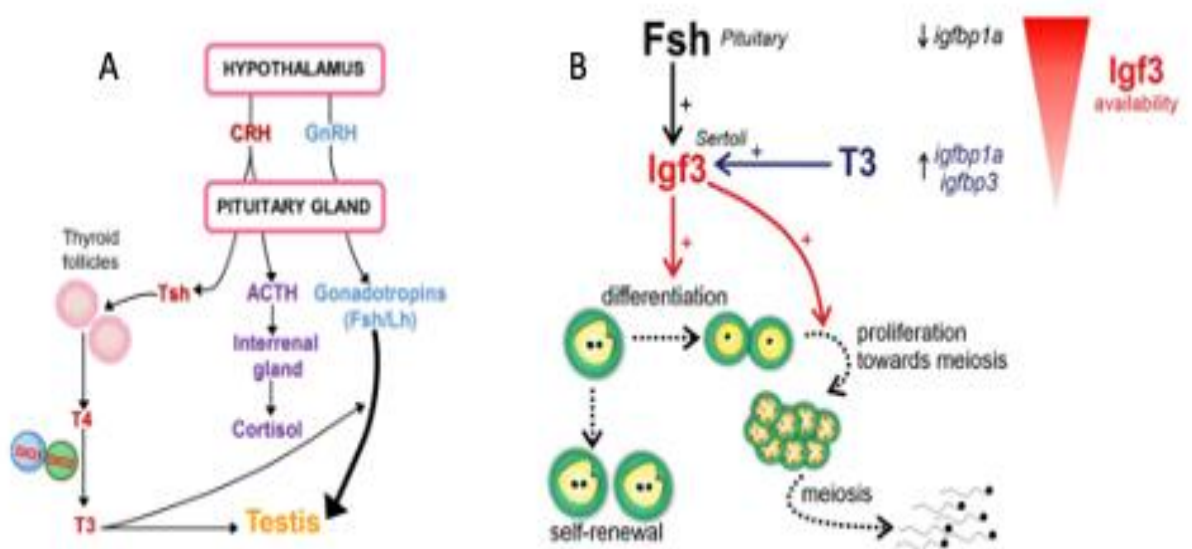
Fonte: Cyr & Eales, 1996

Em trabalho recente com teleósteos por Tovo-Neto e colaboradores (2018), pela Universidade Estadual Paulista (UNESP), traça e esquematiza como há uma influência em fatores de crescimento e que promovem uma estimulação positiva na proliferação de células germinativas no processo da espermatogênese. Nesse trabalho, também foram encontrados

receptores α e β para os hormônios tireoidianos (TR α e TR β) em células de sertoli e leydig (Figura 4, B).

Em estudo mais recente (HORIE *et al.*,2022) com peixe-arroz (*Oryzias latipes*) e peixe-zebra (*Danio rerio*) correlaciona os efeitos da expressão dos genes relacionados a síntese do TSH com a toxicidade de composto químicos presente na maior parte da composição de plásticos, como o ácido perfluorobutanóico (C₈HF₁₅O₂). De acordo com os autores, houve uma redução dos níveis de mRNA para receptor de tireotrofina (TR) α e TR β quando a espécie é exposta a estes compostos. Como resultado, a espécie não teve desenvolvimento adequado dos órgãos reprodutivos e partes locomotivas.

Figura 4: Representação esquemática da atuação dos hormônios tireoidianos na reprodução de teleósteos.



Em inglês, temos a tradução: **Hypothalamus:** Hipotálamo; **Pituitary gland:** Hipofise; **Thyroid follicles:** Tireóide; **Gonadotropins:** Gonadotrofinas; **Testis:** Testículo; **Interrenal gland:** Glândula intrarrenal; **Differentiation:** Diferenciação; **Self-renewal:** Auto Regeneração; **Proliferation towards meiosis:** Proliferação através da meiose; **Meiosis:** Meiose; **Availability:** Disponibilidade. As abreviações são: **TSH:** Hormônio estimulante da tireóide; **GnRH:** Hormônio liberador da gonadotrofina; **GTH:** Hormônio gonadotrófico; **ACTH:** Hormônio adrenocorticotrófico; **Igf3:** Fator de crescimento insulínico-3; **T4 e T3:** Hormônios tireoidianos; **DIO1/DIO2:** Deiodinase; **CRH:** Hormônio liberador da corticotrofina; **igfbp1a:** proteína de ligação ao igf1; **igfbp3:** proteína 3 de ligação ao igf1.

Fonte: TOVO-NETO *et al.*,2018.

Conforme observado na figura 4, **A** representa o processo de síntese hormonal, partindo do hipotálamo até órgãos alvo e **B** a esquematização da gametogênese masculina em teleósteos e atuação do T3 em fatores de crescimento. O FSH está diretamente relacionado ao aumento dos níveis de IgF3, assim como o T3 estimula as células de sertoli a produção de espermatócitos primários e subseqüente proliferação.

Expandindo a discussão a respeito da participação do TSH no metabolismo dos peixes, estudos demonstraram que o tratamento com esse hormônio aumentou os níveis plasmáticos do GH em *Oreochromis niloticus*, *Oncorhynchus mykiss*, *Oncorhynchus kisutch* (DEAL; VOLKOFF, 2020). Segundo os autores, poderiam ser empregadas metodologias para aprimorar e melhorar a taxa de sobrevivência larval e, poderiam até ser empregadas com finalidade reprodutiva. No mesmo trabalho, são apresentadas evidências de que a administração dos HTs pode influenciar na maturação do oócito de *Seriola spp.*

2.2.2.3. Aplicações do TSH e os hormônios tireoidianos em *Arapaima gigas*

Com a síntese teórica apresentada nesse tópico sobre o TSH e os hormônios tireoidianos, pudemos prosseguir nos importantes estudos realizados neste trabalho, assim como os objetivos e metodologias propostos.

Entendendo a importância dos HTs – e partindo de que sua síntese é exclusivamente dependente do TSH – podemos elucidar a importância do estudo hormonal em *Arapaima gigas*. O grupo de pesquisa do centro de biotecnologia do IPEN apresentou nos últimos anos dois trabalhos com hormônios hipofisários da espécie (SEVILHANO *et al.*, 2017; FARIA *et al.*, 2012) o que incentivou a investigar e dar continuidade nos estudos com a proteína em questão.

As várias hipóteses da atuação dos HTs na reprodução justificam parte do projeto, uma vez que a dificuldade de reprodução do *Arapaima gigas* foi previamente discutida no tópico anterior (cap 3, tópico 3.1.1) e há uma relação direta do TSH (cap 3, tópico 3.2.2) no manejo e conservação de espécies de peixes. Ampliar esses estudos para o *Arapaima gigas* visa corroborar com as informações apresentadas.

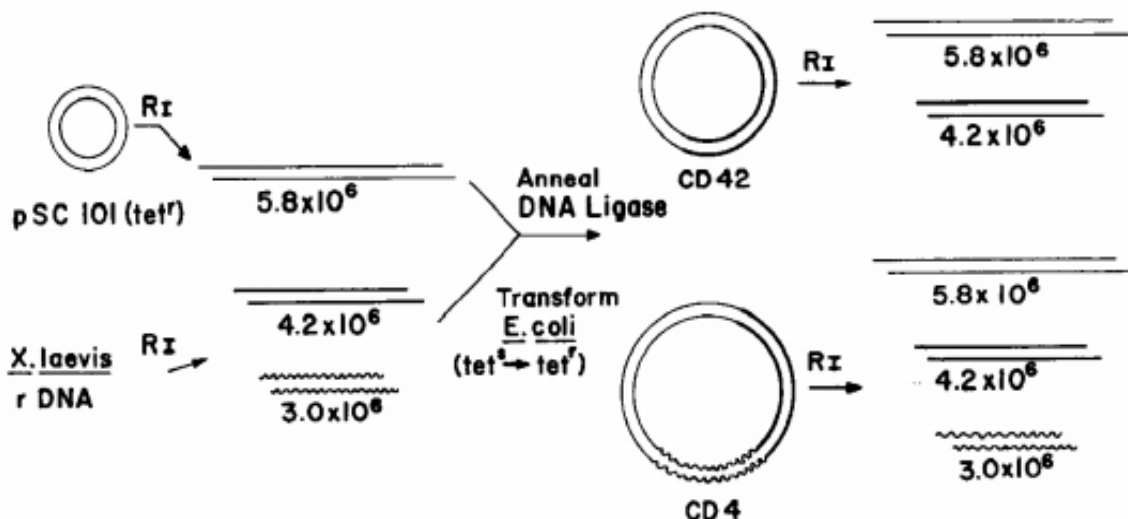
Expressar, caracterizar e purificar o agTSH desdobram nas possibilidades de investigar os comportamentos fisiológicos, propondo entender os mecanismos metabólicos que influenciam diretamente e indiretamente na sua reprodução, crescimento e desenvolvimento, assim como pode ser um potencial alvo de estudo para conservação da espécie, especialmente quando tratamos do cenário toxicológico (metais pesados, agrotóxicos e outros contaminantes fluviais)

2.3. Aplicações da Biologia Molecular, bioinformática e bioquímica estrutural complementa estudos sobre o *Arapaima gigas*

2.3.1. Contexto histórico da clonagem molecular

Apesar de já disponíveis em diversos laboratórios, somente na década de 70, Nathans e Smith (1975), na faculdade de medicina de Johns Hopkins (Baltimore, EUA), conseguiram pôr em prática as ferramentas da clonagem molecular, assim como também aplicar o uso das enzimas de restrição. Nos experimentos que realizaram, perceberam que no interior das bactérias *Escherichia coli*, havia uma classe de enzimas (apeladas de Classe I com suas características) que clivam o DNA em regiões específicas, em sequência de nucleotídeos padrão. Isolando os fragmentos que foram clivados, a partir de uma proteína DNA *ligase*, esses fragmentos foram unidos, formando novas combinações do DNA. Dessa forma, usando as enzimas de restrição adequada no DNA de *Xenopus laevis* e no vetor plasmidial já sintetizado (plasmídeo pSC 101) (Figura 5), conseguiram introduzir os fragmentos em um novo vetor (CD42 e CD4), consolidando assim, a tecnologia do DNA recombinante.

Figura 5 - Esquema de construção do DNA recombinante a partir de duas fontes de DNA diferentes⁴.



Em inglês, temos: **Transform *E. coli***: Transformação de *Escherichia coli*; **Anelamento da DNA ligase**: Anelamento da DNA ligase. As abreviaturas são: **pSC**: Plasmídeo de *Escherichia coli*; **rDNA *X. laevis***: DNA ribossomal de *X. laevis*; **Tetr**: resistência a tetraciclina; **Tet^s**: sensível a tetraciclina.

Fonte: Nathans, & Smith. (1975)

⁴ A clivagem com a enzima de restrição EcoRI (RI) produziu fragmentos que, unidos pela DNA ligase e transformados em *Escherichia coli*, geraram novos vetores plasmidiais.

O avanço das técnicas de clonagem molecular (conhecido também como técnica do DNA recombinante ou manipulação gênica) promoveram um leque de oportunidades para estudos diretamente ligados ao gene de interesse, sendo possível obter inúmeras cópias idênticas em um hospedeiro de alta taxa de replicação. O hospedeiro, ainda mais utilizado dentro dos laboratórios para produção de proteínas recombinantes, é a *Escherichia coli*, bactéria extremamente utilizada pela sua rápida taxa de duplicação (± 20 min) e cultivo relativamente barato.

Demos um salto imenso na ciência. Estamos vendo um crescimento exponencial do uso da metodologia em diversas aplicações no nosso dia a dia, seja em estudos relacionados ao desenvolvimento de vacinas, biofármacos, diagnóstico e até mesmo, para dieta nutricional. A manipulação genética se tornou um grande aliado na medicina, comprovando assim, a importância do seu uso (SHARMA *et al.*, 2014; TIRABASSI, 2014).

Nesse trabalho, daremos continuidade aos estudos realizados com *Arapaima gigas*, aliadas às técnicas da biologia molecular e, especificamente, a manipulação gênica. Abordaremos também as principais estratégias traçadas no laboratório do Centro de Biotecnologia do IPEN nos estudos relacionados aos hormônios hipofisários da espécie.

2.3.2. Tecnologia do DNA recombinante para *Arapaima gigas*

Nos capítulos anteriores, enfatizamos a importância do estudo da espécie, categorizando e elencando os motivos os principais trabalhos publicados em revistas internacionais e nacionais. Também podemos observar que a área de biologia molecular vem sendo um foco principal nos estudos com *Arapaima gigas*. Entre 2016 e 2021, 45 trabalhos são encontrados no PubMed, dentre os quais, 19 artigos trabalham com alguma parte de genômica/proteômica do peixe.

Tomemos o trabalho realizado por Kang Du e colaboradores (2019), pesquisadores da Universidade de Wuerzburg, na Alemanha, que em parceria com outras instituições de pesquisa da China, França e Brasil, sequenciaram por completo o genoma do *Arapaima gigas*. Isso é, em outras palavras, identificar cada sequência de nucleotídeo que compõe o material genético do peixe. Com avanço da tecnologia, hoje as informações presentes nesses arranjos de nucleotídeos são rapidamente mais elucidadas. Enfatizaram a existência de regiões que revelam o grau protetor mais forte no macho, do que na fêmea, cujas

características estão diretamente ligadas ao órgão secretor, localizado na cabeça. Também é evidenciado a presença de *scaffolds* (trechos longos de DNA) específicos diretamente relacionados ao cromossomo Y (relacionado ao sexo masculino) que, podem ser usados para identificar e determinar o sexo do peixe, melhorando assim, o controle da espécie na aquicultura e *in natura*.

Watanabe e colaboradores (2018), em estudos realizados na Universidade do Pará, utilizam a técnica a fim de partir de um transcriptoma (conjunto de transcritos, modificadores do genoma no processo de transcrição) com amostras do fígado e da pele, coletados tanto em machos quanto fêmeas de *Arapaima gigas*. Esse processo, nomeado de “*De novo*”, conseguiu diferenciar a expressão baseada no dimorfismo sexual, trazendo mais uma vez uma nova possibilidade de utilizarmos o genoma como fonte de trabalho, como por exemplo, em aspectos reprodutivos.

No Brasil, temos mais um trabalho realizado por Lucas Torati e colaboradores (2019), pesquisador da Embrapa-Tocantins, em parceria com instituto do Pará e com pesquisadores da Escócia e da Alemanha, no objetivo de avaliar a diversidade genética para preservação da espécie. Focaram o estudo nas modificações nucleotídicas, como o polimorfismo de nucleotídeo único (SNP), que consiste na alteração de único nucleotídeo na sequência genética. Concluiu-se que 448 SNPs estão presentes em mais de 80% dos indivíduos de cinco populações, com maior polimorfismo relacionado a população amazônica e do Solimões, quando comparado ao rio Araguaia. Esses estudos podem prever sensibilidade do *Arapaima gigas* a fatores ambientais, drogas, compostos químicos, como também determinar a origem da espécie. Uma vez que se utiliza muito desses agentes no Brasil em aquicultura, mapear as espécies promove maior controle na captura do seu ambiente natural.

2.3.2.1. Aplicações das metodologias no Centro de Biotecnologia do Instituto de Pesquisas Energéticas e Nucleares

Durante a descrição feita no tópico anterior, podemos perceber que os trabalhos visam contribuir na conservação da espécie, baseada em técnicas utilizadas da biologia molecular. Seria muito extenso listar e discutir todos os trabalhos relacionados ao Pirarucu neste capítulo; entretanto, com artigos conceituados em revistas de alto impacto, descrevemos sucintamente a relação entre os objetivos da pesquisa envolvendo o material genético do *Arapaima gigas*. No grupo de pesquisa do Centro de Biotecnologia do Instituto

de Pesquisas Energéticas e Nucleares, a discussão não foge da conservação da espécie, porém, direciona o trabalho para síntese dos hormônios hipofisários da espécie.

A tecnologia do DNA recombinante, introduzida no início desse capítulo, servirá de base para entendimento das pesquisas realizadas no centro. A identificação dos genes relacionados aos hormônios hipofisários é resultado de um trabalho complexo, atrelado à combinação das técnicas de bioinformática e biologia molecular. Procura-se primeiro identificar as espécies com grau de parentesco próximo (filogeneticamente), para depois procurar, em banco de dados genéticos, sequências depositadas relacionadas ao gene de interesse. Atualmente, esse trabalho está mais simplificado, visto que os avanços ao longo dos anos permitiram maior disponibilidade de informações em banco de dados, como o genoma do *Arapaima gigas*.

Vale ressaltar, no entanto, que o grupo de pesquisa foi pioneiro na identificação dos genes relacionados a subunidade β do hormônio folículo-estimulante (ag- β FSH) e luteinizante (ag- β LH) de *Arapaima gigas* (BARTOLINI *et al.*,2015), assim como a subunidade α comum entre os hormônios hipofisários glicoproteicos (ag- α GTH) (FARIA *et al.*,2013). De extrema relevância para os futuros trabalhos, foi classificado o *Arapaima gigas* em uma árvore filogenética, correlacionando as espécies próximas em nível genômico. Utilizando técnicas da biologia molecular, como a metodologia de Rápida Amplificação das Extremidades do cDNA (RACE) para região 3' e 5', esses trabalhos serviram de base para expressar a proteína de *agFSH* em células de mamíferos (discutido no tópico 2.4 da dissertação) e, para realização deste trabalho.

2.3.2.2. Otimização das técnicas para identificação do gene ag- β TSH

Os últimos trabalhos publicados pelo grupo utilizaram técnicas de amplificação das extremidades da cadeia do cDNA, como mencionado anteriormente, pela metodologia do RACE, descrita em detalhes, no tópico referente a essa metodologia. Entretanto, não podemos descartar outras estratégias para identificar um gene não descrito, como foi o caso do ag- β TSH. A abordagem desse subtópico está justamente na otimização de técnicas, que foram complementadas, para desenvolvimento do trabalho.

O avanço da bioinformática foi fundamental para que pudéssemos traçar e delimitar novos passos aplicados nesta pesquisa. Softwares de livre acesso e uma infinidade de informações que são recuperadas com poucos cliques, podem garantir uma segurança adicional nos resultados obtidos. Tomemos o exemplo da clonagem e caracterização de um

gene novo. Até o presente momento, havíamos apenas repetidos passos para clonagem do ag- β TSH, ag- β LH e ag- β FSH, delimitados em **quatro** etapas:

- I. Identificar espécies filogeneticamente próximas com o gene já publicado.
- II. Alinhar sequências das espécies escolhidas e desenhar *primers* em regiões conservadas do DNA.
- III. Amplificar o gene com reações em cadeia da polimerase (PCR) a partir do mRNA extraído previamente da hipófise de *Arapaima gigas*.
- IV. Sequenciar, caracterizar e prosseguir com a clonagem em vetores de expressão.

No entanto, as etapas **I**, **II** e **III**, apresentam algumas variáveis, criando alguns obstáculos nessa investigação. Por partes, esclareceremos tais dificuldades e informaremos as otimizações feitas no processo a fim de garantir melhor entendimento.

A etapa **I** parte do princípio que o gene alvo, que buscamos no nosso trabalho, esteja publicado com espécies filogeneticamente próximas. Isso nem sempre está disponível ou apresentam apenas uma pequena região conservada, dificultando a etapa **II**. Como esta etapa é dependente da anterior, caso este alinhamento apresente apenas uma pequena região conservada, a limitação do desenho dos primers se dá em dois vieses: uma primeira amplificação, parcial, resultante da orientação *sense* e *antisense* e, das condições do desenho do primer (temperatura de anelamento, estruturas secundárias, etc.). Subsequente, a etapa **III** e **IV** não serão realizadas. Além disso, a penúltima etapa depende da qualidade da extração do RNA, extremamente crucial, caso as duas primeiras etapas tenham sucesso.

Quando pensamos em otimizar o trabalho, duas estratégias são sugeridas, adicionando ou excluindo etapas. A bioinformática foi uma aliada ao adicionar etapas nesse processo, facilitando o caminho. Com os trabalhos publicados (DU *et al.*, 2019; VIALLE *et al.*, 2018), alguns facilitadores entre as etapas foram adicionados:

- A. Manipular e alinhar a sequência encontrada no genoma do *Arapaima gigas* [BLAST, (ALTSCHUL *et al.*, 1990)]
- B. Caracterizar regiões codificantes encontradas no alinhamento a partir da metionina inicial [ORF-Finder (ROMBEL *et al.*, 2002)]

- C. Observar se ocorre o splicing do gene e, caso positivo, construir o CDS (*coding sequence*) [Splign (KAPUSTIN *et al.*, 2008) e ProSplign (KIRYUTIN; SOUVOROV; TATUSOVA, 2007)]
- D. Sintetizar o cDNA em plasmídeo de expressão

Esquematisando e alocando esses facilitadores, teríamos o esquema completo: I e II-A-B-C, III e IV-D. O próximo tópico deste trabalho irá abordar a bioinformática no seu contexto e suas aplicações dentro da biologia molecular.

2.3.3. Bioinformática e aplicações

A terminologia “bioinformática” começou a ganhar relevância em meados da década de 70. Uma técnica conceituada como um processamento da informação para maior entendimento dos organismos vivos, apresenta um dinamismo vivo. Sua compreensão vem de forma construtiva, ao passo de novos estudos publicados ao redor do mundo, definindo uma biologia evolucionária. Um exemplo claro é como no começo no século XXI tivemos o projeto de sequenciamento do genoma humano e, em 2015, a manipulação deste, com a metodologia de edição gênica *Clustered Regularly Interspaced Short Palindromic Repeats* (CRISPR). O mais importante dentro da bioinformática, é a organização da informação. A maneira que é transmitida garante ao próximo pesquisador qual o ponto de partida deve-se seguir com o compilado de dados (HOGEWEG, 2011).

A biologia, portanto, toma outro rumo, com sua implementação. O termo *in silico* ganha espaço, retratando as descobertas realizadas na área, com auxílios de algoritmos cada vez mais aprimorados. Esse reflexo, quando paramos para discutir, é observado na agilidade de novos estudos de desenvolvimento de drogas, mapeamento genético, toxicidade, elucidar rotas biológicas, etc. (RAJPAL, 2005). Os estudos, somados, são imprescindíveis em tempos atuais; uma importante ferramenta que não pode ser deixada de lado quando se trabalha com organismos vivos (YU *et al.*, 2004).

Alguns destes algoritmos utilizados na bioinformática envolvem a manipulação dos ácidos nucleicos DNA e RNA (genômica e transcriptômica), tais como modelos de proteínas (proteômica) e suas interações. Com maiores informações alocadas em base de dados, bioatividade, metabolismo proteico, conformação e disposição espacial, entre muitas possibilidades que existem, cinco desafios são impostos na manipulação: veracidade, variabilidade, volume, velocidade e variedade (CHEN, HUANG, WU, 2017), parâmetros que são levados em consideração a fim de solidificar uma informação.

2.3.4. Bioquímica estrutural de proteínas

Reservamos um tópico dedicado à bioquímica estrutural voltada para proteínas a fim de justificar a implementação das metodologias no presente trabalho.

A área de estudo e determinação da estrutura proteica tem origem na cristalização e uso das técnicas de difração por raios-x. Em 1913, começam os estudos com Max von Laue, William Henry Bragg e William Lawrence Bragg, provando que cristais difratam raios-x, sendo possível determinar os átomos presentes nas estruturas proteicas. Em menos de 50 anos, em 1957, e com 22 anos de trabalho, John Kendrew, em Cambridge (Londres, Inglaterra) usou a técnica de difração para resolver a estrutura da mioglobina, recebendo o prêmio Nobel em 62 (KENDREW *et al.*, 1960). Interessante frisar que atualmente - junho de 2022 -, 65 anos depois, o maior banco de dados de estrutura de proteínas, RCSB, possui 191328 estruturas resolvidas por diferentes técnicas, sendo as três principais compreendendo a difração por raio x (XRD), ressonância nuclear magnética (NMR) e crio microscopia eletrônica (cryoEM). Outras metodologias podem ser empregadas, entretanto, não será o foco da discussão.

Com a introdução desse assunto, perguntamos: qual a importância de explorar a estrutura tridimensional da proteína? Apesar de ser uma resposta extensa e aberta, detalho de maneira objetiva e clara. Hoje, com o avanço científico, tecnológico e acesso às mais diversas técnicas de resolução de estruturas proteicas, conhecer o enovelamento proteico, a disposição espacial dos aminoácidos e seu possível comportamento relacionado a sua atividade, exploram-se as aplicações dessas metodologias no descobrimento de novas drogas, alternativas de mutações a fim de induzir ou reprimir a função proteica e entender quais são os principais aminoácidos que conferem atividade metabólica, carreadora ou sinalizadora.

Entretanto, o processo de cristalização não é simples. Cada proteína possui um comportamento, formação e ordenação dos cristais que podem ser mais complexos ou menos. Não entrarei em grandes detalhes sobre aspectos físicos e matemáticos de ordenação dos cristais, índices de miller e planos tridimensionais pois o presente trabalho propôs implementar metodologias *in silico* na resolução da estrutura proteica. Das mais amplas metodologias com o uso da bioinformática, especialmente hoje com o algoritmo de

inteligência artificial (IA) *AlphaFold2*, temos uma segurança maior em propor modelos fidedignos aqueles resolvidos por cristalografia.

2.3.4.1. AlphaFold2: a revolução da proposta para resolução de estruturas proteicas

Este subtópico apresentará uma descrição do artigo recém publicado na revista *Nature*: "*Highly accurate protein structure prediction with AlphaFold*" (JUMPER *et al.*, 2021) assim como sua importância dentro do campo científico. Até então, havíamos algumas alternativas na resolução de proteínas que não foram cristalizadas, sendo duas principais:

- I. Homologia: A partir de estruturas proteicas já resolvidas, os conceitos de conservação entre espécies e do enovelamento em determinadas regiões, poderiam ser utilizados para resolver a estrutura desconhecida.
- II. *ab initio* (BONNEAU; BAKER, 2001): do latim "do início", apresenta-se com o objetivo de prever a estrutura terciária da proteína sem utilizar a homologia. Nesse caso, apenas com a sequência primária dos aminoácidos conseguimos chegar na estrutura terciária.

A IA *AlphaFold2* é apresentada para inovar nesse quesito. Utilizando arquiteturas de programação relacionada a geometria, propriedades físicas e evolucionárias das proteínas, a linguagem consegue prever as coordenadas tridimensionais dos átomos apenas com a sequência primária dos aminoácidos. A similaridade entre os modelos gerados com aqueles cristalizados em laboratório foi surpreendente, chegando a um nível de identidade muito alto.

Através de um algoritmo denominada *Evoformer*, a informação vai e volta, dentro de um bloco que analisa múltiplos alinhamentos com a sequência (MSA) e procura a melhor solução dentro de N alinhamentos com sequência de aminoácidos em vários organismos. A linguagem neural - da qual se entrará em maiores detalhes - vem do princípio de *transformer* (transformadores, em inglês), introduzido em 2017 em Cornell (VASWANNI *et al.*, 2007) com um princípio de rerepresentar a informação múltiplas vezes e categorizar qual a melhor (dentre as apresentadas) se encaixa para a informação (no caso, a sequência primária de aminoácidos).

Algumas informações são importantes em relação a construção do modelo e sua classificação. Em uma escala de 0 a 100, cada resíduo tem uma pontuação estimada do seu posicionamento em relação ao C α IDDT. Para computar esse valor, o algoritmo recupera de estruturas cristalizadas com o B-factor (entretanto, não segue o mesmo parâmetro de pontuação). Quanto maior o pLDDT, melhor a confiabilidade do seu modelo e das cadeias laterais dos aminoácidos. Maiores detalhes da documentação do AlphaFold podem ser encontrados em <https://alphafold.ebi.ac.uk/faq>.

O uso da ferramenta vem trazendo um mar de possibilidades para resolução de estruturas das quais ainda não foram resolvidas, ampliando principalmente os conhecimentos acerca da biologia evolucionária e dando mais um passo para o desenvolvimento/identificação de fármacos com base na estrutura (*drug-based*). Atualmente – 21 de setembro de 2022 - o repositório de proteínas resolvidas por AlphaFold2⁵ está com 1,000,361 estruturas depositadas dos mais diversos organismos. Um número expressivo de moléculas que vem crescendo exponencialmente.

Ainda assim, não podemos e nem devemos excluir a cristalização e resolução pelas mais diversas técnicas, especialmente quando queremos entender sua conformação na presença de um agente externo, como fármaco, íons, cofatores, DNA/RNA e, também, com relação a suas modificações pós-traducionais, um dos principais processos em células eucarióticas. Essa informação ainda não é possível resolver apenas com o algoritmo de IA. Entretanto, pode ser uma solução na complementaridade das informações a respeito de uma determinada proteína.

2.3.4.2. Aplicações do uso de modelos estruturais de proteínas na biologia computacional

De forma que podemos explorar as propriedades físico-químicas e bioquímicas das proteínas, a dinâmica molecular é uma ferramenta computacional moderna que surge em meados da década de 70 a fim de entender o comportamento de *folding* das proteínas (LEVITT; WARSHEL, 1975). Essa avaliação é feita simulando a proteína em um determinado ambiente e seu comportamento com ligantes, receptores ou até as cadeias laterais dos aminoácidos que compõe sua estrutura.

⁵ AlphaFold Protein Structure Database. Disponível em: <https://alphafold.ebi.ac.uk>. Acesso em 14.jun.22

Sucintamente, de acordo com as técnicas que serão utilizadas nesse trabalho, apresentamos duas técnicas que são aplicadas na dissertação

: I) *Molecular Mechanics with Generalized Born and Surface Area solvation* (MM/GBSA) explora o valor de energia livre ($\Delta G_{\text{bind}} = \text{kcal/mol}$) entre dois ligantes na presença de um solvente. O cálculo de ΔG_{bind} resulta da diferença da energia livre do complexo solvatado ($\Delta G_{\text{complex,solvated}}$) com a soma dos ligantes solvatados ($\Delta G_{\text{receptor,solvated}} + \Delta G_{\text{ligand,solvated}}$) (MILLER *et al.*, 2012). O valor é expresso em kcal/mol.

Já a técnica de *Umbrella Sampling*, a metodologia avalia através de uma reação coordenada (ξ , em Å) e calcula as energias livres durante os intervalos da coordenada da reação, através do potencial de força médio (ΔG_{PMF}), em kcal/mol (KÄSTNER, 2011). O experimento tem como fundamento a simulação aumentando a distância da coordenada, resultando no afastamento das moléculas de interesse, proporcionando uma demanda maior de energia para o sistema se manter estável.

2.4. Expressão de proteínas recombinantes utilizando células de mamíferos

2.4.1. Uso de organismos vivos para produção de proteínas e regulamentação das agências sanitárias

Anteriormente, compreendemos com clareza a utilização da tecnologia do DNA recombinante e as possibilidades que foram adquiridas ao longo do seu desenvolvimento. Como a ciência é dinâmica, acompanhada na maioria das vezes com o avanço das tecnologias, novas perspectivas e metodologias vão sendo adquiridas, otimizando custo, tempo e mantendo a qualidade do produto. Esse capítulo se afunila para o uso de células eucarióticas, mais especificamente, de mamíferos, como sistemas de expressão de proteínas recombinantes. As mais utilizadas, dentro desse sistema, são as originárias de ovário de hamster chinês (CHO) e as de rim de embrião humano (HEK293), sendo que as células CHO, por muitas das vezes, acabam sendo a primeira opção de cultivo e transfecção. Outras opções de células de mamíferos também incluem: fibroblastos renais de hamster bebê (BHK), células humanas de fibrossarcoma (HT-1080) e células de mieloma murino (NS0). A popularidade das células CHO veio em meados dos anos 50 com Theodore T. Puck, enquanto as células HEK293 teve seu primeiro trabalho publicado em 1977, com trabalhos desenvolvidos primeiramente na América do Norte (GRAHAM & SMILEY, 1977).

Apesar de apresentar-se como uma técnica relativamente nova, sofisticada e de alto custo, a transição do sistema em *Escherichia coli* para produção de proteínas heterólogas para organismos eucarióticos foi gradativa. Desde os anos 60, era utilizado os conhecimentos obtidos na construção de vetores (uso de promotores virais, caudas de moléculas para purificação) em virtude do domínio do hospedeiro, ajustado às condições ideais de cultivo na obtenção de proteínas recombinantes. Além disso, os resultados não são insatisfatórios; grandes quantidades de proteína íntegra são expressas e com um custo, atualmente, relativamente baixo. Entretanto, sabemos da complexidade de uma proteína e suas modificações que são realizadas após a tradução e que, apenas células de mamíferos, possuem o mecanismo de glicosilação, uma importante etapa no seu enovelamento (discutido neste tópico). A maior parte da produção de biofármacos está na utilização de células isoladas de eucariotos a fim de sustentar uma possibilidade de tratamento de hormônios, enzimas, fatores de coagulação, assim como produção de vacinas.

O órgão sanitário *Food and Drug Administration* (FDA) que regula a produção dessas proteínas e o *Therapeutic Biologics Applications* (BLA), delimitam etapas e parâmetros na produção a fim de garantir maior segurança quando o produto é lançado no mercado. Dentre esses parâmetros, na produção de vacinas, por exemplo, enfatizamos os qualitativos, que compreende: porcentagem máxima de impurezas, organismo de origem, processo de produção, propriedades físico-químicas, atividade biológica, ligação com o alvo, produto de referência e estabilidade⁶. Em junho de 2021, o FDA e o *Center for Drug Evaluation and Research* (CDER) atualizam o compilado de moléculas bio-terapêuticas aprovadas. São 355 patentes produzidas por grandes indústrias farmacêuticas com diferentes concentrações e apresentações, disponibilizadas no *Purple Book* da agência. Aproximadamente 66 biomoléculas recombinantes originadas de *Escherichia coli* são regularizadas pelo FDA (GILEADI, 2017). Por fim de comparação, a Agência Nacional de Vigilância Sanitária (ANVISA) possui uma regulação RDC 55/2010 que está bem alinhada com a agência sanitária dos Estados Unidos (FDA) e da Europa (EMA), com estudos compreendendo desde sua química e produção até estudos clínicos e sua aprovação. Hoje, são 14 os principais medicamentos utilizados dentro do cenário nacional (PRIVATO, MARTINEZ, SCHMIDT, 2020).

Nos próximos tópicos se pretende discutir e especificar quais são os principais usos em cada tipo de célula mencionado (HEK293 e CHO), uma vez que durante o trabalho, os dois sistemas de expressão foram utilizados. Antes de adentrarmos, a colocação mais importante é de que não haverá discussão entre vantagem e desvantagem, relacionadas às células. Há o melhor sistema de expressão para a proteína alvo, ou seja, aquela que o pesquisador/indústria quer obter no seu resultado final.

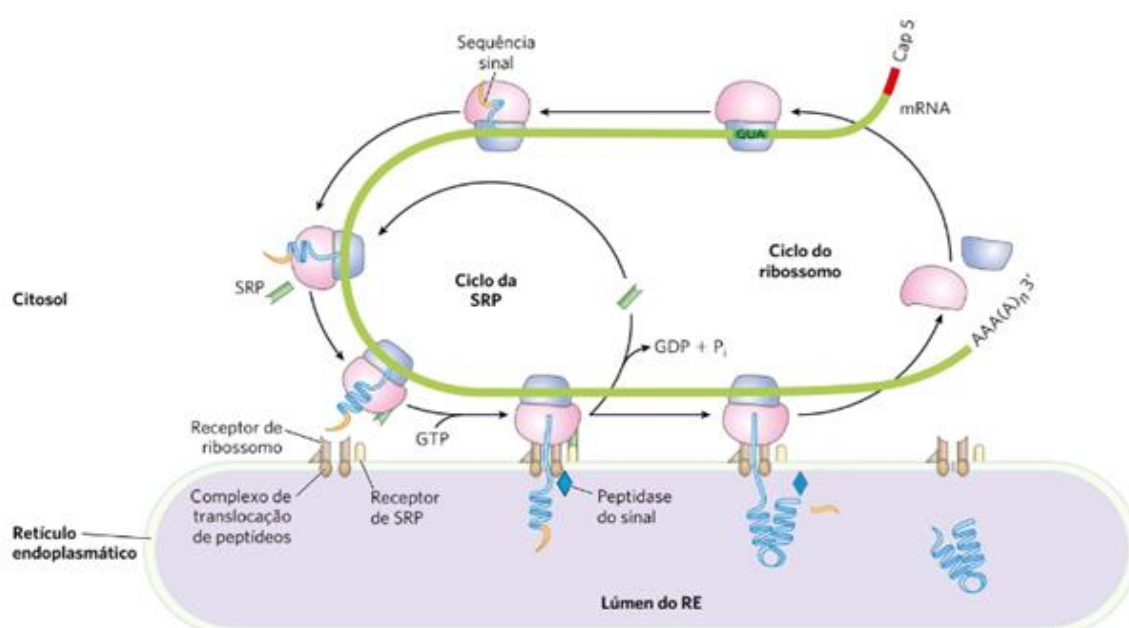
2.4.2. Processo de expressão proteica em eucariontes

O processo de síntese de proteínas difere em eucariotos e procariotos, uma vez que a organização celular de ambos é bem diferente, gerando processos pós-traducionais mais complexos. A maioria das proteínas eucarióticas, no processo de tradução, tem o endereçamento final no retículo endoplasmático (RE), especificamente em ribossomos aderidos à organela em questão. Esse direcionamento é feito por meio de um *signal-peptide* (uma sequência de 13-36 aminoácidos), em que, ribossomos livres, no citosol, começam a

⁶ Disponível em: www.fda.gov/regulatory-information/search-fda-guidance-documents/characterization-and-qualification-cell-substrates-and-other-biological-materials-used-production. Acesso em: 18 de out. 2021

sintetizar essa sequência e, à medida que começa a emergir no lúmen do RE, o *signal-peptide* se liga a um *signal-recognition particle* (SRP), fundamental para o alongamento posterior da cadeia de aminoácidos em questão (Figura 6). A figura detalha o início da tradução do RNA mensageiro (mRNA) com o *signal-peptide* até o momento em que o ribossomo adere ao retículo endoplasmático e dissocia, reciclando a organela e finalizando a tradução. Todo esse processo de alongamento da cadeia proteica tem início nos ribossomos livres e término no RE, região na qual os principais processos pós-traducionais ocorrem e determinam a atividade específica da proteína ou do peptídeo (NELSON & COX, 2018).

Figura 6 – Síntese e endereçamento proteico.



Abreviações: GTP – Trifosfato de guanosina; GDP – Difosfato de Guanosina; P – Fósforo; RE: Retículo Endoplasmático

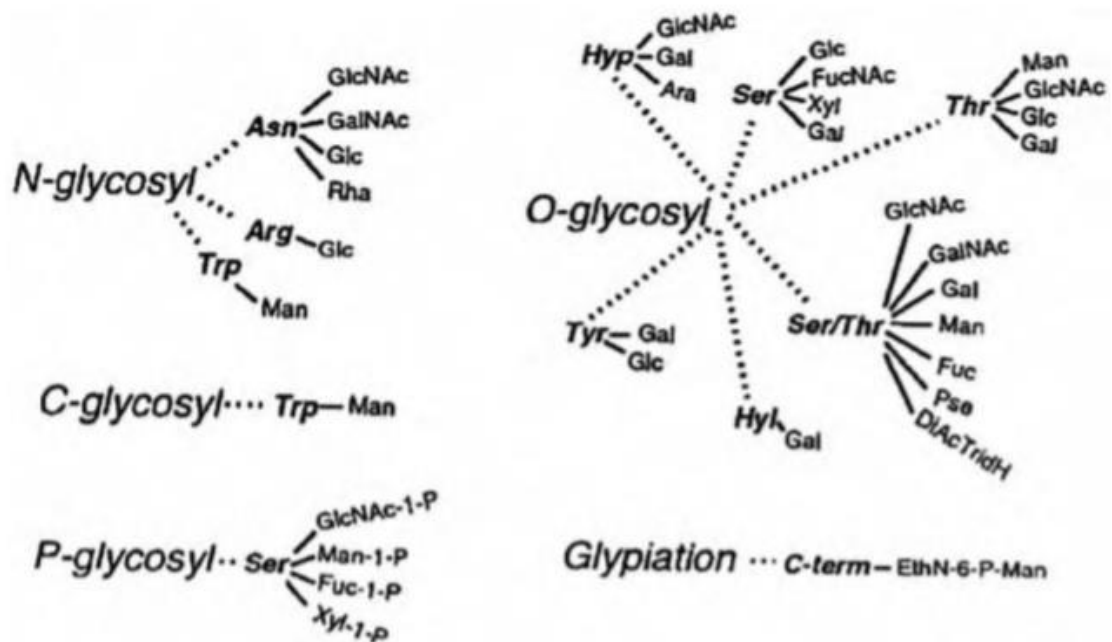
Fonte: Adaptado de Nelson & Cox, 2018.

Grande parte dessas proteínas contém grupos covalentes ligados à sua estrutura. As modificações pós-traducionais (MPTs) são parte importante do processo de finalização da síntese proteica. (ALBERTS *et al.*, 2010). Compreende-se as modificações como fosforilação, acetilação, metilação, acilação, glicosilação, sulfatação, formação de pontes dissulfeto, ubiquitinação, hidroxilação, lipidação (MANN; JENSEN, 2003) e, recém descoberta, a sumoilação – adição do grupo SUMO (*Small ubiquitin-like modifier*) (WILKINSON; HENLEY, 2010). Todos os processos citados acima determinam uma modulação proteica, da qual, nenhuma é mais ou menos importante. Daremos importância ao processo de glicosilação – especificamente o tipo N – apenas para conhecimento sobre uma etapa que é presente na subunidade β que compõe o hormônio estudado desse trabalho.

2.4.2.1. Modificações pós-traducionais: glicosilação

O processo de glicosilação, presente em organismos do tipo eucariotos e arqueas, consiste na adição de oligossacarídeos à proteína recém-sintetizada no RE. Existem cinco diferentes tipos: N-glicosilação, O-glicosilação, C-glicosilação, gliptação e fosfoglicosilação (HAYNES, 1998) variando conforme a posição do açúcar na cadeia proteica (**Figura 7**).

Figura 7 - Os cinco diferentes tipos de glicosilação encontrados na literatura.



Fonte: Walker-Nasir *et al.*, 2007.

Nota-se que, o aminoácido Asn em N-glicosilação e Ser/Thr em O-glicosilação possuem maiores números de glicanos, por serem mais comuns. O prefixo “N” se refere a ligação do oligossacarídeo em um resíduo do átomo de nitrogênio da asparagina (SCHWARZ; AEBI, 2011) e “O” pela ligação no átomo de oxigênio de serina ou treonina (TOSHIMA; TATSUTA, 1993). A glicosilação do tipo “C” é a adição da cadeia de açúcar no átomo de carbono, processo que raramente acontece (DE BEER *et al.*, 1995).

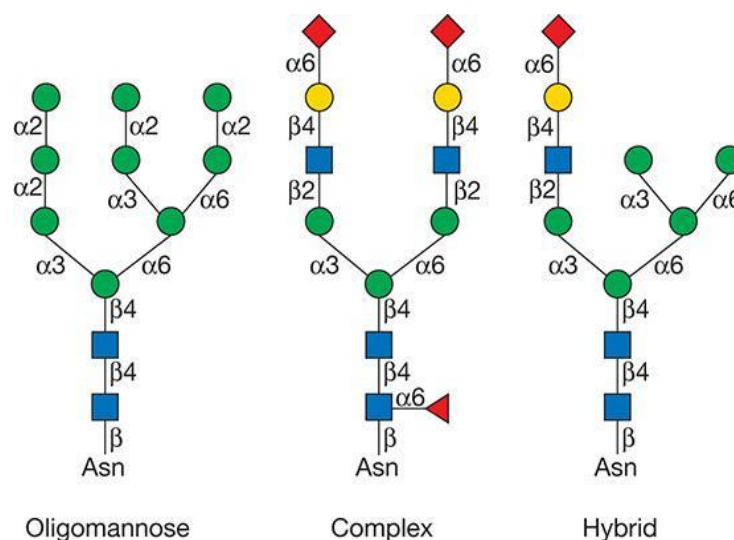
Não menos comum que o tipo N e tipo O, a gliptação é a adição do grupo ancorador glicosilfosfatidilinositol (GPI) no carbono c-terminal da proteína. A estrutura contém uma região fosfolipídica, com diversas funções de adesão celular, reconhecimento imune, entre outros (COUSSEN *et al.*, 2001). O processo de fosfoglicosilação ocorre apenas em parasitas

e alguns tipos de fungos, onde há a transferência de oligossacarídeos ao resíduo livre de serina ou treonina por ligação fosfodiéster (HAYNES, 1998).

As formas variáveis de glicosilação mencionadas acima determinam funções físico-químicas e biológicas à proteína, dentre elas: a estabilização da conformação proteica; modificação da massa, viscosidade e tamanho; regulação da sinalização intracelular e sua localização; modificação nos processos imunológicos (LIS; SHARON, 1993).

A glicosilação do tipo N é a mais comum, ocorrendo em cerca de 70% das proteínas. Consiste na transferência de um núcleo pentassacarídeo ($\text{Man}_3\text{GlcNAc}_2$) ao resíduo de asparagina, formando a estrutura base denominada *sequon* (BUTLER, 2006). Ainda no retículo endoplasmático, a partir de uma molécula de fosfato de dolicol, ocorrem sucessivas adições de cadeias glicosídicas serão adicionadas e, após as ações de glicosidases e glicosiltransferases, o N-glicano maturo pode assumir três diferentes formas: o tipo Oligomanose, possuindo apenas resíduos de Man ligados a sua estrutura; o tipo complexo, contendo um “antena” ligado a estrutura base ($\alpha 6$) e, o tipo híbrido, que pode conter uma prolongação de resíduos de manose em uma parte e outra parte com resíduos de GlcNAc (STANLEY *et al.*, 2017) (Figura 8) em que todas as estruturas possuem um resíduo base de $\text{Man}_3\text{GlcNAc}_2$.

Figura 8 – Diferentes tipos de N-Glicanos.



Em inglês, temos a tradução: **Oligomanose:** Oligomanose; **Complexo:** Complexo; **Híbrido:** Híbrido. As formas geométricas representam: **Quadrado azul:** N-acetilglucosamina; **Círculo verde:** Manose; **Círculo amarelo:** Galactose; **Triângulo vermelho:** Fucose; **Losango vermelho:** Ácido siálico.

Fonte: STANLEY *et al.*, 2017

2.4.2.2. Distúrbios congênitos da glicosilação (CDG)

A deficiência ou alteração nos carboidratos geram efeitos notórios na atividade biológica da proteína. Ajit Varki (1993) publicou um estudo em que reuniu diversos efeitos negativos em diferentes estruturas proteicas, reforçando a importância e impacto dos açúcares, assim como seu estudo. Essas deficiências são classificadas em um grupo como distúrbios congênitos da glicosilação (CDG): são doenças genéticas do tipo autossômica, rara, com a maioria dos casos resulta de uma hipoglicosilação afetada pelo processo de N-glicosilação (JAEKEN, 2010). Dentre os efeitos negativos provocados pela doença, podemos exemplificar pacientes que apresentaram microcefalia, epilepsia focal e doença congênita no coração; todos os pacientes apresentaram uma variação no gene de um tipo de oligotransferase (BRYANT *et al.*, 2020).

2.4.2.3. Estudos que abordam a glicosilação em proteínas de peixes

Contextualizando os processos de síntese proteica e da modificação pós-traducional em questão, poucos são os trabalhos publicados que evidenciam a importância da glicosilação de proteínas em peixes - especificamente quando se trata de hormônios ligados à reprodução (KAMEI *et al.*, 2003; KASUTO; LEVAVI-SIVAN, 2005). Os trabalhos publicados destacam a importância da comparação do produto nativo com o produto sintético (no que diz respeito a glicosilação e sialilação) para melhor entendimento da função biológica *in vivo* (KAZETO *et al.*, 2008). Dessa forma, contribuiria para futuros tratamentos, com objetivos determinados, seja no melhoramento da reprodução em cativeiro ou no incremento de massa e, até mesmo, numa desova mais acelerada.

2.4.3. Componentes do meio de cultura sem soro (*Serum-Free*) para células de mamífero

Uma síntese bem elaborada publicada por Li e colaboradores (2021) reforça a importância do estudo do meio de cultura utilizado. Nos últimos 20 anos, essa busca de compreensão está diretamente relacionada com a produção de biomoléculas, especialmente na síntese de anticorpos (um dos mais importantes ramos dentro da indústria biotecnológica) e como afeta no padrão de glicosilação das proteínas. São oito os componentes presentes no meio de cultura sem soro (*serum-free*); hoje, podemos também encontrar a terminologia sem proteína (*protein-free*). O armazenamento do meio de cultura (controle de temperatura, em ambiente sem luz, hermeticamente fechado) é imprescindível, sabendo que sua estabilidade depende desses fatores.

Os oito componentes serão descritos resumidamente a seguir. As fontes de energias, majoritariamente compostas por açúcares (glicose, fucose, manose), fornecem energia para cultura celular; os aminoácidos, sendo a principal fonte de nitrogênio, também atuam como fonte energética e de síntese de asparagina e glutamina; temos também, as vitaminas, cuja principal função é manter o crescimento celular, monitorando o metabolismo intracelular.

Os lípidos que constituem parte da membrana celular, são importantes como reserva energética e sinalização celular e podem até contribuir em maiores níveis de glicosilação. Os elementos traços, como ferro (Fe), zinco (Zn), manganês (Mn), níquel (Ni), regulam atividades enzimáticas, atuando muitas vezes como co-fatores. Em culturas em suspensão, as células são submetidas a rotação constante, então, se dá a importância de um anti-atrito. O surfactante *plutonic-F68* é comumente usado nesses casos.

Concluindo, sais inorgânicos, com sódio (Na), potássio (K), fósforo (P), ajudam a manter a pressão osmótica celular, característica individual para cada cultura e organismo. Os ácidos nucleicos são expressos normalmente no meio; quando o caminho de síntese é bloqueado (por exemplo, células DHFR - *dihydrofolate reductase-deficient*), é adicionado hipoxantina e timidina, que regulam a síntese do DNA e podem até mesmo influenciar na glicosilação. Antioxidantes e ácidos carboxílicos podem ser utilizados a fim de regular a expressão proteica.

Vimos, portanto, neste subtópico a importância do meio de cultura no desempenho e desenvolvimento do trabalho. Cada componente utilizado possui uma função particular ou específica, porém, sinérgica quando em conjunto, constituem um ambiente favorável para sua proliferação. Organismos vivos *in vitro* exigem uma cautela adicional nos experimentos, garantindo o sucesso do produto final.

2.4.4. Células de ovário de Hamster Chinês (CHO)

2.4.4.1. Contexto histórico e sua utilização

Uma busca rápida em base de dados com o termo “CHO cells” e rapidamente nos deparamos com inúmeros trabalhos, publicações e dissertações que utilizaram o organismo com geralmente o mesmo foco: expressão proteica. Apenas em 1987 com a regularização da primeira proteína recombinante produzida (ativador do plasminogênio tecidual - rTPA), que trabalhos focando melhorar e otimizar sua expressão foram sendo desenvolvidos. Linhagens celulares e modificações genéticas estão sendo realizadas em um ritmo acelerado, com a

batelada de informações que são adquiridas e disseminadas. Não nos surpreende que a engenharia de células e tecidos seja um campo novo, mas com grande potencial. Produtividade, qualidade e estabilidade são os que grandes grupos de pesquisadores e empresas buscam entregar.

Poucos são os obstáculos quando utilizamos as células CHO. Apresenta-se algumas vantagens, classificadas rapidamente como: crescimento massivo de células em meio quimicamente definido e livre de soro, uma replicação segura e habilidade de produzir proteínas recombinantes com modificações pós-traducionais similares (FISCHER, HANDRICK, OTTE, 2015). Compreende o sistema mais empregado na produção de anticorpos monoclonais, proteínas recombinantes humanas para diferentes tipos de tratamentos e, direcionado também, na produção de vacinas (NYON *et al.*, 2018; ELDI *et al.*, 2017; BRAZZOLI *et al.*, 2005).

Inúmeros meios de culturas foram adaptados para a linhagem celular. Antigamente, suas composições variam com adição de soro (humano ou de feto bovino), plasma ou extrato de tecidos, influenciando no desenvolvimento celular e aumentando o risco para contaminação; eliminando esses fatores e definindo novos suplementos, os meios sem soro (*serum-free*) ganham espaço com pequenas adições fundamentais de moléculas que garantem os três fatores mencionados: estabilidade, produtividade e qualidade (RITTACO, WU, KHETAN, 2018).

2.4.4.2. Células CHO no Centro de Biotecnologia do Instituto de Pesquisas Energéticas e Nucleares

A expertise e pioneirismo no Brasil com uso de células para expressão de proteínas recombinantes começa no final dos anos 90. Desde então, produções foram sendo realizadas com o organismo vivo, até a última publicação do grupo de pesquisa, em 2019, com caracterização dos N-glicanos da Prolactina humana expressas em CHO aderidas em placas. Antes de discutirmos em outro tópico desse capítulo sobre as células HEK293, busco traçar uma linha do tempo das linhagens celulares utilizadas no CEBIO para produção de hormônios hipofisários, para concluirmos nossa abordagem dentro da linhagem celular.

O uso das células CHO no centro de biotecnologia tem início em 2000, com a produção da prolactina humana - hPRL - que, até então, só tinha sido expresso em *Escherichia coli* (SOARES *et al.*, 2000). Dois anos depois, já seria possível sintetizar hormônios heterodiméricos, como o TSH humano - hTSH - (PERONI *et al.*, 2002). Desde

então, outros trabalhos que envolviam mais as etapas de caracterização e purificação dos hormônios foram ganhando relevância: determinação por cromatografia de fase reversa do hTSH (OLIVEIRA *et al.*,2003), purificação com apenas duas etapas do hTSH (MENDONÇA *et al.*,2005), caracterização de N-glicanos (OLIVEIRA *et al.*,2008) e preparações hormonais com linhagens de CHO (HELLER *et al.*,2010). Em relação às produções com *Arapaima gigas*, não há trabalhos com CHO.

2.4.5. Células de rim de embrião humano 293 (HEK293)

Esse tópico seguirá a mesma apresentação feita anteriormente com as células CHO, dividido em dois subtópicos. Fundamentado em periódicos, livros e artigos, discutimos o isolamento e utilização do sistema e sua importância no contexto da biotecnologia e seu uso no Centro de Biotecnologia do IPEN na produção de hormônios recombinantes humanos e de *Arapaima gigas*.

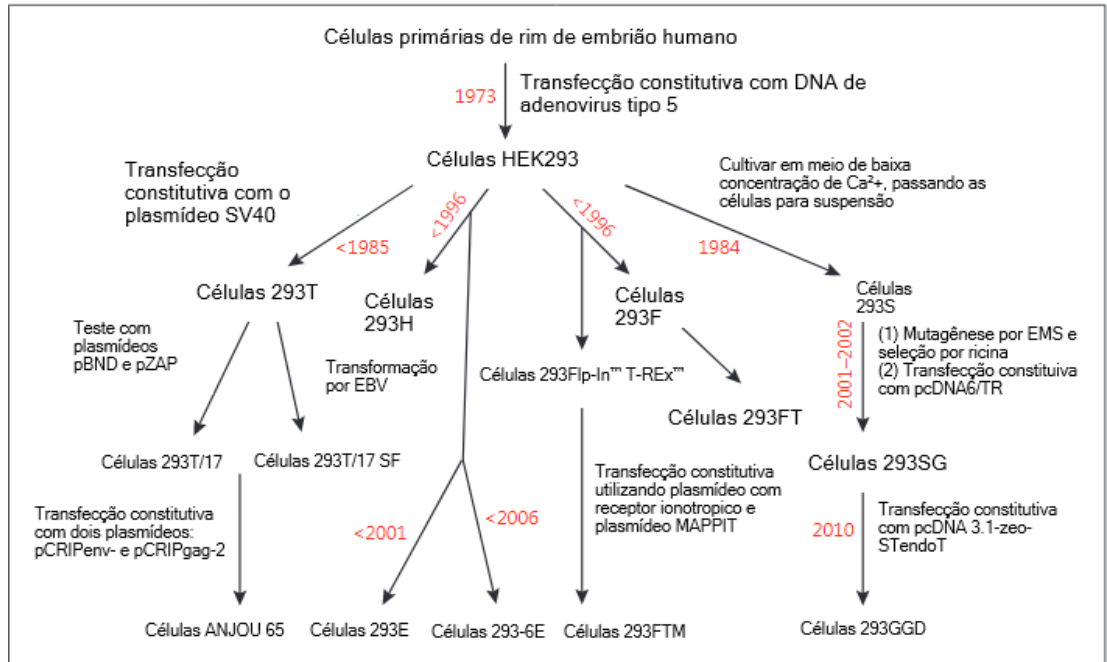
Paralelamente com os sistemas citados anteriormente, as células HEK293 vem ganhando um espaço no cenário da produção de biomoléculas. Brevemente, fazendo um censo quantitativo em outubro de 2021 na base de dados *Web of Science* (WOS), mais de 14.000 estudos estão relacionados com o organismo (com um número crescente vindo desde 2001; em 2020, tivemos 895 publicações, 5,29% a mais do ano anterior). A mesma pesquisa com células CHO está atribuída a mais de 114.000 trabalhos publicados (2020 com 5.883 publicações, 1,44% maior em relação ao ano anterior). Em suma, essa busca por optar em também incluir nos estudos células humanas, está em: otimizar os estudos com proteínas recombinantes (majoritariamente humanas) através da investigação no metabolismo celular que aumente sua produção, comparação do processo de glicosilação e sua interferência final na sua atividade biológica e interação das células com fatores extrínsecos adicionados ao meio de cultura.

2.4.5.1. Contexto histórico e sua utilização

As células de rim de embrião humano são uma linhagem imortalizada que têm, integrada em seu cromossomo, uma certa quantidade de genoma de adenovírus, alterando o seu ciclo celular. Esse foi o primeiro estudo realizado em 1973 para criação da linhagem 293, expandindo rapidamente em 11 anos, onde alguns testes com meio de cultura e transfecções constitutivas geraram novas linhagens que são utilizadas massivamente até hoje. Na figura abaixo (Figura 9) podemos observar o fluxograma com os anos destacados

em vermelho e seu principal acontecimento que originou a variação da célula 293 (HU *et al.*, 2018).

Figura 9: Histórico da utilização das células HEK293.



Fonte: Traduzido de Hu *et al.*, 2018.

Apesar de produzir modificações pós-traducionais similares ao humano, a célula contém um mecanismo que não é encontrado nos humanos: Galactose- α -1,3-galactose (α -gal) e ácido N-glicolilneuramínico (NGNA) (DUMONT *et al.*, 2016). Esses oligossacarídeos causam uma resposta imunogênica, sendo um empecilho para indústrias farmacêuticas e institutos de pesquisas não optarem pela sua utilização para produção de biomoléculas.

A eficiência da utilização de uma célula humana, como a HEK, é a sua alta capacidade de duplicação e principalmente, atua como uma ótima plataforma na expressão de proteínas de membrana e canais iônicos. Quando comparado com as células CHO, esse resultado final é alcançado mais rapidamente. Suas desvantagens são mínimas e superáveis, como a formação de agregados (*clusters*) - dado sua alta taxa de replicação - e contaminação viral humana específica (THOMAS; SMART, 2005), podendo ser contornada com eliminação desses agentes e adaptações nas condições de cultivo celular.

2.4.5.2. Células HEK293 no Centro de Biotecnologia do Instituto de Pesquisas Energéticas e Nucleares

Diferente das células CHO, as células HEK293 tiveram uma implementação recente no laboratório. Apenas três trabalhos, com início em 2017, utilizaram o sistema para proteínas recombinantes: produção do hTSH em células aderidas (SANT'ANA *et al.*, 2017), hormônio folículo estimulante de *Arapaima gigas* em suspensão (SEVILHANO *et al.*, 2017) e hPRL em células aderidas (SILVA *et al.*, 2019). Vale ressaltar que esse trabalho propõe expressar o agTSH em células de mamíferos, que teve expressão maior em HEK293.

3. Materiais e métodos

Do ponto de vista metodológico o projeto de pesquisa foi iniciado em maio de 2020 e teve afetada sua parte prática em laboratório devido às restrições de mobilidade e de acesso causadas pela pandemia SARS-CoV-2. Diante das recomendações sanitárias adotadas, parte de nossa pesquisa foi realizada de forma remoto. Assim, realizamos o adensamento teórico sobre o assunto e as aplicações das técnicas da bioinformática sobre o objeto estudado. Com o retorno das atividades presenciais em laboratório, a realização dos experimentos propostos no período que equivale ao segundo semestre de 2021 à julho de 2022.

A seguir, listamos os tópicos tratados na pesquisa com seus respectivos materiais e métodos utilizados no projeto.

3.1. Hipófises de *Arapaima gigas*

Dez hipófises foram gentilmente coletadas e doadas das instalações da piscicultura Raça (Anexo 1), sendo que sete dessas coletadas no período matutino e três no período noturno. A razão da coleta a noite foi baseada nas hipóteses que o hormônio de crescimento (GH) seja liberado em maior quantidade à noite, outro hormônio estudado pelo grupo. Após abertura do crânio, a hipófise, alojada na sela túrcica, foi retirada com auxílio de uma pinça e, imediatamente, despejada dentro de um criotubo de 1,5 mL. As hipófises foram pesadas em balança semi analítica, contendo sempre cerca de aproximadamente 0,1 mg. A hipófise então ficou imersa em 1 mL de *RNAlaterTM Stabilization Solution* (Thermo Fisher Scientific, Waltham, Massachusetts, EUA). A solução protege o RNA inativando qualquer *RNAase* (enzima responsável por degradar RNA) por tempo indeterminado em temperaturas de -20 °C. O material então foi rotulado e imerso em nitrogênio líquido para armazenamento e transporte.

O material foi transportado via terrestre de Canabrava do Norte (Mato Grosso) para Palmas (Tocantins) e depois para São Paulo (São Paulo) em kit térmico UN 1845 (Polar Store, São Bernardo do Campo, São Paulo), contendo gelo seco, seguindo toda a documentação necessária para transporte de material biológico (UN3373-B).

No momento em que as amostras chegaram às instalações do Centro de Biotecnologia (CEBIO), todos os criotubos foram armazenados em ultra freezer a -80 °C.

3.2. Extração de RNA da hipófise coletada de *Arapaima gigas*

O processamento da hipófise foi realizado com auxílio dos estudantes seguindo as orientações do kit *RNeasy mini* (Qiagen, Hilden, Nordrhein-Westfalen, Alemanha).

O criotubo, contendo a hipófise e a solução *RNAlaterTM*, foi descongelada em gelo. Após essa etapa, com auxílio de luvas criogênicas impermeáveis, a hipófise foi retirada e colocada em um cadinho de porcelana previamente resfriado, despejando quantidade suficiente de nitrogênio líquido para congelamento do tecido. Em seguida, com auxílio do homogeneizador de tecidos *Pyrex[®] Tissue Grinder – 7 mL, body O.D. × H 15mm × 117 mm* (Corning Inc., Nova Iorque, EUA), foi triturada e reduzida a pó.

Após redução total do tecido em questão, 600 µL de *Buffer RLT* (com 2-Mercaptoetanol) foi adicionado e homogeneizado com uma seringa equipada com agulha 20g para obtenção de uma solução viscosa. Essa amostra foi acondicionada em um microtubo de 1,5 mL e centrifugada na velocidade máxima por três minutos. O sobrenadante resultante foi transferido para um novo microtubo de 1,5 mL.

Homogeneizado com 700 µL de álcool 70% recém preparado, a solução foi transferida para a coluna *RNeasy spin column* alocada em um tubo de coleta, centrifugada por quinze segundos a 10000 × rpm. O sistema consiste em uma pequena coluna com um filtro que prende as moléculas de RNA até sua eluição.

O conteúdo centrifugado, no tubo coletor, foi desprezado e 700 µL de *buffer RWI* foram adicionados na coluna e realizada outra centrifugação por quinze segundos a 10000 × rpm. O conteúdo centrifugado mais uma vez foi desprezado. Duas centrifugações sucessivas a essa etapa com 500 µL de *buffer RPE* a 10000 × rpm por quinze segundos e, em seguida, dois minutos, foram realizadas para a lavagem do material. O conteúdo filtrado nessas etapas acabou sendo desprezado.

A coluna, após sucessivas etapas de lavagem, foi alocada em um novo microtubo de 1,5 mL e 50 µL de água livre de *RNase* foi adicionado no centro de sua membrana. A eluição ocorreu na centrifugação por um minuto a 10000 × rpm, sendo o filtrado reintroduzido na mesma coluna e eluído novamente para maior rendimento do processo.

3.3. Amplificação rápida das extremidades de cadeia (5'/3' RACE)

Nesta parte prática, utilizamos os kits comerciais 5' RACE System for Rapid Amplification of cDNA Ends, version 2.0 e 3' RACE System for Rapid Amplification of

cDNA Ends (Cat. nº 18374058 e 18373019, respectivamente, Invitrogen™, ThermoFisher Scientific, Waltham, MA, EUA).

O protocolo de amplificação para a região 5' e 3' do cDNA percorrem caminhos distintos, uma vez que, a extremidade 5' necessita o desenho de dois primers específicos (GSP) na orientação *antisense* e na extremidade 3' um primer específico *sense*. Os primers foram desenhados com base na sequência obtida *in silico* (tópico 3.5). Utilizamos o programa BioEdit (HALL, 1999) e o *OligoAnalyzer tool*, ferramenta gratuita disponível e fornecida pela empresa IDT⁷ para construção dos primers. Os primers específicos estão listados a seguir (Tabela 3) e foram adquiridos pela empresa Exxtend (Paulínia, SP, Brasil).

Tabela 3: Lista dos primers utilizados

Primers	Sequência (NT)	T _m (°C)	Finalidade
GSP 1 TSH (P1)	TGCTACAGTAACAGT	42.4	5' RACE
GSP 2 TSH (P1)	ATGTTGGTGTCCCGCGAATA	57.3	5' RACE
GSP TSH	AACACCACCATCTGCATGGG	59.3	3' RACE

Fonte: Autor da dissertação

Nesse sentido, a metodologia para 5' RACE exige que os *primers* antisense (3' → 5') estejam alinhados da seguinte forma:

- GSP1: Com uma distância maior de 300 pares de base (pb) da extremidade 5' mRNA. Deve conter uma baixa temperatura de melting (T_m: ~ 42 °C) e entre 16-20 pb.
- GSP2: Um *primer* em uma região já amplificada (*nested*) que irá anelar na extremidade 3' (com relação ao cDNA) do *primer* GSP1 ou em região anterior (*upstream*) ao *primer* GSP1.

O experimento consiste em quatro etapas: a primeira, com síntese da primeira fita de cDNA a partir do RNA (1-5µg) para um mRNA específico (nesse caso, utilizando o GSP1 relativo ao ag-βTSH) com ação da enzima *SuperScript™ II RT*, fornecida no kit informado acima; a segunda, com uma purificação do material de dNTPs (bases de nucleotídeos degenerados) e GSP1 não incorporados em coluna S.N.A.P., também fornecida no kit acima;

⁷ Disponível em: <https://www.idtdna.com/pages/tools/oligoanalyzer>. Acesso em: 14 de julho de 2022.

a terceira, com uma adição de uma cauda homopolimérica na região 3' e a última seguida da amplificação por reação em cadeia da polimerase (PCR) com o *primer* GSP2.

A metodologia utilizada para a região 3' toma vantagem a partir da cauda poli-A já presente em mRNA. O *primer* senso (5' → 3') sintetizado precisa apenas conter as seguintes orientações:

- GSP: *Primer* específico para seu produto de interesse contendo de 18-22 nucleotídeos.

A partir da conversão direta do RNA em cDNA com ação da enzima *SuperScript™ II* RT (utilizando, nesse caso, um *primer* adaptador, fornecido também pelo kit), apenas uma amplificação direta de PCR com o *primer* GSP amplifica até o término da cadeia.

Os resultados descritos são visualizados em gel de agarose (1%) descritos no subtópico a seguir.

3.3.1. Reação em cadeia da polimerase (PCR) e eletroforese em gel de agarose

Para aplicação da metodologia descrita abaixo, a reação em cadeia da polimerase (PCR) foi realizada com o sistema *Platinum™ SuperFi™ II DNA Polymerase* (12368010, Invitrogen™, ThermoFisher Scientific, Waltham, MA, EUA) e a agarose com o *Certified™ Molecular Biology Agarose* (Bio-Rad Laboratories, Hercules, CA, EUA).

O protocolo do PCR é descrito com a adição dos seguintes componentes:

- 10 µL de 2X *Platinum™ SuperFi™ II PCR Master Mix*
- 0,5 µM de *primer* na orientação sense
- 0,5 µM de *primer* na orientação antisense
- 5-100 ng de DNA ou 0,1-10 ng de plasmídeo
- q.s.p. (quantidade suficiente para) 20 µL de Água livre de nuclease

Em seguida, após homogeneização e centrifugação das amostras, são programados em um termociclador (*iCycler Thermal Cycler 2x48 PCR Machine*, Bio-Rad Laboratories, Hercules, CA, EUA) os seguintes parâmetros:

1. Desnaturação (1x)	98 °C30 seg
2. Amplificação (25-35x)	98° C.....5-10seg
	60 °C.....10 seg
	72°C.....10-30seg
3. Extensão (1x)	72°C.....5min
	4 °C..... ∞

Resumindo as etapas e os processos relacionados a temperatura, temos: 98 °C para desnaturação e abertura da fita dupla de DNA, 60 °C a de anelamento que está de acordo com o *primer* e 72 °C a extensão da nova cadeia de DNA. A temperatura de 4 °C nesse caso é apenas para conservar a amostra até sua utilização.

O gel de agarose (1%) é feito diluindo em um Erlenmeyer de vidro 0,5g de agarose em 50 mL de tampão TAE 1x (Tris-Acetato-EDTA), esquentando em micro-ondas por trinta segundos ou até a solução se apresentar transparente. Em seguida, quando não estiver mais saindo vapor da solução, adiciona-se 1 µL de brometo de etídio (1mg/mL) (E7637, Merck, Darmstadt, Alemanha) para visualização das bandas.

A análise é feita em uma cuba eletroforética com imersão do gel em tampão TAE 1x. As amostras obtidas por PCR são aplicadas no gel previamente pronto da seguinte maneira: 5 µL da amostra + 1 µL de 6X TriTrack DNA *Loading Dye* (ThermoFisher Scientific, Waltham, MA, EUA). O marcador de peso molecular utilizado nessa metodologia foi GeneRuler 100 bp DNA Ladder (SM0241, ThermoFisher Scientific, Waltham, MA, EUA). Utilizando uma fonte elétrica, a corrida eletroforética é feita por uma hora com 80 V de corrente.

Para visualização do gel utiliza-se luz ultravioleta. O resultado foi foto documentado com o aparelho *Alliance Q9 Advanced* (UVItec Ltd., Cambridge, GB).

3.4. Sequenciamento de Sanger

O sequenciamento das bandas visualizadas em gel de agarose descritos no tópico anterior foi realizado no Centro de Pesquisa e Diagnóstico Molecular de Doenças Genéticas, do departamento de biofísica na Universidade Federal de São Paulo (UNIFESP) – Escola Paulista de Medicina (São Paulo, SP). São utilizados 30-200 ng de produto de PCR com os primers utilizados na reação na concentração de 10 µM.

O sistema de sequenciamento utilizado pelo centro é do tipo Sanger feito no equipamento *3500 xL Genetic Analyzer* (ThermoFisher Scientific, Waltham, MA, EUA).

3.5. Análise filogenética por máxima verossimilhança (ML)

A árvore filogenética da sequência correspondente ao ag- β TSH foi construída através do método máxima verossimilhança (ML) com árvore. A análise consiste é uma estimativa de uma distribuição que busca maximizar a maior probabilidade dentro do modelo escolhido entre várias distribuições. Foram utilizadas 38 sequências de aminoácidos com diferentes espécies de peixes (incluindo a do *Arapaima gigas*) depositado em banco de dados. Apenas a região codificante desses aminoácidos foi utilizada. Todo o processo foi realizado através do software de livre acesso MEGA X (KUMAR *et al.*, 2018).

O primeiro passo consiste em um alinhamento de todas as sequências, utilizando o Clustal Ω . Em seguida, o próprio software determina qual o melhor modelo de substituição a ser usado. Segundo as informações obtidas (descritas na tabela do Apêndice A) e seguindo os valores para BIC (Critério de informação Bayesiana) e AICc (Critério de informação Akaike), o melhor modelo de substituição e utilizado nessa análise foi o Jones-Taylor-Thornton com frequência de aminoácidos (JTT+F; BIC: 10016,74683; AICc: 9402,175522). A metodologia é um aperfeiçoamento de automação derivada da máxima parcimônia (MP) assumindo que as sequências alinhadas sejam em ordem de identidade > 85% (WHELAN, GOLDMAN, 2001; JONES, TAYLOR & THORNTON, 1992).

A árvore foi construída com 500 replicadas (*bootstraps*) para cada táxon incluso na árvore (FELSENSTEIN, 1985). Por fim, a construção foi realizada com o algoritmo *Subtree-Pruning-Regrafting* (SPR), que desmembra uma parte da melhor árvore construída e esta é alocada em outras árvores já construídas. O algoritmo então busca em todas as subárvores até o resultado não superar a melhor verossimilhança das topologias já construídas. Seus parâmetros de busca foram “*search_level=1*” e as árvores iniciais foram obtidas através de 10 replicadas (NEI; KUMAR, 2000).

3.6. Obtenção da região codificante *in silico*

Abordado na dissertação ([ver tópico 2.3.2.2](#)) como uma estratégia das técnicas de obtenção da região codificante, a identificação *in silico* complementa os resultados obtidos no processamento da hipófise e utilização dos kits para amplificação do ag- β TSH. Dessa forma, podemos ter uma segurança adicional no processamento das informações.

Assim descrito no título do subtópico, o trabalho foi realizado através de softwares de livre acesso para sistema operacional Windows e dados publicados acerca do objeto de estudo, que facilitaram e foram fundamentais no desenvolvimento dos dados. Dividido em quatro etapas, abordamos e concluímos as metodologias empregadas na obtenção do gene e sua construção dentro do vetor plasmidial.

3.6.1. Alinhamento de sequências cDNA β TSH de espécies contra o genoma do *Arapaima gigas*

O genoma completo do *Arapaima gigas* (em formato de *scaffolds*) foi depositado no banco de dados do NCBI (referência de acesso: GCA_900497675.1) em maio de 2018. O próprio sistema desenvolvido pela companhia nos permite realizar o alinhamento de sequências de nucleotídeos dentro do genoma (BLASTn). Utilizamos quatro sequências de espécies diferentes: *Anguilla anguilla* (XM_035387135.1), *Danio rerio* (NM_181494), *Anguilla japonica* (AB175833) e *Scleropages formosus* (XM_018725541).

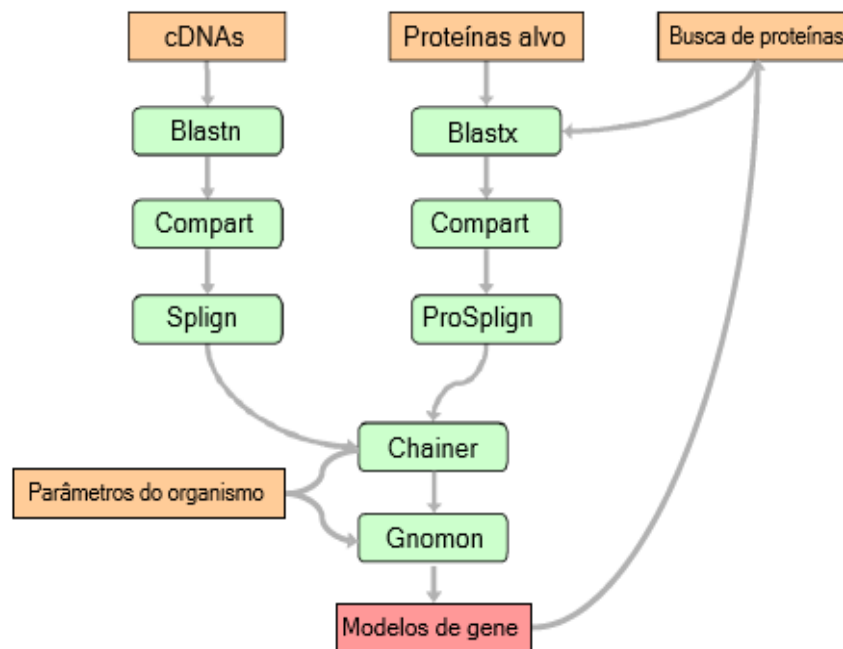
3.6.2. Identificação da região codificante pela metodologia GNOMON

3.6.2.1. Ferramenta de predição gênica GNOMON.

Como já mencionado no segundo tópico deste capítulo, o genoma do Pirarucu, com diferentes metodologias de sequenciamento (que englobam a *Next-Generation Sequencing - NGS*), permitiu que houvesse otimização do processo da identificação de gene, alvo deste trabalho.

Entretanto, não só o acesso a base de dados, utilização de programas e predição do modelo foram suficientes para sua conclusão. Estes estudos, seguiram um padrão, aplicado nesse projeto. A ferramenta *Gnomon*, desenvolvida para o *National Center for Biotechnology Information* (NCBI), em 2010, contém um fluxograma (Figura 10) que delimita as etapas para obtenção de um gene *in silico* (SOUVOROV *et al.*, 2010). Cada passo será explicado a fim da maior compreensão do processo e concluímos o capítulo com considerações tomadas acerca do processo.

Figura 10: Fluxograma das etapas do método de predição Gnomon.



Fonte: Traduzido de Souvorov *et al.*, 2010.

3.6.2.2. cDNA e alinhamento múltiplo BLASTn.

O fluxograma que acompanhamos no processo é o caminho descrito à esquerda da **figura 10**. Uma vez que a base de dados consultada é genômica - podendo também ser proteômica -, os algoritmos citados a seguir não apresentam tantas diferenças ao serem discutidos.

Entende-se por fita complementar do DNA (cDNA) o gene alvo transcrito em RNA mensageiro (mRNA) e convertido novamente em DNA. Essa conversão acontece na complementaridade das bases nitrogenadas U-A, G-C-, A-T e C-G (ressaltando a exclusividade da uracila - U - em RNA) com ação da enzima transcriptase-reversa. A importância dessa etapa é identificar genes, convertidos em cDNA, de espécies filogeneticamente próximas ao *Arapaima gigas*, ao passo de prosseguirmos para a próxima etapa do processo: *Basic Local Alignment Search Tool-nucleotide* (Blast).

O Blast é um algoritmo desenvolvido e hospedado no NCBI, de livre acesso e interface própria, que busca, a partir de uma sequência (nucleotídeos ou aminoácidos), similaridades em outras já depositadas em seu banco de dados. Também é permitido ao

usuário que é possível identificar sequências de cDNA em genomas de diversas espécies (MCGINNIS; MADDEN, 2004). Nesse caso, em genes que são expressados em apenas uma região cromossomal facilitam os estudos de predição gênica, uma vez que a localização será exclusiva e todo o trabalho posterior será realizado a partir desse resultado.

3.6.2.3. **Compart e Splign**

Seguindo o texto base proposto para metodologia GNOMON e as próximas etapas descritas no fluxograma, o Compart e Splign apresentam-se como ferramentas de análise do resultado obtido no BLAST.

O Compart tem como função analisar os *hits* do BLAST, na finalidade de localizar duplicação dos genes. Dessa forma, aproxima as sequências alvos do genoma, caso seja necessário. No caso desta dissertação, essa etapa não foi necessária, uma vez que o gene codificante do ag- β TSH não é expresso em mais de uma região.

Na próxima etapa, Splign/ProSplign, há a construção de alinhamentos segmentados. Usualmente é localizado no genoma um gene codificante para uma proteína com regiões interpostas, chamadas de íntrons. Essas regiões (com um número não definido de nucleotídeos) não são transcritas e traduzidas em proteínas, sofrendo um processo chamado de *splicing*, onde acabam sendo removidas. Esses segmentos no Splign/ProSplign podem ser juntados e assim construído a região codificante do gene.

3.6.2.4. **Chainer e Gnomon**

As últimas etapas do fluxograma estão relacionadas diretamente com parâmetros relacionados ao organismo, para criar uma anotação final do gene predito. Entretanto, essas etapas são reservadas e direcionadas para construções longas. As construções de regiões que apresentam *splicing* alternativo e possíveis variantes utilizam a linguagem de programação em C++ disponibilizada gratuitamente pelo NCBI.

Dessa forma, concluímos a fundamentação teórica por trás da predição *in silico* com base na sequência genômica. A abordagem facilitará a compreensão dos resultados obtidos seguindo o mesmo fluxograma e raciocínio. O próximo tópico deste capítulo será descrever o fundamento teórico por trás da bioquímica estrutural de proteínas.

Exportando o *scaffold540* do genoma (GCA_900497675.1) em arquivo FASTA que, contém a sequência de nucleotídeos para ag- β TSH, manipulamos o material genético com

auxílio dos programas: *Unipro UGENE*, *Splign*, *ProSplign* e *BioEdit*. Utilizamos para referência e norteamento dos estudos a sequência obtida de *Scleropages formosus* (XM_018725541), espécie filogeneticamente mais próxima do *Arapaima gigas*.

3.7. Construção de modelo tridimensional do agTSH

O modelo tridimensional do heterodímero foi construído com auxílio do programa AlphaFold e sua implementação pelo ColabFold (MIRDITA *et al.*, 2022), ao qual seu código pode ser acessado livremente, uma vez que utiliza uma máquina virtual na síntese dos modelos. Optamos por utilizar as sequências de aminoácidos sem o peptídeo sinal, ag- α TSH:

"NFNVGCEECKLKENKYFSRLGAPIFQCMGCCFSRAYPTPLRSKKTMLVPKNITSE
ATCCVAKEVKRLITLNNVRLNHTDCHCNTCYHKS" e ag- β TSH:
"FCALTDYTLTYVEKAGCDFCVAINTTICMGFCYSRDTNIVGLAGKRLLLQRTCTYR
SVEYHTTALPGCHQHADPFFSYPVALDCYCSTCDTGSHDCTHKGDDNSAQCAK
PLLHIYPYPGQSNHI", avaliado pelo software SignalP - 6.0 (TEUFEL *et al.*, 2022), a fim de obtermos um resultado mais próximo do natural. Os parâmetros que selecionamos no processo foram descritos no Apêndice B desta dissertação, em linguagem de programação Python.

Os resultados são apresentados na página do programa e automaticamente gera um arquivo (.zip) com os cinco modelos tridimensionais (.pdb).

3.8. Análise dos modelos tridimensionais e sequência de aminoácidos do agTSH

Para análise do melhor modelo tridimensional (ranqueado anteriormente por pontuação pLDDT), o software PyMOL 2.0 e Maestro 2022.2 auxiliaram a visualização das estruturas e disposição espacial dos aminoácidos. Dessa forma, conseguimos ter dimensão do tamanho estereoquímico da proteína. Com o software PDBsum (*Pictorial database of 3D structures*), conseguimos estudar propriedades físico-químicas e estruturais, como Ramachandran, carga, topologia e intervalos em sua estrutura, o que é de extrema importância no estudo da ligação com a subunidade alfa e também, na interação solvente/receptor.

3.9. Construção dos complexos

O domínio extracelular do receptor do TSH de *Scleropages formosus* (sfTSHR, NCBI XM_018732307.1) foi construído com a mesma abordagem apontada em 3.7, e o

melhor modelo foi utilizado para construir o docking envolvendo o agTSH. A construção dos complexos agTSH: sfTSHR e agTSH: hTSHR foram feitos por alinhamento estrutural da estrutura cristalizada do hormônio folículo estimulante humano ligado ao seu receptor (hFSH:hFSHR, PDB: 4AY9). Um sistema de controle do hTSH:hTSHR foi feito de forma semelhante, com a estrutura da subunidade β do hTSH foi construída com AlphaFold2 e estruturas cristalizadas foram recuperadas tanto da subunidade α do hTSH como do hTSHR (PDB: 4AY9 e 3G09, respectivamente).

No final, obtemos o modelo que possui a menor energia livre de ligação no complexo. Dessa forma, podemos prosseguir com a dinâmica molecular.

3.10. Cálculo de energia livre de ligação (MM/GBSA) entre agTSH + sfTSHR

As dinâmicas foram realizadas com os complexos agTSH: hTSHR, agTSH:sfTSHR e hTSH:hTSHR sendo protonados no pH=7.4 em servidor (H++, <http://newbiophysics.cs.vt.edu/H++/>). Então, os complexos foram inseridos em caixas de geometria octaédrica ($\alpha=90^\circ$, $\beta=104.90^\circ$ e $\gamma=90^\circ$) preenchidas com moléculas de água TIP3P e contra-íons para suficiente neutralização da carga (*tleap*, Amber20). Todos os complexos tiveram suas energias minimizadas, com aquecimento gradiente de 10-300K NVT (N: números de partículas, V: volume do sistema e T: temperatura, constante) e equilíbrio de 1 bar com 300K NPT (N: números de partículas, P: pressão do sistema e T: temperatura, constante). A dinâmica molecular ocorre em campo de força ff14SB em trajetória de 1 μ s.

3.11. Avaliação da força de ligação por *umbrella sampling*

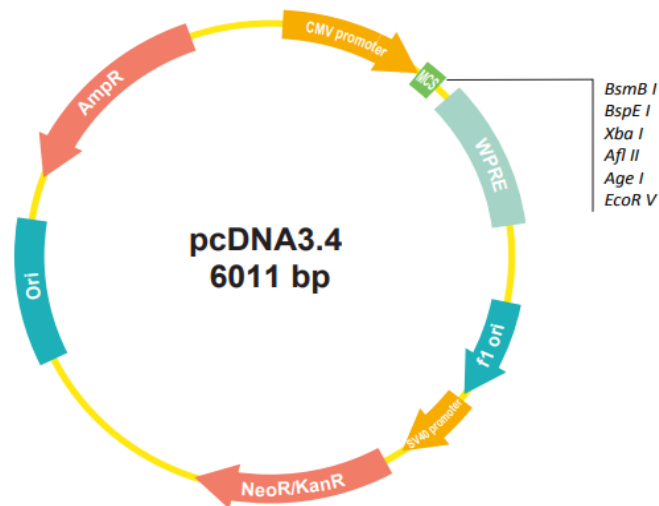
Os resultados obtidos pela dinâmica molecular foram cuidadosamente orientados para que o TSH seja distanciado do receptor no eixo z sem maior conflito estereoquímico. A fim de preservar os eixos rotacionais e translacionais do receptor e do hormônio, três partículas fictícias (D1, D2 e D3) foram adicionadas na caixa de solvatação de maneira não colinear e mantidas fixas de acordo com as restrições harmônicas. A distância entre D1 e o $C\alpha$ dos TSHs foram restritas ao longo do eixo z a cada 1 Å com constante harmônica de 10 kcal/mol/Å².

O método de análise de histograma *Weighted histogram analysis* (WHAM) foi empregado para obtenção dos potenciais de força médio (PMFs) ao longo do eixo selecionado. Maiores detalhes do protocolo são referidos no trabalho mencionado (GISMENE *et al.*, 2022).

3.12. Transformação de bactérias *Escherichia coli* DH5 α com plasmídeo pcDNA 3.4 + ag- β TSH e sua purificação

Uma cepa de bactéria *Escherichia coli* DH5 α quimicamente competente, previamente armazenada no laboratório à - 80 °C, foi transformada com 200 ng de plasmídeo pcDNA 3.4 + ag- β TSH (Figura 11). A transformação é feita pelo método de choque térmico, no qual consiste na transferência do DNA plasmidial através da abertura rápida dos poros da membrana celular.

Figura 11 - Mapa do vetor pcDNA3.4 com suas respectivas regiões que conferem características ao plasmídeo.



Em ordem alfabética: **AmpR**: Resistência ao antibiótico ampicilina; **CMV promoter**: promotor de citomegalovírus; **F1 ori**: origem de replicação de fago; **MCS**: Sítio múltiplo de clonagem - região onde está inserido o gene ag- β TSH e as enzimas de restrição que são utilizadas para clonagem; **NeoR/KanR**: Resistência aos antibióticos neomicina e kanamicina; **Ori**: origem de replicação; **SV40 promoter**: região promotora de vírus vacuolante símio 40; **WRPE**: Elemento regulatório pós-transcricional de vírus de hepatite WoodChuck

Fonte: GenScript Biotech (Piscataway, NJ, EUA).

O protocolo começa com a adição do plasmídeo purificado em 200 μ L de volume com bactérias. A amostra então fica 20 minutos em gelo para depois sofrer o choque térmico, saindo do gelo para 37 °C por dois minutos. Em seguida, são adicionados 600 μ L de meio luria-bertani (LB) autoclavado sem antibiótico e, por uma hora, a amostra fica sob agitação constante à 37 °C. Gradativamente, são incorporados os plasmídeos e recuperada a membrana celular.

As bactérias então transformadas são semeadas por estriamento com alça de drigalski, adicionado 100 μ L da reação em placa de petri com meio LB + ampicilina 0,1% e

analisadas após o período de incubação *overnight* (16-18h) a 37 °C. Com adição de glicerol, as bactérias transformadas remanescentes são armazenadas a -80 °C.

No dia seguinte, as colônias que cresceram são inoculadas em 100 mL de meio líquido LB autoclavado, novamente a 37 °C, porém, com agitação constante de 150 rpm. A placa então é descartada em lixo branco.

O terceiro dia consiste na extração do plasmídeo das bactérias transformadas que foram inoculadas no meio líquido LB. Seguimos o protocolo de purificação do produto comercial *NucleoBond Xtra Midi kit transfection-grade plasmid DNA* (Macherey-Nagel, Düren, Nordrhein-Westfalen, Alemanha) para obtermos o plasmídeo em grande quantidade, suficiente para prosseguirmos com a transfecção transiente. O resultado final foi quantificado em espectrofotômetro e o plasmídeo armazenado em freezer a -20 °C.

3.13. Cultura de células de mamífero e transfecção transiente com lipídeo catiônico

A cultura de células de mamíferos foi realizada em suspensão para células de rim de embrião humano (HEK293F⁸). Em um erlenmeyer de 125 mL com tampa de filtro 0,22 µM, as células são descongeladas em 30 mL de meio de cultura específico, sem antibiótico, e incubadas a 37 °C, 8 % CO₂ e 150 rpm até sua confluência para transfecção. As células HEK293 exigem uma densidade para transfecção de 3,0 x 10⁶ células/mL e ≥ 95 % de viabilidade. Duas produções foram realizadas: a primeira para rastreabilidade e a segunda para purificação da amostra. Durante o crescimento das células, a densidade (células/mL) e viabilidade (em porcentagem) são determinadas por câmara de Neubauer, utilizando o método de distinção por azul de tripan.

Em uma primeira análise, utilizamos metade do valor final para transfecção dos plasmídeos (alfa e beta). Todos os reagentes foram adquiridos pela empresa *Thermo Fisher Scientific* e o protocolo, otimizado, para maiores rendimentos da produção. Após chegar nas condições desejadas para a transfecção, o protocolo pode acabar sendo alterado.

⁸ O sufixo “F” significa “*Freestyle*”, a primeira linhagem adaptada em suspensão comercializada pela empresa *ThermoFisher Scientific*.

3.13.1. Transfecção transiente de células HEK293 - ensaio inicial

As células na densidade final de $3,0 \times 10^6$ células/mL com ≥ 95 % de viabilidade foram ressuspendidas em 25,5 mL de meio de cultura pré-aquecido em um novo erlenmeyer. Então, 30 μ g dos plasmídeos foram diluídos em um falcon (A) de 15 mL contendo 1,5 mL de Opti-MEM e, cuidadosamente, homogeneizados. Em outro falcon (B) contendo 1,5 mL de Opti-MEM, 81 μ L de lipídio *Expifectamine293* foram adicionados e, após homogeneização, incubado a temperatura ambiente por cinco minutos. Depois, o conteúdo dos falcons foi misturado (B→A), homogeneizado e incubado a temperatura ambiente por 20 minutos. Esse complexo plasmídeo + lipídio foi adicionado aos 25,5 mL do meio de cultura contendo as células e incubado nas condições de 32 °C, 8% CO₂ e 150 rpm. Após o período de overnight (17-22h), os *enhancers* 1 e 2 são adicionados, sendo 150 μ L e 1,5 mL respectivamente, totalizando 30 mL finais de meio de cultura transfectado. Uma amostra de 1 mL do meio condicionado foi coletada nos dias 03, 05 e 07 após a adição dos *enhancers* e, armazenados a -80 °C. A cultura se manteve durante sete dias para coleta do meio e as células descartadas, posteriormente.

3.14. Análise qualitativa da produção em SDS-PAGE

A eletroforese em gel de poliacrilamida (SDS-PAGE) é uma metodologia específica para avaliação qualitativa de proteínas. Com princípios similares a metodologia da eletroforese em gel de agarose, as principais diferenças observadas entre as técnicas são:

- A orientação do sistema é vertical, com duas fases de preparação (empilhamento, com um pH mais negativo; resolução, com um pH mais básico). Avaliando o peso molecular da proteína analisada, a malha de porcentagem do gel pode ser ajustada para melhor visualização (8-15%). Além disso, se faz necessário o uso de tampão específico de tris-glicina gelado para condução de eletricidade.
- A amostra precisa ser diluída e fervida por três minutos à 85 °C em um tampão contendo dodecil-sulfato de sódio (SDS), Tris e Glicina. Usualmente esse tampão está concentrado 4x. Nesse caso, optamos por preparar as amostras através da precipitação por ácido tricloroacético (TCA) através do protocolo já definido *Novagen pET System Manual, 10th edition, pag. 44*
- As etapas de visualização do gel seguem três etapas importantes de fixação, com metanol; coloração, com *coomassie brilliant blue g 250* e, por último, revelação, com ácido acético.

- Caso seja necessário, uma etapa de coloração com nitrato de prata pode ser realizada. O protocolo utilizado está descrito em *Multiphor II Electrophoresis system – User Manual – Pharmacia LKB Biotechnology – Sweden, 1990*. Páginas 184 e 187.

As amostras coletadas no dia 3,5 e 7 após a transfecção foram preparadas seguindo as diluições apontadas acima e aplicadas em um gel de 15%. O volume de amostra aplicado foi 10 μ L, sendo as condições de corrida com voltagem e tempo de 180 V e 45 min, respectivamente. O tampão de Corrida utilizado foi o Tris-Glicina 1x.

3.15. Análise do meio condicionado em RP-HPLC

O meio condicionado resultante do quinto dia ($V_0 = 30$ mL) após a transfecção foi concentrado e filtrado ($V_f = 5$ mL) em tubos 3.0 kDa MWCO Amicon® Ultra Centrifugal Filters (Merck, Rahway, NJ, EUA). Após essa etapa, o concentrado foi aplicado em sistema de cromatografia líquida de alta performance por fase reversa (RP-HPLC).

Detalhando as propriedades do ensaio, temos: I) fase estacionária constituída por uma pré-coluna de Sílica acoplada a uma coluna C4 Grace Vydac® (W.R. Grace Company, Columbia, MR, EUA) a 20 °C e II) fase móvel, líquida, com dois tampões A (Fosfato de sódio 50mM pH 7,0) e B (Acetonitrila PA). Com fluxo de 0,5 mL/min e um *loop* (momento em que a amostra é misturada com os tampões) de 2000 μ L, os detalhes da corrida foi definido em cinco etapas: 1) 15% tampão B de 0 à 7 min; 2) gradiente de 15 a 60% do tampão B de 7 à 45 min; 3) 60% do tampão B de 45 à 50 min; 4) gradiente de 60 a 15% do tampão B de 50 a 55 min; 5) 15% do tampão B até 60 min. Foi programado para que o equipamento realizasse coletas de 0,5 mL/min do minuto 30 até o 50. Todo ensaio foi realizado no equipamento LC-20A (Shimadzu corp, Kyoto, KY, JP).

Para o ensaio, utilizamos o controle positivo (hormônio luteinizante humano, hLH, NIH) na concentração de 5 μ g/5 μ L, controle negativo, aplicando 5000 μ L da amostra concentrada e 4000 μ L do quinto dia concentrado após a transfecção transiente em células HEK293. Os cromatogramas resultantes das amostras foram gerados em absorbância 220 nm. As coletas foram armazenadas em geladeira a 4 °C *overnight* (16-18h).

3.16. Análise quantitativa dos *pools* em HPSEC

As coletas realizadas nos picos resultantes dos cromatogramas em RP-HPLC ($V_0 = 2$ mL) foram concentradas e dialisadas ($V_f = 500$ μ L) com tampão fosfato de sódio 20mM

pH 7,0 em tubos 3.0 kDa MWCO Amicon® Ultra Centrifugal Filters (Merck, Rahway, NJ, EUA). O concentrado foi aplicado em sistema de cromatografia líquida de alta performance exclusão molecular (HPSEC).

Detalhando as propriedades do ensaio, similar a metodologia de RP-HPLC, o sistema possui a fase estacionária com apenas uma coluna acoplada TSKgel® 2000 (Tosoh, Tóquio, JP) em temperatura ambiente e, sua fase móvel, com um tampão fosfato de sódio 20mM pH 7,0. Com fluxo de 1,0 mL/min e *loop* de 500 µL, o programa detalhado orienta para coletar 0,5 mL/min a partir do minuto 10 ao 20. O tempo total de corrida foi de 40 minutos. Todo ensaio foi realizado no equipamento LC-20A (Shimadzu corp., Kyoto, KY, JP). Os cromatogramas resultantes das amostras foram gerados em absorvância 220 nm. Para fins de referência, um padrão de hormônio pituitário humano foi utilizado na análise

4. Resultados

A partir do estudo publicado por Vialle e colaboradores (2018), mencionado durante a dissertação, sobre a publicação do genoma de *Arapaima gigas*, foi possível identificar e manipular os dados, através de softwares disponibilizados gratuitamente, na região codificante do gene da subunidade beta do hormônio estimulante da tireoide de *Arapaima gigas* (ag-TSH). Enfatizamos que o TSH é necessariamente formado pela interação proteica não covalente entre as subunidades alfa e beta, exigindo no primeiro momento, a metodologia em torno da subunidade beta, uma vez que nunca foi estudada para essa espécie. Dessa forma, o processo complementa os estudos realizados *in vitro*, sustentando e corroborando os resultados previamente obtidos. Vale pontuar que os modelos propostos são previsões que tomam como base outros organismos e dados já publicados; no entanto, se tratando de um ambiente acadêmico, tais informações são valiosas e inéditas.

O tópico que será descrito a seguir foi dividido em etapas com figuras para melhor entendimento do processo. Os programas utilizados durante esse processo foram: Basic Local Alignment Tool - Nucleotide (BlastN) e Protein (BlastP) (CHEN *et al.*, 2015); UGENE Integrated Bioinformatic Tools (OKONECHNIKOV *et al.*, 2012); BioEdit – Nucleics (HALL., 2011); PREDDIMER (Polyansky *et al.*, 2014); Splign e SplignPro (KAPUSTIN *et al.*, 2008); PROCHECK (Laskowski *et al.*, 1993); Alphafold (Jumper *et al.*, 2021); SignalP 6.0 (TEUFEL *et al.*, 2022); ClusPro 2.0 (COMEAU *et al.*, 2004); InterPro (BLUM *et al.*, 2020); Geneious basic (KEARSE *et al.*, 2012); PyMOL (DELANO., 2002); Ramachandran plot – Zlab server (ANDERSON *et al.*, 2005); ClustalΩ (SIEVERS *et al.*, 2011); PDB2PQR – APBS server (UNNI *et al.*, 2011); MEGA11 (TAMURA, STECHER, KUMAR, 2021); Expasy (DUVAUD *et al.*, 2021).

4.1. Estudos *in silico* para ag-βTSH e agTSH: identificação genômica

Como previamente introduzido, tomamos como base o trabalho publicado que auxiliou o estudo a ser conduzido com maior precisão. O genoma do *Arapaima gigas* publicado (Código de acesso NCBI: GCA_900497675.1) possui 667.351.951 nucleotídeos e 60.052 *scaffolds*. São informações contidas dentro do DNA que até então, não haviam sido utilizadas como uma estratégia no trabalho. Porém, a árvore filogenética das espécies resultantes desse sequenciamento completo em contraste com o *Arapaima gigas*, revelou que o *Scleropages formosus*, possui um grau de parentesco mais próximo do que as outras

ali existentes. Essa informação foi valiosa no sentido de precisão: utilizando do algoritmo *BLAST* e a sequência de β TSH do *Scleropages formosus* (Código de acesso NCBI: XM_018725541), conseguimos localizar em que posição do genoma do *Arapaima gigas* a sequência codificante do ag- β TSH se encontra. Ora, se as espécies são filogeneticamente próximas, o alinhamento resulta em uma posição única no gene. Visualizamos esse resultado que determinou o *scaffold540* como a região de expressão para β TSH com 85% de identidade cobrindo 127 nucleotídeos

4.1.1. Confirmação do scaffold com outras sequências correspondente ao gene

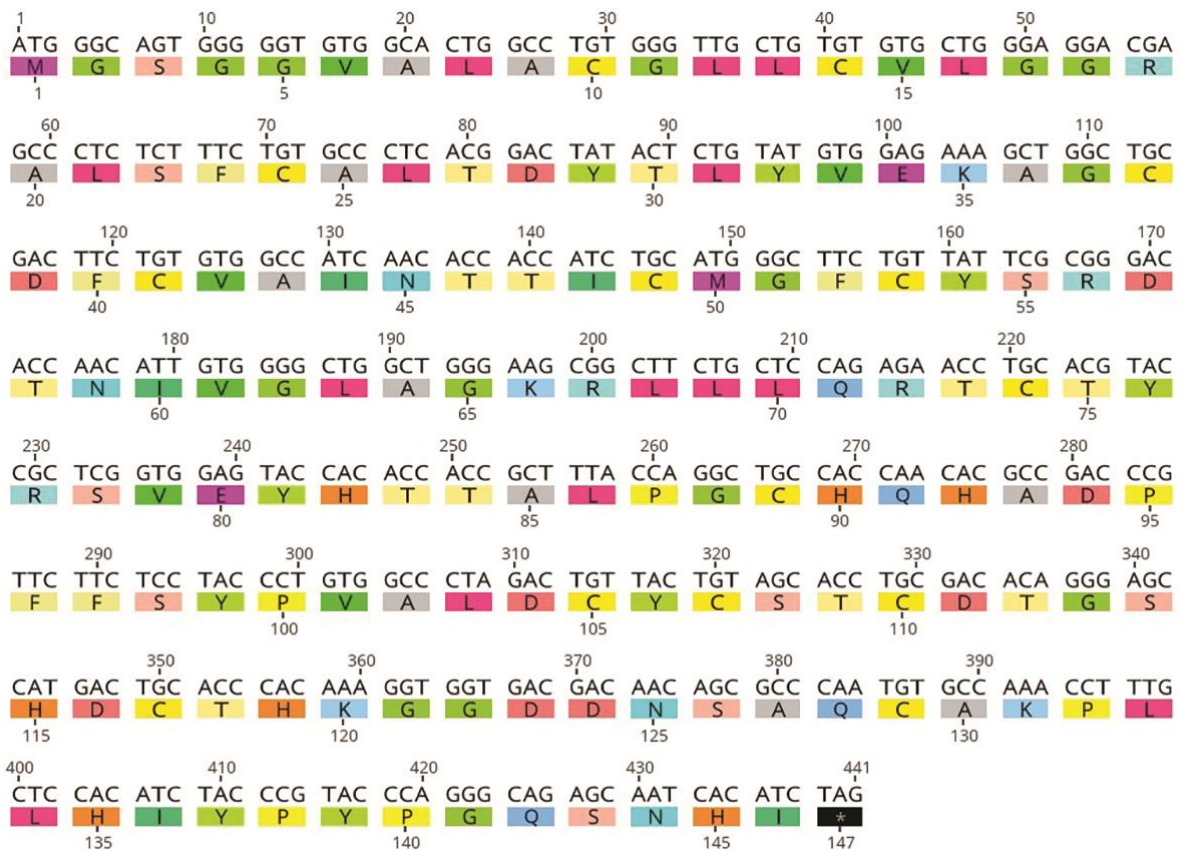
Para o argumento de que o *scaffold540* é único e exclusivo para sequência de nucleotídeos que expressa o β TSH, foram selecionadas três sequências da proteína com espécies filogeneticamente próximas ao *Arapaimas gigas*: *Anguilla anguilla* – (NCBI: XM_035387135.1), com 88% de identidade; *Danio rerio* (NCBI: NM_181494), com 95% de identidade e *Anguilla japonica* (NCBI: AB175833), com 88% de identidade. Todas as sequências alinharam em regiões que estão dentro do *scaffold540*. Dessa forma, exportamos a sequência completa contida dentro do *scaffold* em formato FASTA (UFQX01000656.1).

4.1.2. Identificação da região codificante contrastando com *S. formosus* (Splign & SplignPro)

A metodologia de predição de proteínas *in silico* Gnomon (**ver tópico 3.6.2.**) nos permite alinhar a sequência de *Scleropages formosus* (NCBI: XM_018725541) com o *scaffold540* (previamente exportado em arquivo em formato FASTA para definirmos qual a porcentagem que obtemos para região codificante. Quando feito o alinhamento, o resultado obteve 100% de cobertura e permitiu identificar os códons que são traduzidos na proteína desde a metionina inicial (ATG) até o códon terminal (que nessa situação, está descrito como TAG).

O resultado, desse processo, nos gerou 441 nucleotídeos traduzidos em 146 aminoácidos, dentre os quais, 124 constituem o peptídeo maduro (Figura 12).

Figura 12: Sequência de nucleotídeos e aminoácidos correspondente à região codificante da subunidade beta do TSH de *Arapaima gigas*.



Fonte: Autor da dissertação

Através do software mencionado na metodologia (ver tópico 3.4), está representado o peptídeo sinal dos aminoácidos (M1-S22), início da cadeia madura (F23) e término (I146). O sítio de N-glicosilação (N45) está de acordo com o descrito na literatura. Para esses argumentos serem sustentados dentro o modelo previsto (como o número de cisteínas que formam as pontes dissulfeto, por exemplo), propomos nos subtópicos deste item alguns resultados para corroborar a teoria.

4.1.3. Evidência para proposta *in silico*: porcentagem de identidade dos aminoácidos

Com a própria ferramenta utilizada para identificação do scaffold540, submetemos a sequência de aminoácidos a ferramenta BlastP, exclusivo para proteínas:

MGSGGVALACGLLCVLGGRALSFCALTDYTLTYVEKAGCDFCVAINTTICMGFCYS
RDTNIVGLAGKRLLLQRTCTYRSVEYHTTALPGCHQHADPFFSYVALDCYCYCSTC
DTGSHDCTHKGDDNSAQCAKPLLHIYPYPGQSNHI

Obtivemos um resultado de múltiplos alinhamentos que foram utilizados para construção da árvore filogenética. Para facilitar o entendimento, a **tabela 4** discrimina as espécies com maior similaridade, destacadas em negrito e em ordem decrescente para β TSH. Todas as sequências são espécies de peixe com a exceção das últimas correspondente a *Homo sapiens* e *Mus musculus*.

4.1.4. Evidência para proposta *in silico*: assinatura da subunidade beta

A assinatura é um padrão que existe dentro dos hormônios glicoproteicos na subunidade beta, que possuem 100-140 aminoácidos com doze cisteínas conservadas, todas envolvidas em pontes dissulfeto e, as primeiras cisteínas (N-terminal) estão envolvidas na ligação com a subunidade α e cinco restantes são localizadas na região (C-terminal) (PIERCE; PARSONS, 1981; HARTREE; RENWICK, 1992; LAPTHORN *et al*, 1994).

Seguindo essa hipótese, o modelo proposto atende as duas maneiras: a proteína madura (desprezando o peptídeo sinal de 21 aminoácidos) contém 126 aminoácidos e doze cisteínas conservadas, envolvidas nas pontes dissulfeto.

Tabela 4: Identidade (em porcentagem) da cadeia β de *Arapaima gigas* com 38 espécie

Nome comum	Nome científico	Similaridade	Número GenBank
Pirarucu	<i>Arapaima gigas</i>	-	OP125854
Aruanã-dourado	<i>Scleropages formosus</i>	81.60	XP_018581057
Tarpão indo-pacífico	<i>Megalops cyprinoides</i>	70.40	XP_036387986.1
<i>Brown grouse</i> ⁹	<i>Paramormyrops kingsleyae</i>	68.80	XP_023693336.1
Sável	<i>Alosa alosa</i>	64.52	XP_048111532.1
Arenque denticulado	<i>Denticeps clupeioides</i>	64.46	XP_028854856.1
Enguia japonesa	<i>Anguilla japonica</i>	64.29	Q7ZZV4.1
Sábalo americano	<i>Alosa sapidissima</i>	63.71	XP_041945432.1
Enguia europeia	<i>Anguilla anguilla</i>	63.49	XP_035243023.1
Salmão do atlântico	<i>Salmo salar</i>	63.48	NP_001117000.1
Arenque do atlântico	<i>Clupea harengus</i>	62.90	XP_041945431.1
Truta arco-íris	<i>Oncorhynchus mykiss</i>	62.90	XP_021425667.1
Garoupa gigante	<i>Epinephelus lanceolatus</i>	59.68	XP_033482843.1
Peixe maior	<i>Syngnathus acus</i>	58.82	XP_037107826.1
Enguia elétrica	<i>Electrophorus electricus</i>	58.54	XP_026875009.1
<i>Common coral trout</i>	<i>Plectropomus leopardus</i>	58.06	XP_042353710.1
Peixe-leite	<i>Chanos chanos</i>	56.45	XP_030633387.1
<i>Rhinoceros horn fish</i> ¹⁰	<i>Sinocyclocheilus rhinoceros</i>	56.45	XP_016395895.1
<i>Anshui clove fish</i> ¹¹	<i>Sinocyclocheilus anshuiensis</i>	56.45	XP_016312618.1
Pungitius	<i>Pungitius pungitius</i>	56.10	XP_037346720.1
Carpa	<i>Cyprinus carpio</i>	55.65	XP_018973503.2
Barbilho da linha dourada	<i>Sinocyclocheilus grahami</i>	55.65	XP_016143109.1
Tambaqui	<i>Colossoma macropomum</i>	55.60	XP_036426973.1
Peixe-arqueiro	<i>Toxotes jaculatrix</i>	55.60	XP_040889707.1
<i>Fathead minnow</i>	<i>Pimephales promelas</i>	54.84	XP_039548072.1
Peixe-dourado	<i>Carassius auratus</i>	54.47	XP_026111649.1
Bagre-americano	<i>Ictalurus punctatus</i>	54.10	XP_017335429.1
Bagre listrado Raphael	<i>Pangasianodon hypophthalmus</i>	54.10	XP_026777430.1
Robalo riscado	<i>Morone saxatilis</i>	53.50	XP_035518185.1
Bodião-Reticulado	<i>Labrus bergylta</i>	53.23	XP_020486388.1
<i>Burton's Mouthbrooder</i>	<i>Haplochromis burtoni</i>	52.10	XP_014195782.1
<i>Chinese large-mouth catfish</i>	<i>Sirilus meridionalis</i>	51.70	XP_046710878.1
Peixe-zebra	<i>Danio rerio</i>	49.70	NP_852471.1
Carpa-do-limo	<i>Ctenopharyngodon idella</i>	49.70	ACF33511.1
Esturjão chinês	<i>Acipenser sinensis</i>	49.60	AHV85521.1
Esturjão siberiano	<i>Acipenser baerii</i>	48.80	CAB93505.1
<i>Mummichog</i>	<i>Fundulus heteroclitus</i>	45.80	XP_012720582.2
Peixe-napoleão	<i>Cheilinus undulatus</i>	43.10	XP_041636575.1
Humano	<i>Homo sapiens</i>	40.80	AAA36782.1
Camundongo	<i>Mus musculus</i>	40.10	AAA40494.1

Fonte: Autor da dissertação

⁹ Traduzido do dialeto checo: **Rypoun hnědý**

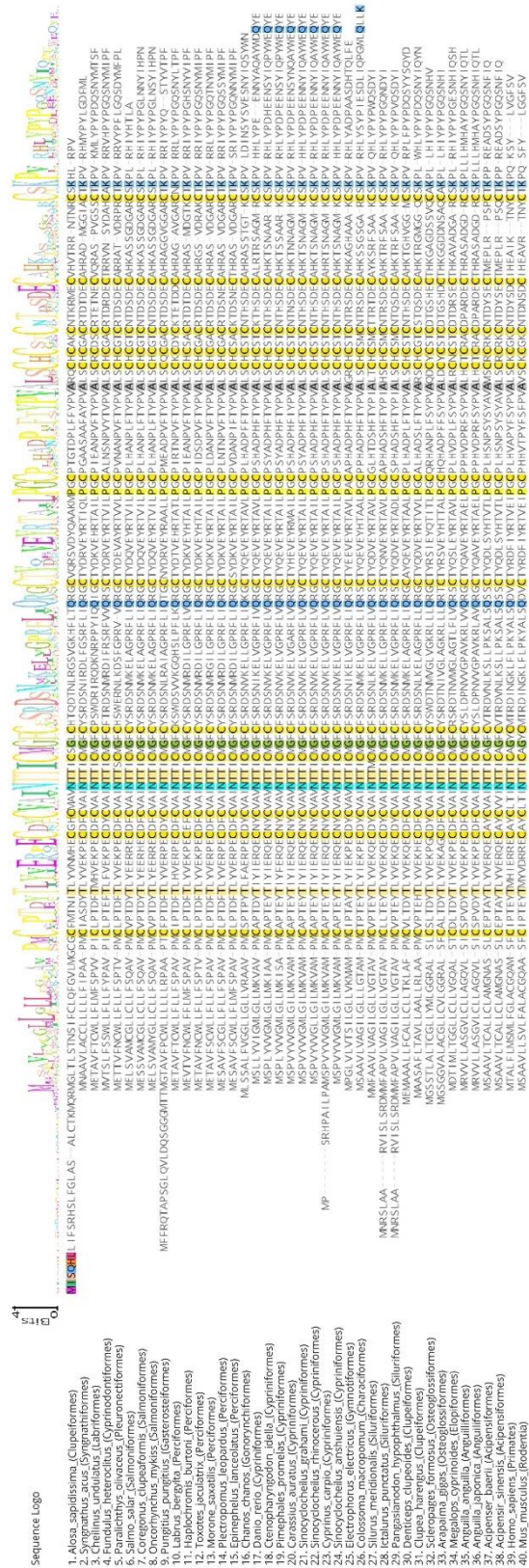
¹⁰ Traduzido do dialeto chinês mandarim: **犀角金線魮**.

¹¹ Traduzido do dialeto estoniano: **Anshui sõõrhuul**.

4.1.5. Evidência para proposta *in silico*: conservação dos aminoácidos e domínios

Propondo uma abordagem pouco usual destacamos abaixo alguns domínios restritos a subunidade beta (NCBI *Conserved Protein Domain Family*: cd00069 e InterPro *Classification of protein families*: IPR018245) (Figura 13). Nessa família de proteínas, os motivos descritos são: sítio de ligação com o receptor (V61, T112, G113, D116, T118 e H119), região de dimerização (L31, E34, I48-V61, T73, C74, C110, T112, H115-H119) e nó de cistina (C24, C38, C41, C49, C53, C74, C89, C105, C107, C110, C117 e C129), envolvendo doze cisteínas interligadas (LAPTHORN *et al.*, 1994; HEARN; GOMME, 2000; GARCIA-CAMPAYO; BOIME, 2001). A similaridade entre os aminoácidos está em 100% e realçadas com cores. As propriedades físico químicas recuperadas com o software ProtParam – ExPASy indicam que a glicoproteína deduzida tem um pI teórico de 6,34 e massa molecular de 15,7 kDa. É notável um maior teor de leucina, glicina e cisteína, cada uma a 10,3%, 10,3% e 9,6% (Mol% dos aminoácidos), respectivamente. O total de resíduos com carga positiva é de 9 e a carga negativa é de 11.

Figura 13: Alinhamento múltiplo de seqüências (ClustalΩ) envolvendo a seqüência de ag-βTSH (33) com mais 39 espécies destacando os aminoácidos com 100% de identidade.

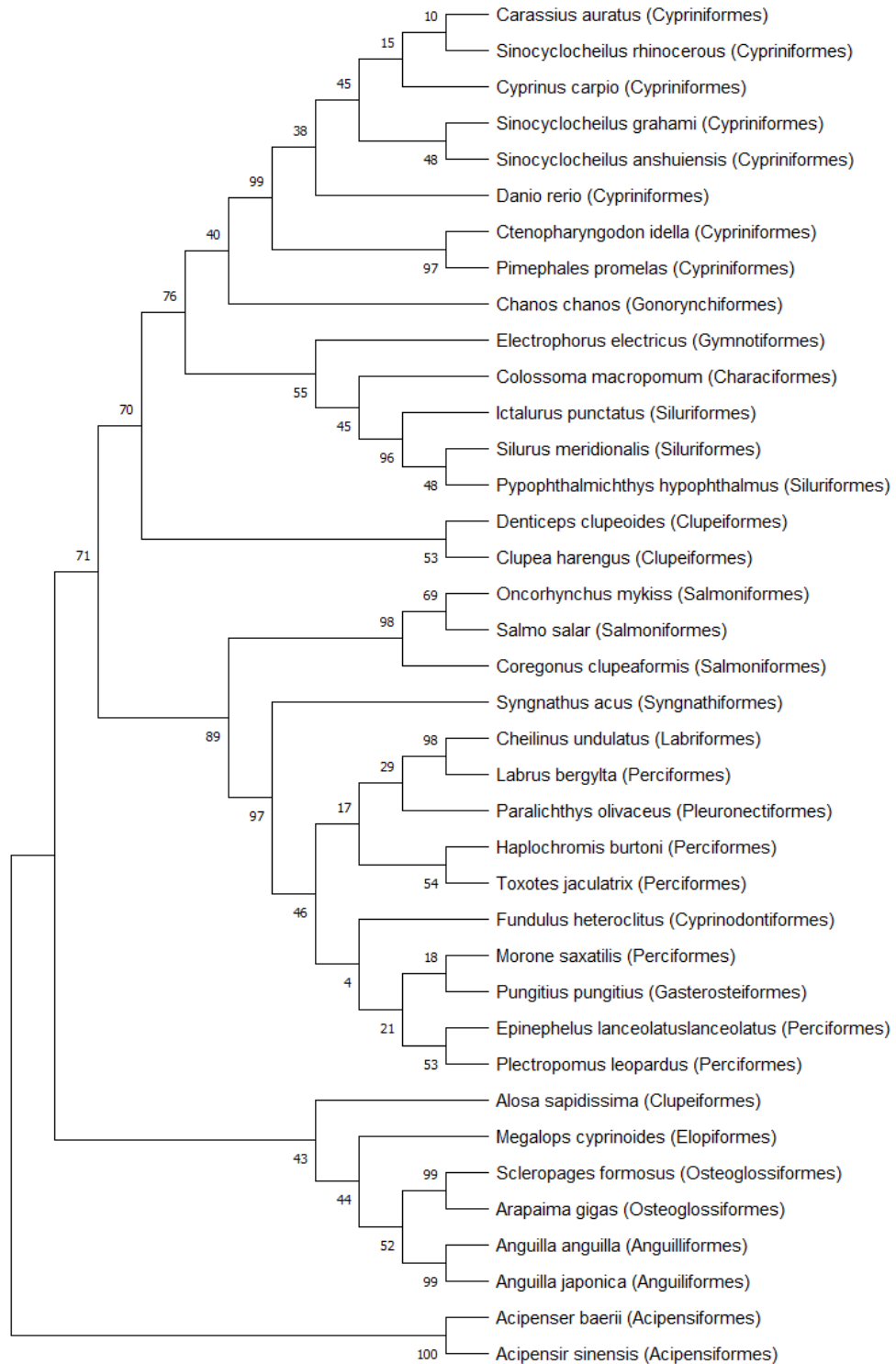


Fonte: Autor da dissertação

4.1.6. Árvore filogenética

A árvore filogenética construída pelo método de máxima verossimilhança é apresentada abaixo (Figura 14).

Figura 14: Árvore filogenética construída com sequências de β TSH.



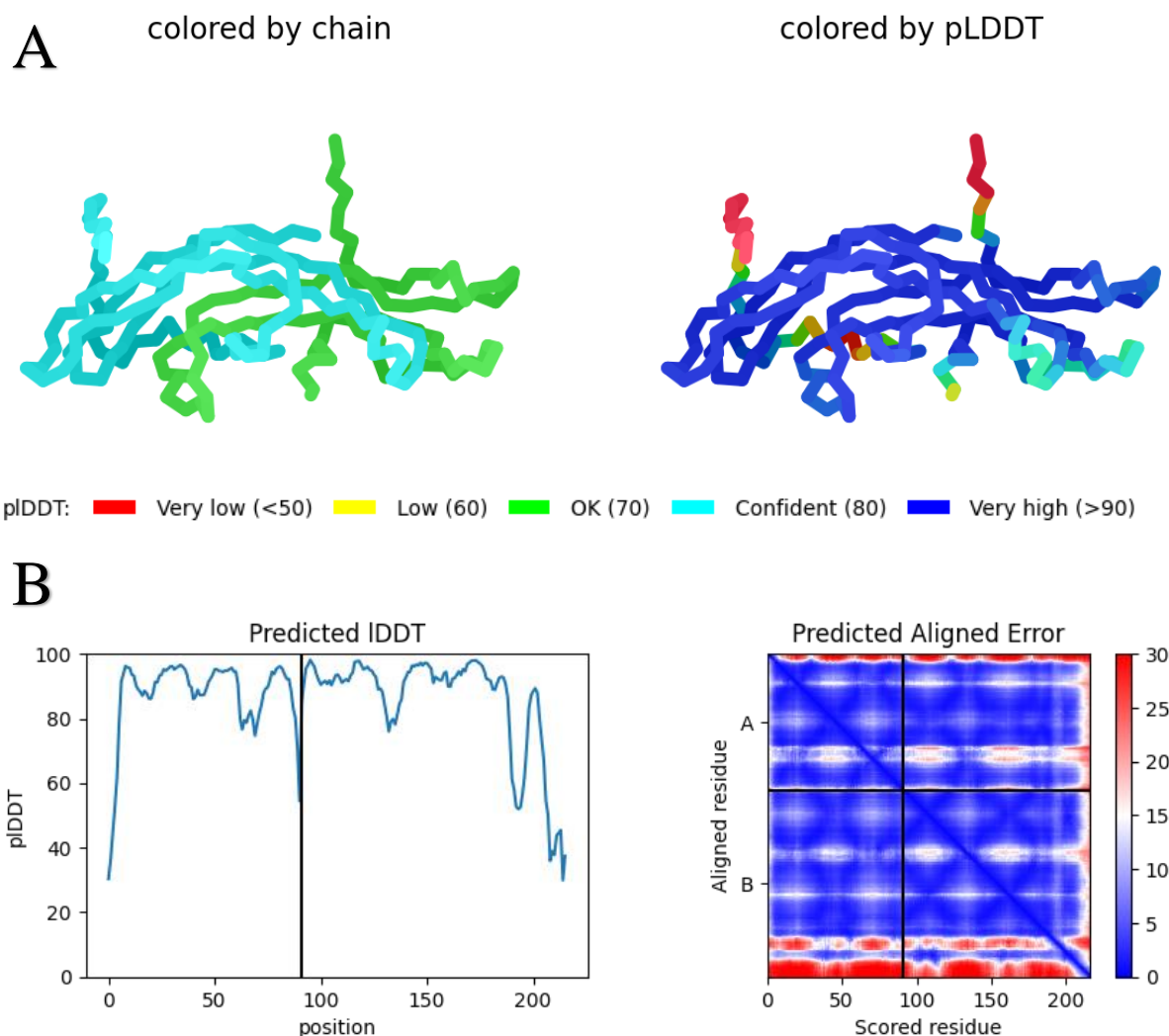
A árvore foi enraizada nas espécies que estão dentro da ordem Acipensiformes. O valor de *log* para verossimilhança foi na ordem de -4603.86. A árvore inicial para busca heurística é obtida pelo método de máxima parcimônia, com cálculo das taxas evolucionárias das espécies por distribuição gamma ($G = 1.0371$). No final, 182 posições foram realizadas para compor a árvore.

4.1.7. Construção do modelo tridimensional para agTSH e cálculo de RMSD com proteínas cristalizadas

O modelo tridimensional(3D) proposto neste trabalho foi construído pelo algoritmo *AlphaFold*, usando a ferramenta ColabFold (MIRDITA *et al.*, 2022). Optamos por excluir o peptídeo sinal ao submeter a construção, uma vez que a proteína madura não contém essa região traduzida.

Dentre os cinco modelos que são gerados, o de número 4 obteve a maior pontuação pLDDT: 86.17. Os resultados obtidos pela metodologia, assim como o modelo tridimensional que será explorado nos próximos tópicos, podem ser visualizados abaixo (Figura 15). Estima-se que o peso molecular do hormônio está em torno de 24 kDa com base na sequência primária de aminoácidos (não está incluso nesse peso molecular a modificação pós-traducional de glicosilação).

Figura 15: Modelo tridimensional *in silico* do agTSH a partir da metodologia *AlphaFold*.

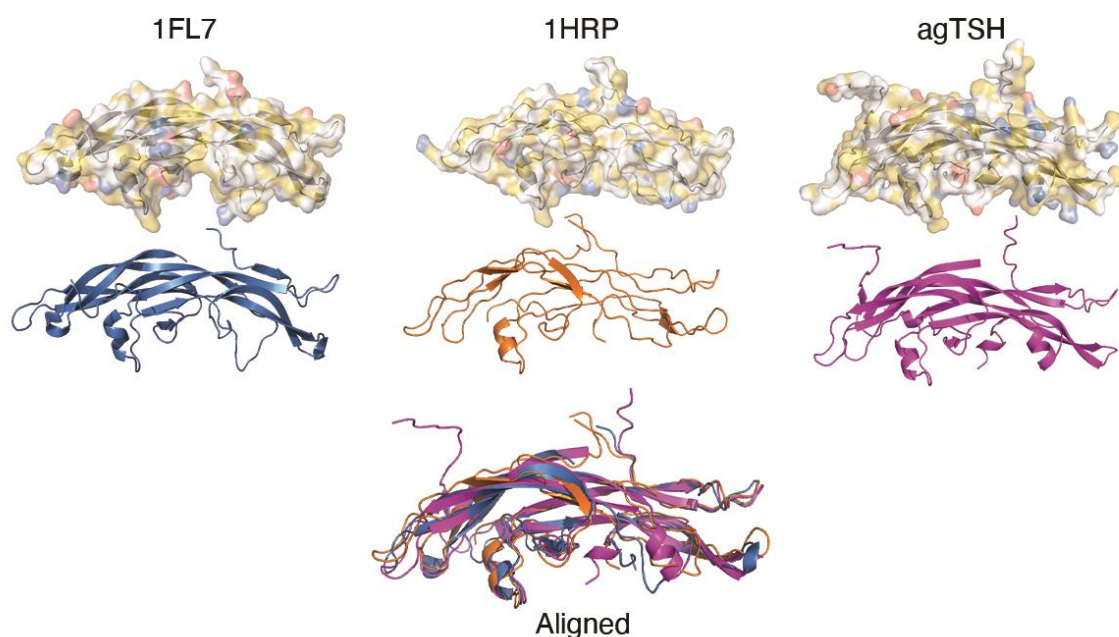


(A) Estrutura *in silico* predita colorida por cadeia (verde-água: subunidade alfa, ciano: subunidade beta) e colorida por pLDDT. A métrica de pontuação pLDDT varia de <50 para modelos não confiáveis (vermelho) a >90 para modelos mais confiáveis (azul escuro) (B) Pontuação por IDDT, classificando a pontuação pLDDT por resíduo e matriz de alinhamento por pares, a direita. Quanto mais azul, a posição relativa dos resíduos e suas ligações são mais confiáveis.

Fonte: Autor da dissertação

O cálculo da distância entre átomos por RMSD (*root-mean-square deviation*) (Figura 16) entre modelos já publicados no banco de dados com o agTSH foi de: 0.830 Å (150/150 átomos) para 1FL7 e 1.095 (152/152 átomos) Å para 1HRP. O cálculo e os modelos apresentados foram realizados com o programa PyMOL.

Figura 16 – Alinhamento do modelo de agTSH (magenta) com proteínas humanas cristalizadas.



1FL7 representa o cristal do FSH humano (azul) e 1HRP o cristal da Gonadotrofina Coriônica humana (laranja). A representação em superfície foi colorida com script YRB (HAGEMANS *et al.*, 2015). Em inglês, *aligned* entende-se alinhados.

Fonte: Autor da dissertação

4.1.8. Propriedades físico-químicas e espaciais para agTSH

Analisar as propriedades físico-químicas e espaciais da proteína nos norteiam para compreendermos melhor sua interação com receptor, disposição de cargas em um determinado pH, comparação ao padrão evolucionário e de conservação e, nesse caso, sua influência na formação do heterodímero e isoladamente, o comportamento de *folding*. Nesse primeiro momento, expomos apenas os resultados para no tópico da discussão da dissertação, possamos retomar e concluir nossas considerações.

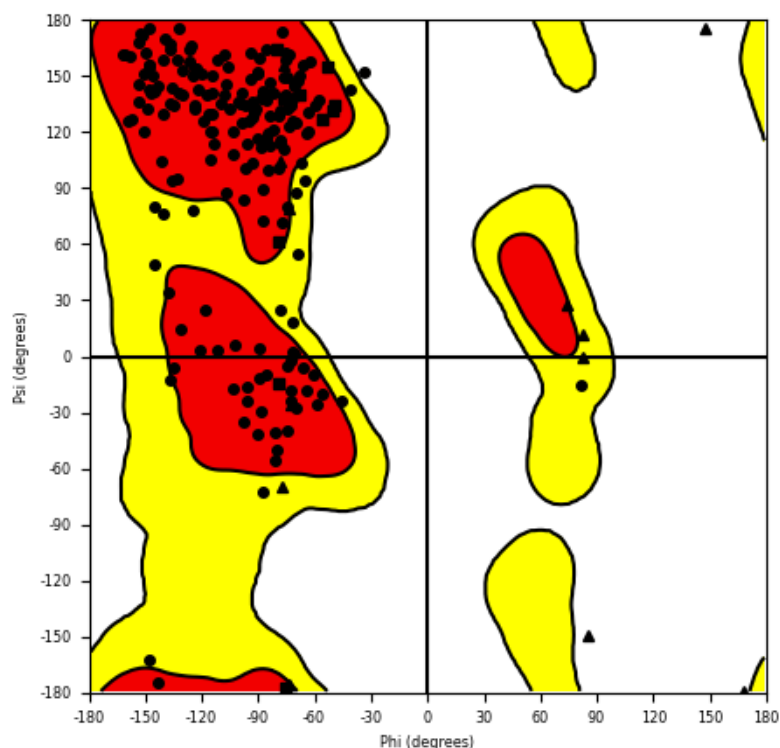
4.1.9. Pontos de Ramachandran para agTSH

O gráfico de Ramachandran permite a visualização dos ângulos diédricos Ψ (psi) x ϕ (phi) das ligações peptídicas dos aminoácidos. Essas ligações, podendo ser alfa-hélice ou folha-beta, compõe a estrutura tridimensional da proteína (Kleywegt, 1995). Todos os aminoácidos são representados com exceção dos casos especiais: prolina e glicina. Esses aminoácidos possuem particularidades diferentes dos 20 existentes: a prolina possui na sua

estrutura, um anel pirrolidina (C_4H_9N) que se liga a estrutura anterior da proteína, limitando a sua conformação. A glicina por sua vez não tem um C_β , e seu C_α não é quiral, induzindo a severos conflitos na conformação e no gráfico de ramachandran por sua livre movimentação espacial (HO; BRASSEUR, 2005).

Considerando os 212 aminoácidos que compõem a proteína madura, 95,755% dos aminoácidos plotados estão em regiões altamente favoráveis, 2,830% em regiões permitidas e 1,415% em regiões não permitidas. Quatro aminoácidos (cadeia α N1, S91 e cadeia β S1 e I125) não foram plotados no gráfico. Os aminoácidos em regiões não permitidas são: cadeia α G29 e cadeia β G30, G41. Os valores obtidos foram recuperados pelo servidor ZLab (<https://zlab.umassmed.edu/bu/rama/>) e o gráfico construído no software Maestro 13.3.121 (Schrödinger, LLC, Nova Iorque, Estados Unidos) (Figura 17).

Figura 17: Gráfico de ramachandran para agTSH.



Em vermelho, as regiões são mais favoráveis para formação de estruturas secundárias alfa-hélices ($+120^\circ/+180^\circ$ Psi e $-60^\circ/-150^\circ$ Phi) e folhas-beta ($+30^\circ/-60^\circ$ Psi e $-60^\circ/-150^\circ$ Phi). Regiões em amarelo são regiões permitidas e o restante regiões não permitidas. Em inglês, *degrees* representa graus.

Fonte: Autor da dissertação

4.1.10. Topologia proteica para agTSH

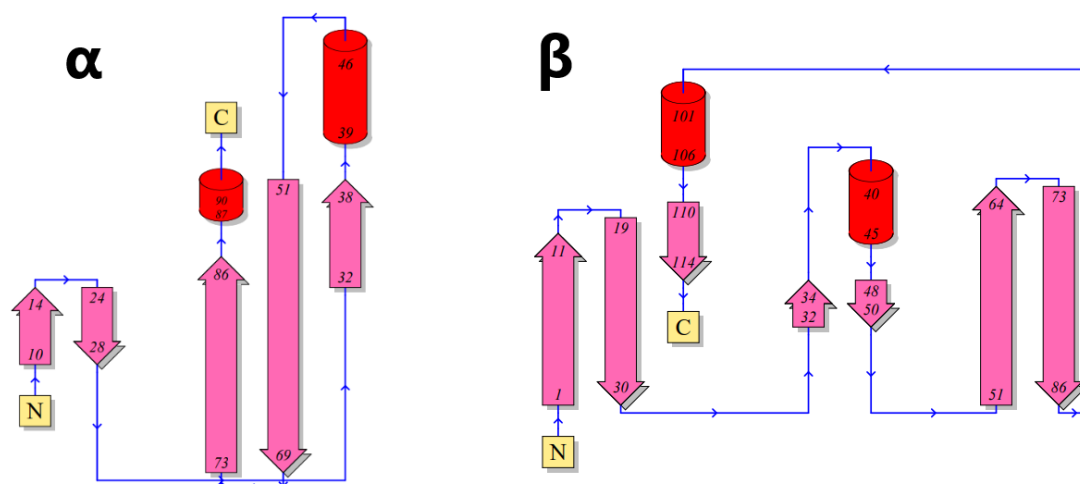
O entendimento da topologia da proteína está relacionado às estruturas secundárias e orientação a partir da sequência de aminoácidos. Dessa forma, simplificamos a estrutura

tridimensional e clarificamos o entendimento da orientação de aminoácidos e curvatura de enovelamento. Sempre estaremos observando a partir do começo da cadeia (N-terminal) até seu término (C-terminal), em um sentido único (para proteínas que seguem topologias lineares). A apresentação em 2D pode adotar diferentes representações tanto para folhas-beta quanto alfa-hélice (GEIS-ASTEGGIANTE; LEE; FENSELAU; 2015).

Segundo a classificação dada no banco de dados CATH Protein Structure Classification (*Class, Architecture, Topology and Homologous Superfamily*) (KNUDSEN; WIUF, 2009), o agTSH se enquadraria na classificação: **C: 2** (Predominante beta), **A: 2.10** (Fitas), **T: 2.10.90** (Citocinas da subunidade beta, nó de cistina) e **H: 2.10.90.10** (Citocinas com nó de cistina).

Os dados de topologia do agTSH foram obtidos no servidor pdbsum (LASKOWSKI *et al.*, 2018) com a sequência construída no servidor *Alphafold*. O diagrama que ilustra o arranjo das folhas betas na configuração beta (folha pregueada) e posições das alfa-hélices são apresentados na figura abaixo (Figura 18).

Figura 18: Diagrama da topologia para as cadeias α e β .



As setas representam as estruturas das folhas beta e os cilindros as alfas hélices. “C” e “N” são as extremidades C-terminal e N-terminal, respectivamente.

Fonte: Autor da dissertação

No mesmo servidor, os motivos de estruturas secundárias calculadas são apresentados no programa PROMOTIF (v 3.0, Gail Hutchinson's). Maiores detalhes da documentação (como a descrição dos aminoácidos envolvidos em diferentes motivos, análise de ligante e conflitos estereoquímicos) podem ser encontrados no site

“www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPage.pl?pdbcode=index.html” com o código r002 e senha “160506”.

Além dos dados da topologia, apresentamos uma região importante nomeada *seat-belt* entre a décima cisteína e décima segunda cisteína da subunidade β nos hormônios glicoproteicos. Os aminoácidos perpassam a subunidade alfa e está envolvida na associação entre as subunidades, ligação com o receptor e sua ativação (GROSSMAN *et al.*, 1993). Sua representação está apontada em duas frentes: sem transparência e com transparência em superfície (Figura 19).

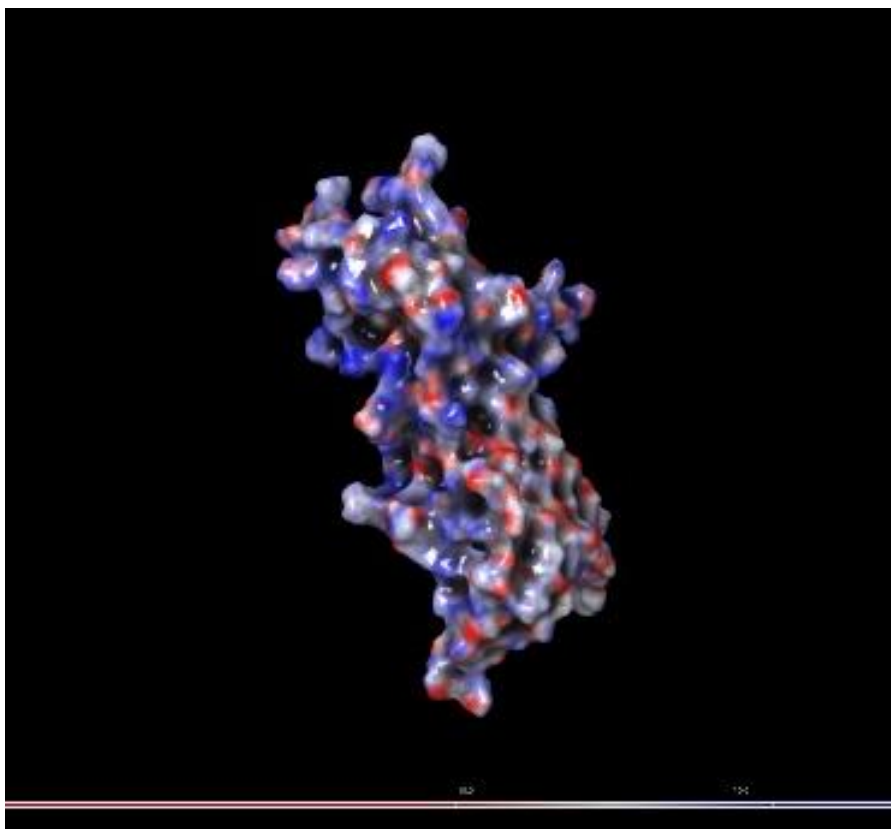
4.1.11. Valores teóricos dos parâmetros eletrostáticos de agTSH

Os parâmetros eletrostáticos de uma proteína são classificados de acordo com a cadeia lateral dos aminoácidos. Essa influência dos valores do pKa e suas ligações de hidrogênio e van der Waals refletem no *folding* em determinado pH. Utilizando os servidores Adaptive Poisson-Boltzmann Solver (APBS) e PDB2PQR podemos calcular a solvatação teórica dos aminoácidos que compõe o agTSH. O software Maestro 13.3.121 (uso acadêmico) foi utilizado para representação do potencial eletrostático com superfície do agTSH.

Os dados descritos (Tabela 5) foram obtidos a partir dos seguintes valores: pH 7.0 e propka 3.4.0 (LI; ROBERTSON; JENSEN, 2005), campo de força AMBER (*Assisted Model Building with Energy Refinement*), excluindo água. A representação colorida (Figura 23) destaca entre azul (mais negativo) a vermelho (mais positivo). Para os parâmetros eletrostáticos, usamos a constante dielétrica do soluto como 1,0, a constante dielétrica do solvente em 80,0, o raio do solvente em 1,4 Å, temperatura de 298,0 K e extensão da régua de -10,0 à +10,0 Å.

O valor do pI (C_T) pode ser calculado através da fórmula de Henderson-Hasselbalch: $C_T = C_N + C_P$, do qual C_N representa as somas das cargas negativas ($C_N = \sum_{i=1}^n \frac{-1}{1+10^{pK_{Ni}-pH}}$) e C_P as cargas positivas ($C_P = \sum_{i=1}^n \frac{1}{1+10^{pH-pK_{Pi}}}$). De acordo com os valores obtidos para cada aminoácido na predição, o pI da proteína é de 9,40. O cálculo do valor da energia livre de *folding* (kcal/mol) para os respectivos pH seguem formulas descritas (FANG, 2012). De acordo com os resultados observado, o pH que favoreceu a menor energia livre para *folding* (-3,90 kcal/mol) está próximo de 9 (9,3). O sumário da documentação e os resultados apresentados são descritos no Apêndice C dessa dissertação.

Figura 20: Representação do agTSH com superfície colorida por potencial eletrostático Poisson Boltzman.



A coloração azul são aminoácidos com cargas negativas e em vermelho os aminoácidos em carga positiva.

Fonte: Autor da dissertação

Tabela 5: Varredura teórica dos valores para *fold*ing do agTSH.

pH	Energia livre para <i>fold</i> ing (kcal/mol)	Carga da proteína (<i>unfolded</i> / <i>folded</i>)
0	16,87	33,00/32,98
1	16,77	32,98/32,78
2	15,90	32,76/31,38
3	12,43	31,00/27,48
4	7,92	24,05/21,42
5	4,98	16,25/14,12
6	1,32	12,13/9,02
7	-2,01	7,00/5,49
8	-3,31	4,12/3,51
9	-3,89	1,72/1,47
10	-3,18	-6,13/-4,25
11	1,64	-18,06/-13,47
12	7,18	-23,64/-20,33
13	11,17	-28,78/-26,18
14	13,63	-30,72/-29,78

Fonte: Autor da dissertação.

4.1.12. Dinâmica molecular e cálculos de energia livre de MM-GBSA e umbrella sampling para agTSH com diferentes receptores

A dinâmica molecular do melhor modelo construído para agTSH foi aplicada de acordo com a metodologia descrita e os valores de energia de ligação (ΔG_{bind}) e potencial de força média (ΔG_{PMF}) descritos de acordo com o receptor complexado (Tabela 6).

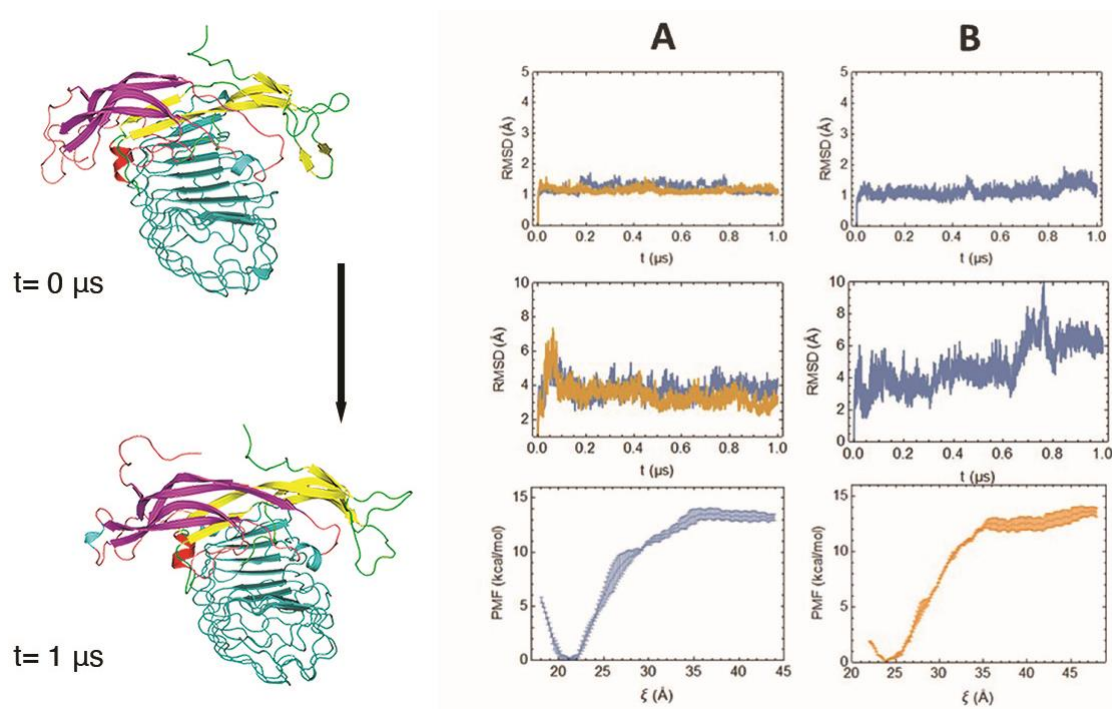
Tabela 6: Valores obtidos da dinâmica molecular para agTSH.

Modelo	ΔG_{bind} (kcal/mol)	RMSD (Å)	ΔG_{PMF} (kcal/mol)
I. hTSH + hTSHR	$-79,9 \pm 0,1$	3,2 – 4,4	$13,3 \pm 0,4$
II. agTSH + sfTSHR	-73,0	5,8	$13,1 \pm 0,5$
III. agTSH + hTSHR	$-38,0 \pm 2$	6,3 – 12,6	$5,6 \pm 0,6$

Fonte: Autor da dissertação

Os resultados sugerem que em comparação com o modelo humano (I) os valores são próximos na simulação utilizando o receptor de *Scleropages formosus* (II), com uma diferença de 8,64% para ΔG_{bind} , 24,14% para RMSD e 1,50% em ΔG_{PMF} . Uma vez que o modelo do agTSH com receptor humano (III) também foi proposto e seus valores descritos, fica claro que o modelo apresenta uma instabilidade maior, com valores fluando na ordem de duas vezes mais. Os gráficos obtidos das simulações e o modelo do complexo agTSH:sfTSHR estão representados abaixo (Figura 21).

Figura 21: Perfis dos resultados obtidos por dinâmica molecular e *umbrella sampling*.



A esquerda da imagem, temos a simulação do agTSH e a parte extracelular do receptor de Aruanã-dourado (*S. formosus*) em *umbrella sampling* na trajetória de 1 μ s. Nas colunas a direita, os dois primeiros gráficos na vertical – de cima para baixo - **(A)** são os perfis de RMSD da estrutura da cadeia lateral dos aminoácidos do receptor e hormônio humano, e o último gráfico o perfil para *umbrella sampling* do complexo. **(B)** Entende-se a mesma descrição para A da cadeia lateral dos aminoácidos do receptor de *S.formosus* e *A.gigas*, com o gráfico de *umbrella sampling* do complexo. A cor azul e laranja representam replicadas quando sobrepostas.

Fonte: Autor da dissertação

4.1.13. Tabela de informações recuperadas com Expasy

Dada a disponibilidade das ferramentas disponíveis no portal de recursos Expasy (DUVAUD *et al.*, 2021) para melhor compreensão do hormônio em si, a tabela abaixo (Tabela 7) reúne informações teóricas que podem ser adquiridas com a sequência primária dos aminoácidos e sua estrutura resolvida por inteligência artificial (*Alphafold*). Os resultados descritos excluem o peptídeo sinal discriminado pela metodologia já descrita na dissertação.

Tabela 7: Informações recuperadas com base na sequência primária de agTSH.

Propriedade	Valor	Referência
Número de aminoácidos	215	ProtParam
Índice de instabilidade	48.46	ProtParam
Massa molecular ¹²	24129.66 Da	PeptideMass
Ponto isoelétrico	8.17	PeptideMass
Energia livre para folding	pH 9,3: -3,9 kcal/mol	APBS-PDB2PQR

Fonte: Autor da dissertação

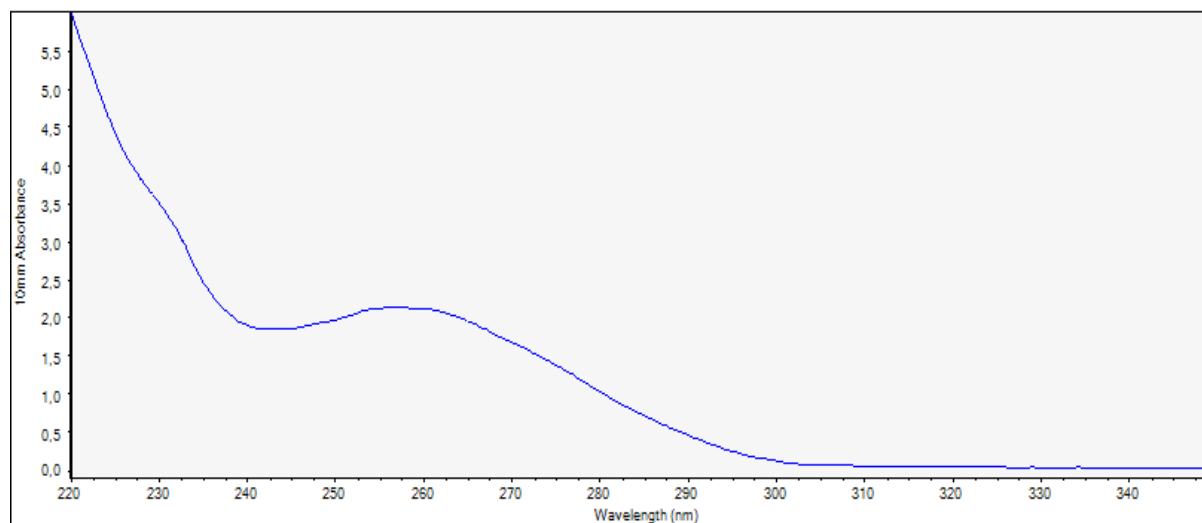
4.2. Síntese do hormônio recombinante agTSH: atividades práticas

4.2.1 Quantificação da extração de RNA total a partir da hipófise de *Arapaima gigas*.

O material genético mRNA extraído da hipófise foi quantificado por espectrofotometria *NanoDrop 2000*. Após calibrar o equipamento com tampão Tris-HCl pH 7,5, aplicamos um microlitro da amostra (50 µL totais), resultando a densidade de 84,0 ng/µL de RNA, em uma razão de pureza ($A_{260/280}$) dentro do esperado, com valor de 2,09 (Figura 22). A leitura do espectro foi realizada com varredura entre 220nm e 350nm e o comprimento de onda para análise foi utilizado em 260nm.

¹² A massa molar do heterodímero teórico – sem modificações pós-traducionais - corresponde a soma da massa das subunidades α (10,4 kDa) e β (13,7 kDa). Os valores foram obtidos pelo mesmo software.

Figura 22: Cromatograma da análise de 1 μ L do mRNA extraído da hipófise de *Arapaima gigas*.

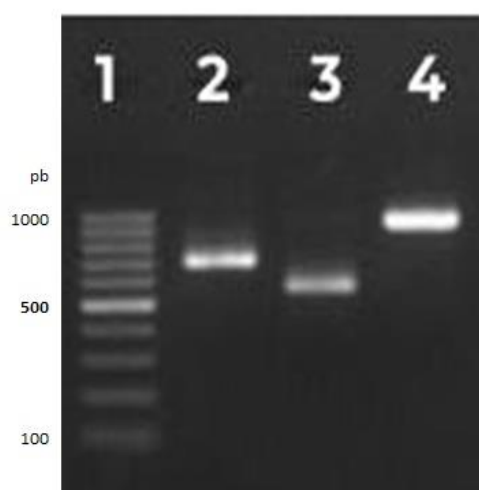


Fonte: Autor da dissertação

4.2.2 Amplificação das extremidades de cadeia (RACE)

O resultado da amplificação das extremidades da cadeia para região 3' utilizando o primer GSP TSH (AACACCACCATCTGCATGGG) foi visualizado por eletroforese em gel de agarose no poço 3 (Figura 23).

Figura 23: Eletroforese em gel de agarose para 3' RACE.



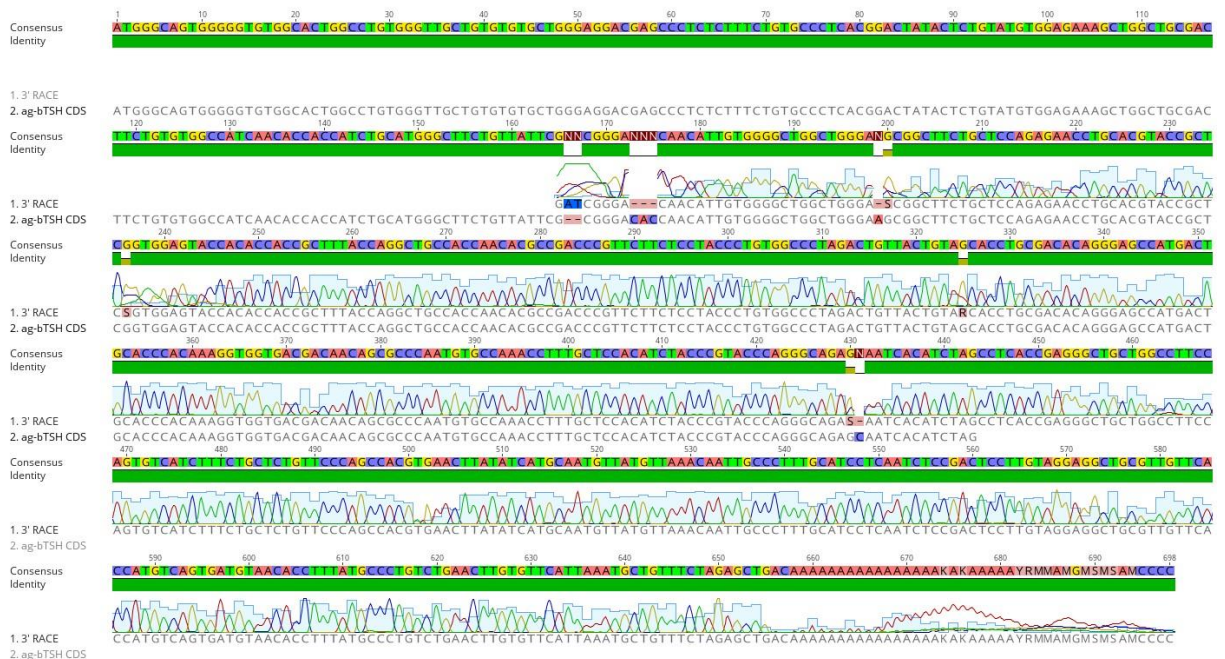
Com indicação das bandas, temos: (1) Marcador de peso molecular 1000-100 pb. (2) Controle agLH (700 pb). (3) ag- β TSH (650pb). (4) agGH (1000 pb).

Fonte: Autor da dissertação

4.2.3 Sequenciamento de SANGER e comparação com proposta *in silico*

O sequenciamento das amostras realizado no Centro de Pesquisa e Diagnóstico Molecular de Doenças Genéticas, do departamento de biofísica na Universidade Federal de São Paulo (UNIFESP) – Escola Paulista de Medicina (São Paulo, SP) para região 3' (Figura 24) do gene ag- β TSH corrobora com o resultado do método de predição *in silico*. A figura revela um perfil de eletroferograma definido para as bases nitrogenadas, indicando uma boa qualidade do sequenciamento e os nucleotídeos alinhados confirma nossa hipótese apresentada anteriormente.

Figura 24: Eletroferograma e alinhamento do sequenciamento do RACE correspondente a região 3' do ag- β TSH com a predita *in silico*.



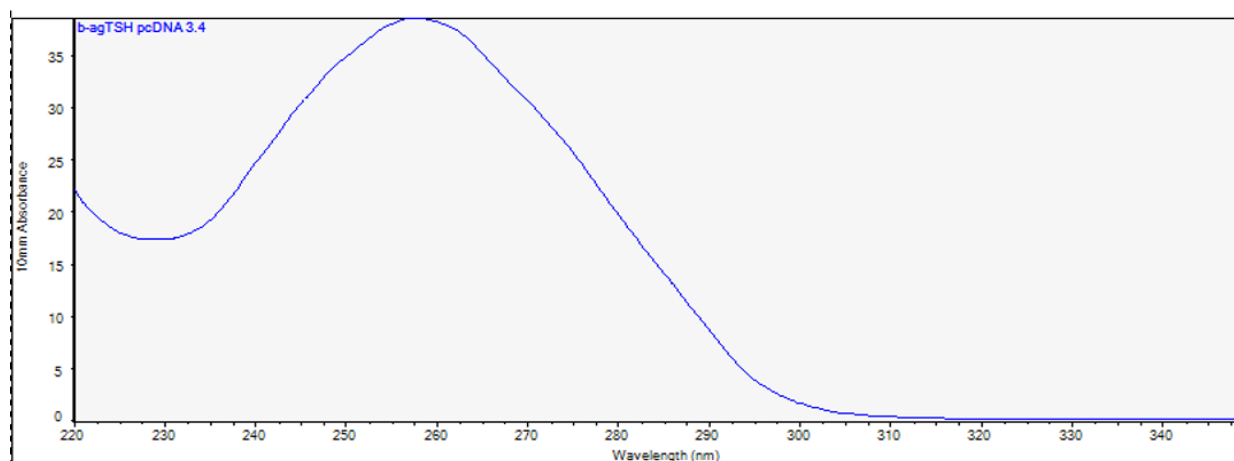
Fonte: Autor da dissertação

4.2.4. Quantificação do plasmídeo pcDNA 3.4 + ag- β TSH amplificado em *Escherichia coli*

O plasmídeo pcDNA 3.4 + região codificante do ag- β TSH (com adição da região de Kozak (ACC) antecedendo o peptídeo sinal (ATG) foi amplificado e purificado de acordo com a metodologia descrita anteriormente. 300 μ L de água livre de nuclease foram utilizados na diluição do DNA, resultando na concentração de \cong (aproximadamente) 2 μ g/ μ L e razão de absorvância $A_{260/280}$ 1,90, quantificado por espectrofotometria (Figura 25). A leitura do

espectro foi realizada com varredura entre 220nm e 350nm e o comprimento de onda para análise foi utilizado em 260nm.

Figura 25: Cromatograma da análise de 1 μ L do plasmídeo pcDNA 3.4 + ag- β TSH amplificado *Escherichia coli* e purificado

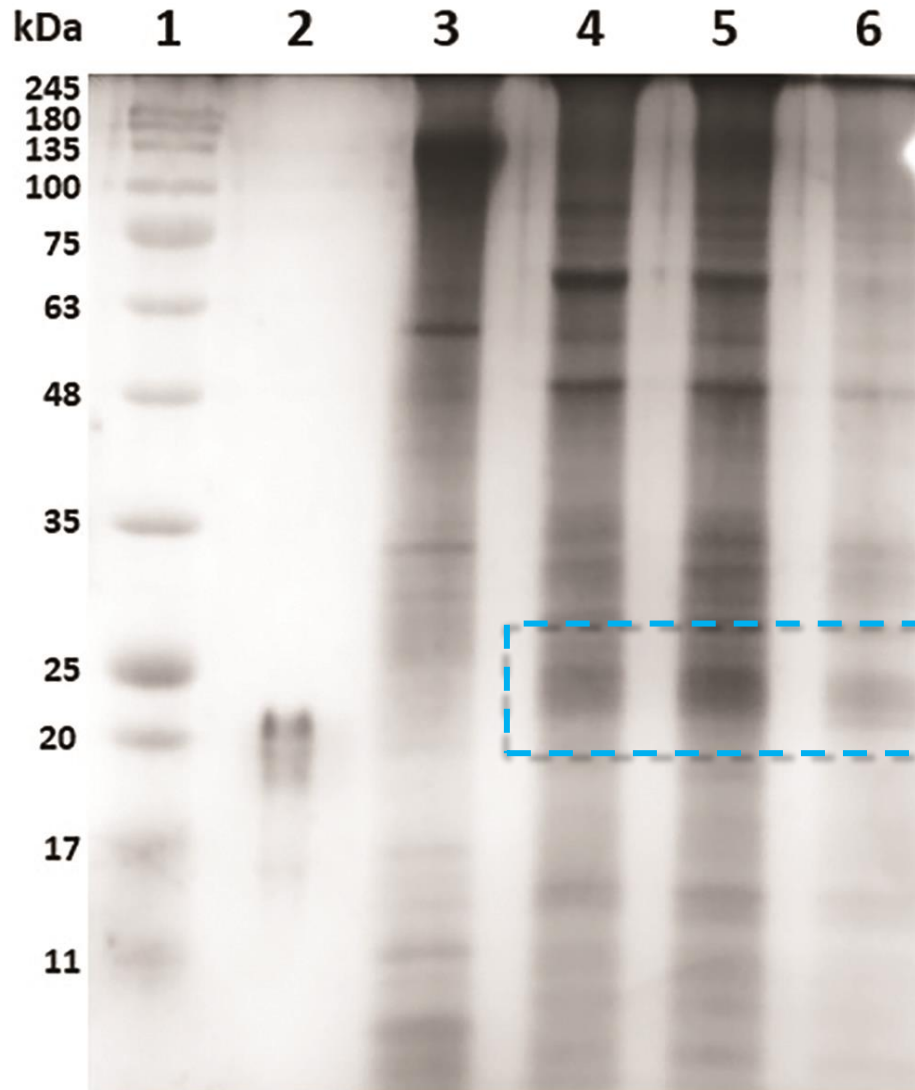


Fonte: Autor da dissertação

4.2.5. Avaliação por SDS-PAGE da produção de agTSH em células HEK293 cultivadas em suspensão

A análise do primeiro ensaio, fundamental para determinação do melhor dia de síntese do hormônio recombinante, é realizada por eletroforese em gel de poliacrilamida (SDS-PAGE). Foram analisados os dias três, cinco e sete após a transfecção. O gel foi revelado utilizando a coloração com nitrato de prata (Figura 26).

Figura 26: Eletroforese em gel de poliacrilamida 15% referente a produção piloto de agTSH em células HEK293.



(1) Marcador de massa molecular BLUEye Prestained Protein Ladder (Sigma Aldrich, St Louis, MI, EUA) – 5 μ L; (2) Thyrogen (hTSH) 600 ng (3) Meio condicionado de células transfectadas sem DNA – controle negativo (coletado no quinto dia) – 20 μ L (4) Meio condicionado de células transfectadas no terceiro dia – 20 μ L; (5) Meio condicionado de células transfectadas no quinto dia – 20 μ L; (6) Meio condicionado de células transfectadas no sétimo dia – 20 μ L. O retângulo indica a posição do agTSH heterodímero glicosilado.

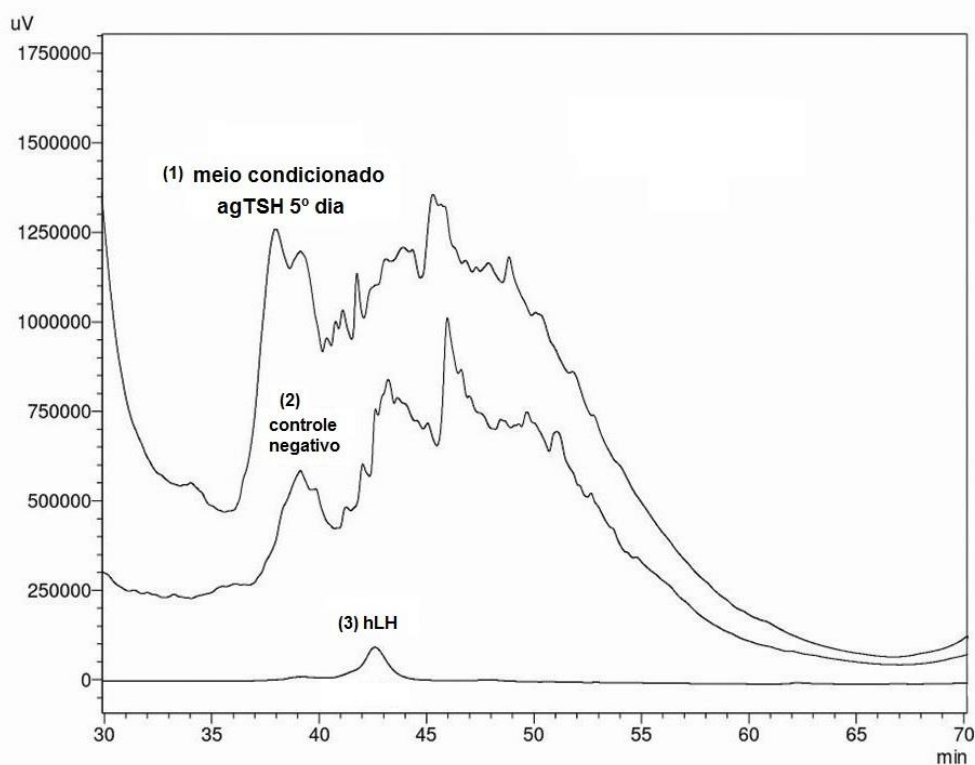
Fonte: Autor da dissertação

Dessa forma, avaliando-se a intensidade das bandas na posição esperada de aproximadamente (\approx) 25 kDa, foi considerado como o melhor dia de coleta o quinto dia após transfecção transiente para produção de agTSH em células HEK293.

4.2.6. Quantificação por RP-HPLC do agTSH expresso em célula HEK293

Após análise por SDS-PAGE, uma cromatografia líquida de alta eficiência por fase reversa (RP-HPLC) comparando com o controle negativo (Figura 27) ajuda a visualizar o *background* da produção do agTSH em células HEK293. O tempo de retenção suspeito para retenção do hormônio recombinante está entre 37-46 min, com uma maior definição dos picos resultantes. Entre esses minutos, dois pools #P1 (37-40 min) e #P2 (40-45 min) foram coletados. As amostras foram precipitadas por TCA e visualizadas em SDS-PAGE 15% com nitrato de prata, demonstradas no próximo tópico

Figura 27: Comparação entre os três cromatogramas obtidos em RP-HPLC.



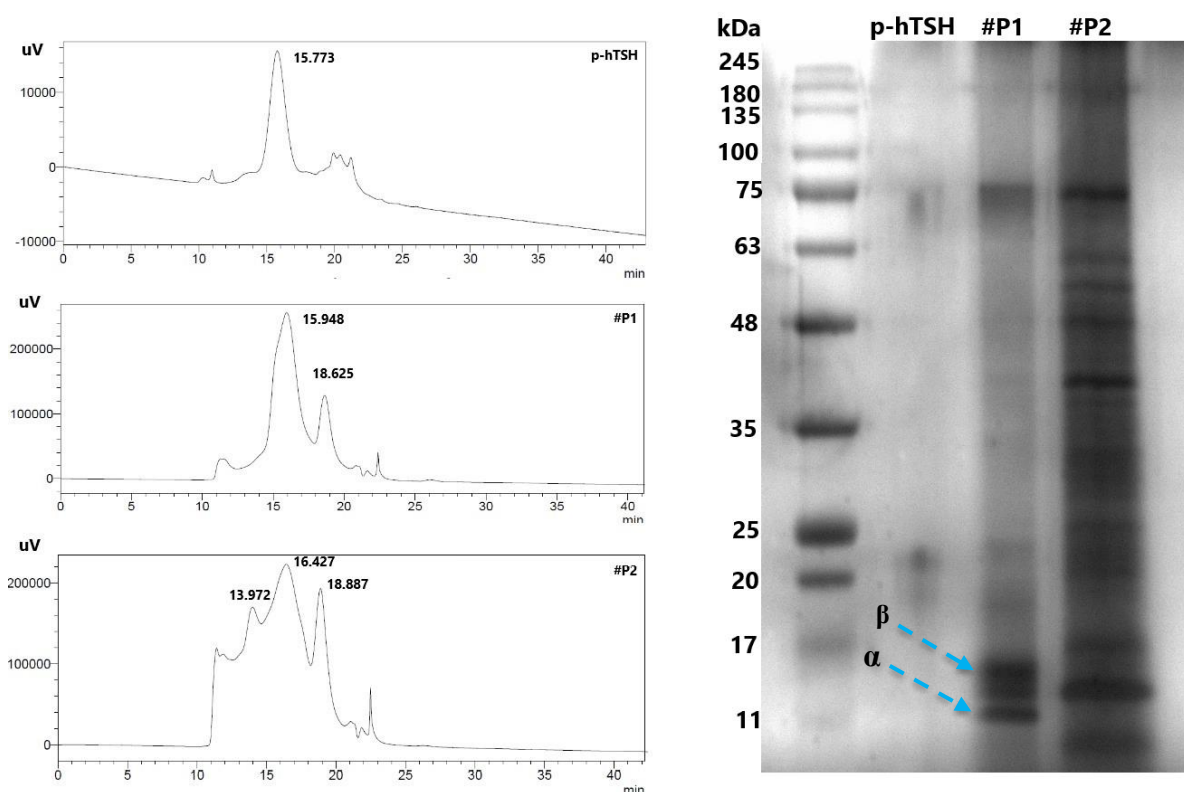
(1) Curva obtida aplicando 5000 μL do meio condicionado concentrado ($V_0 = 30 \text{ mL}$) coletado de células HEK293F transfectadas com αGTH e βTSH após cinco dias a 37°C , 130 rpm e 8% CO_2 ; (2) Curva obtida aplicando 4000 μL do meio condicionado concentrado ($V_0 = 30 \text{ mL}$) coletado de células HEK293F transfectadas sem DNA e (3) Curva obtida aplicando-se hLH hipofisário com tempo de retenção de 42,59 min.

Fonte: Autor da dissertação

4.2.7. Análise dos *pools* em HPSEC do agTSH

Uma fração de 70 μL das amostras #P1 e #P2, após tratamento previamente mencionado, foi visualizado em SDS-PAGE e, o restante da amostra (500 μL) de cada um dos *pools* foi injetado em HPSEC (Figura 28).

Figura 28: Cromatogramas da análise em HPSEC e SDS-PAGE dos pools #P1 e #P2 coletados da primeira etapa de purificação em RP-HPLC



Nos cromatogramas da esquerda, a legenda no canto superior direito indica: (**p-hTSH**) o hTSH hipofisário (5 $\mu\text{g}/5\mu\text{L}$), com tempo de retenção (T_R) 15.773 e área relativa (A_R)=1569 u.a.; (**#P1**) o pool coletado em RP-HPLC (#37-40 min) concentrado para 500 μL e (**#P2**) o pool coletado em RP-HPLC (#40-45 min) com as frações. A direita, SDS-PAGE 15%, condição não-reduzida, gel revelado por nitrato de prata. Da esquerda para direita, temos: marcador massa molecular BluEye PreStained (Sigma-Aldrich), hTSH hipofisário, pool #P1 e pool #P2.

Fonte: Autor da dissertação

Há uma proximidade do tempo de retenção do padrão humano e de *Arapaima gigas*. O tempo de retenção difere em apenas 0,161 minutos. Conseguimos visualizar as subunidades dissociadas no #P1 entre 17 e 11 kDa. A presença de duas formas próximas

para subunidade β reflete em possíveis formas glicosiladas do hormônio que alteram minimamente seu peso molecular.

4.2.8. Rendimento estimado da produção

O processo de análise partindo de uma produção do meio condicionado apresentou uma recuperação esperada dentro das metodologias aplicadas. As concentrações calculadas estão descritas (Tabela 8) abaixo. Estima-se que há uma perda de 45,5% no processo de concentração e aplicação na coluna. Entretanto, o produto (agTSH) está com 95% de pureza, analisando-se a área sob a curva obtida por HPSEC.

Tabela 8: Concentração de agTSH estimados em HPSEC após coleta em RP-HPLC.

	T_R (min)	Área (u.a.)	Concentração de agTSH ($\mu\text{g/mL}$)
Pool #P1 Coletado da RP-HPLC	15,948	31580	54,86
Pool #P1 Coletado da HPSEC	15,934	1962	25,01

T_R = tempo de retenção; **(u. a.)** = Área sob a curva em unidades arbitrárias; **Concentração de agTSH**: Valores calculados em relação a área sob a curva referente ao padrão de hTSH hipofisário (5 μg).

Fonte: Autor da dissertação

5. Análise e discussão dos resultados

Os resultados deste trabalho serão discutidos em duas frentes: uma em relação à bioquímica estrutural do agTSH com foco na bioinformática e outra na expressão transiente do hormônio em células humanas HEK293F cultivadas em suspensão.

De forma geral, os objetivos do trabalho foram alcançados. Foram introduzidas metodologias extremamente atuais com relação à bioquímica estrutural de proteínas que auxiliaram e nortearam o desenvolvimento da dissertação. Além disso, propusemos trabalhar com um hormônio cujos dados não estão depositados em base de dados. Desde sua identificação, conseguimos iniciar os estudos de expressão e caracterização. Adiantamos que se faz necessário prosseguir na adaptação das melhores condições de expressão e caracterização do agTSH recombinante, tais como ensaios imunológicos e biológicos que comprovem a atividade hormonal. Entretanto, para fins de estudos de conservação e

preservação do *Arapaima gigas*, o hormônio obtido já é um alvo consolidado de estudo com potencial para futuras abordagens.

5.1. Sequência primária do ag- β TSH

Esse tópico tem como finalidade abordar as informações recuperadas da árvore filogenética proposta nesse trabalho (ver tópico 4.3). A mais recente árvore filogenética publicada foi com relação ao genoma de *Arapaima gigas* (VIALLE *et al.*, 2018). No trabalho publicado, sustenta-se o argumento que os Osteoglossomorpha (superordem da qual o *Arapaima gigas* faz parte) é um grupamento irmão aos Elopomorpha (onde fazem parte *Anguilla anguilla* e *Megalops cyprinoides*, por exemplo). Ambos os grupos mencionados anteriormente tem raízes em Clupeocephala. Esses resultados não divergem da construção da árvore filogenética para ag- α GTH (FARIA *et al.*, 2013), onde os mesmos resultados são confirmados.

Apesar de alguns valores de distâncias evolutivas na árvore filogenética serem inferiores a 70 (adequado para inferiçã), podemos observar que os resultados descritos na dissertação corroboram com essas árvores filogenéticas previamente indicadas no parágrafo anterior. Dessa forma, temos mais uma comprovação na identificação dos nucleotídeos correspondente ao ag- β TSH.

As sequências gênicas correspondente ao posicionamento das glicoproteínas apresentam localizações distintas no cromossomo de diferentes espécies. Em humanos (*Homo sapiens*), foi clonada e identificada no cromossomo 1; em camundongo (*Mus musculus*), no cromossomo 3; em peixe-zebra (*Danio rerio*), cromossomo 6 (<https://www.ncbi.nlm.nih.gov/gene>). Entretanto, todos os 529 genes e suas localizações apresentaram uma única região no DNA que é responsável pela expressão do β TSH.

A sequência primária dos aminoácidos da subunidade β compartilha motivos específicos que tem similaridade com outras espécies (além de peixes). Destes motivos, o mais notório para elucidar a sequência e que nos deu segurança para continuação dos estudos foi o nó de cistina, formado por doze cisteínas das quais derivam seis pontes dissulfeto, importantes na estabilização da proteína (MCDONALD; HENDRICKSON, 1993.; SUN; DAVIES, 1995). As cisteínas envolvidas são: ²⁴C-⁷⁴C; ³⁸C-⁸⁹C; ⁴¹C-¹²⁹C; ⁴⁹C-¹⁰⁵C; ⁵³C-¹⁰⁷C; ¹¹⁰C-¹¹⁷C. O motivo estrutural pode ser encontrado na base de dados Pfam 35.0 *cys_knot* (pfam: 00007) ou na base de dados do InterPro “IPR006208”. Podemos também com a sequência primaria identificar a sequência de aminoácidos “ICMGFCYSRDTNIV” (ver resultados, tópico 4.2.3) envolvida no sítio de dimerização,

assim como outros motivos estruturais já descritos e fundamentados. Além disso, o sítio de N-glicosilação (NTT) é presente nas 40 espécies que foram alinhadas. Esse resultado é melhor explorado no próximo tópico com elucidação dos resultados e complexos de ligação ao receptor de *Scleropages formosus*.

A subunidade β dos hormônios glicoproteicos apresentam diferenças em relação as suas cargas. A cadeia madura do TSH possui 124 aminoácidos, sendo 74 aminoácidos polares (48 não carregados e 26 carregados) e 50 aminoácidos hidrofóbicos; FSH são 108 aminoácidos, distribuído em 69 aminoácidos polares (45 não carregados e 24 carregados) e 40 aminoácidos hidrofóbicos e, por fim, o LH com 117 aminoácidos, tem sua distribuição de resíduos em 71 aminoácidos polares (47 não carregados e 24 carregados) e 46 aminoácidos hidrofóbicos. Levando a discussão para um sistema de análise por cromatografia baseada na interação hidrofóbica, o TSH é a glicoproteína que apresenta menor tempo de retenção entre os LH, de forma intermediário e FSH com maior tempo de retenção, refletindo menor hidrofobicidade entre esses hormônios. Os dados corroboram com a distribuição apresentada nesse trabalho e com o publicado com o agFSH (SEVILHANO *et al.*, 2017).

Todos esses dados foram compilados e catalogados no NCBI e já há um número de acesso (OP125854, vide Anexo 2) que será liberado com a publicação do artigo científico.

5.2. Estrutura quaternária do agTSH

Iniciando os pontos que serão apresentados com base na estrutura quaternária da proteína, isso é, o enovelamento da interação das duas subunidades ag- α GTH e ag- β TSH, temos uma estrutura que apresenta um grau de similaridade alto com modelos de hormônios humanos cristalizados *in vitro*. Portanto, existe um ponto positivo ao resolvermos estruturas com metodologia *ab initio* (desde o início): seus modelos gerados podem ser interessantes já na primeira abordagem teórica. O modelo pode ser acessado em <https://modelarchive.org/doi/10.5452/ma-plm8v>

Uma das regiões estudadas na subunidade β do TSH é o *seat-belt*, presente entre a décima e décima segunda cisteína. No caso do *Arapaima gigas*, esta é uma região de 19 aminoácidos entre a ^{110th}C e ^{129th}C. Já no *Homo sapiens*, corresponde a região entre ^{88th}C e ^{105th}C. Representa uma identidade de 30% apenas entre as 40 espécies alinhadas apresentadas nessa dissertação. Entre a C^{10th}-C^{11th} (^{110th}C-^{117th}C), a sequência “DTGSHD” é majoritariamente carregada (50%) e entre a C^{11th}-C^{12th} (^{117th}C-^{129th}C) “THKGGDDNSAQ”

apresenta a mesma frequência para resíduos polares carregados e não carregados (36,4%). Quando comparamos a mesma região *seat-belt* para ag-βLH (Número de acesso GenBank: AJO68014.1) e ag-βFSH (Número de acesso GenBank: AIA09918.1) as cargas diferem em: I) FSH C^{10th}-C^{11th} com incidência maior de resíduos polares não carregados (66,4%) e C^{11th}-C^{12th} com prevalência de resíduos hidrofóbicos (55,6%), II) LH C^{10th}-C^{11th} com resíduos polares não carregados (50%) e C^{11th}-C^{12th} com resíduos hidrofóbicos (44,4%).

Os aminoácidos da cadeia α, R41, T45, M46, L47 e V48 são considerados os resíduos importantes “*hot spots*” para ligação com a subunidade beta. Na cadeia β, resíduos que são considerados importantes são: T96 e D100. A sequência em si apresenta similaridade de 68% com a sequência de *Homo sapiens*. Nesse caso, diferenças nas sequências das cadeias beta dos TSH humano e de *Arapaima gigas*, como por exemplo, I/T96 e I/D100 (humano/*A.gigas*), podem contribuir para a perda de afinidade do agTSH pelo hTSHR. Os resultados obtidos a partir das simulações de dinâmica molecular do complexo agTSH:hTSHR sugerem que o hormônio de *Arapaima gigas* não é capaz de interagir com o receptor humano de TSH.

Essa resolução do modelo e sua fidedignidade apresentam duas grandes vantagens com base nos resultados obtidos: a identidade da sequência de aminoácidos do hormônio quando alinhado com outras espécies (ver capítulo 4, tópico 4.2.1., tabela 4) e a metodologia extremamente nova para construção das estruturas proteicas (ver capítulo 2, tópico 2.3.4.1). Dada as suas vantagens de simulações *in silico*, podemos supor que o produto recombinante tem potencial para avaliação da atividade biológica, uma vez que os parâmetros obtidos na ligação agTSH + sfTSHR apresentam uma diferença mínima quando comparado ao modelo humano (hTSH + hTSHR).

5.3. Expressão, caracterização e purificação do agTSH

Por fim, com relação a expressão transiente, podemos afirmar que a metodologia foi eficiente na obtenção do hormônio recombinante. Como reportado anteriormente para o hormônio folículo estimulante de *Arapaima gigas* – agFSH - (SEVILHANO *et al.*, 2017), os níveis de rendimento foram similares: agTSH 25 mg/L vs agFSH 28 mg/L.

A escolha do vetor plasmidial pcDNA 3.4 (GenScript Biotech, Piscataway, NJ, EUA) aparenta ser uma ótima escolha na expressão de proteínas recombinantes, da qual inserimos a região codificante obtida *in silico* sem nenhuma otimização de códons. Tal ferramenta (sem otimização de códons) pode ser testada com o objetivo de melhorar os níveis de expressão,

podendo ser uma opção de estratégia para ser utilizada nos trabalhos futuros com essas ou outras proteínas de *Arapaima gigas*.

Com relação aos trabalhos do grupo de pesquisa reportando a expressão do TSH humano, o valor obtido no presente trabalho (25mg/L) foi baixo, mas, em relação às culturas com células aderidas, a produtividade foi alta. Seguindo as preparações recombinantes do laboratório, temos: 1,4 mg/L (DAMIANI *et al.*,2009), 0,6 mg/L (OLIVEIRA *et al.*,2007), $3,2 \pm 0,64$ mg/L (MENDONÇA *et al.*,2005) e 17,8 mg/L (PERONI *et al.*,2002). Vale ressaltar que os valores citados são apenas produções em placas – excepcionalmente Peroni e colaboradores (2002) em biorreator – que acomodam no máximo 10 mL e, houveram produções que chegaram a 10L de meio condicionado. O estudo com agTSH foi feito em apenas uma produção em suspensão, com volume final de 30 mL.

Ainda assim, necessita ser avaliado as metodologias de caracterização e purificação, uma vez que o *background* de proteínas não desejadas – visualizadas no cromatograma – dificultam a identificação do heterodímero. Uma dissociação das subunidades foi relatada e visualizada em SDS-PAGE. Esse relato já foi evidenciado por outro trabalho do grupo (CARVALHO *et al.*, 2009) com o TSH humano que, apesar também da dissociação, a reassociação das subunidades consegue manter a atividade biológica da proteína, sendo também uma possível estratégia para que futuros ensaios possam ser realizados.

Uma possibilidade evitar a desassociação das subunidades seria adequar as metodologias de purificação para um sistema menos agressivo, como a troca iônica seguida de exclusão molecular.

Visando a aplicação *in vivo* do hormônio, uma série de ensaios são necessários antes de um escalonamento na produção, os quais não foram possíveis realizar no período delimitado. Estes ensaios são de cunho imunológico e biológico (*in vitro*), como ELISA (*Enzyme-Linked Immunosorbent Assay*), *western blotting* e até radioimunoensaio (RIE). Há descrita outras metodologias mais recentes que podem ser empregadas para mensurar a atividade biológica *in vitro* (SENDAK *et al.*,2002) e trabalhos que possam ser referência na avaliação *in vivo*, como já introduzido para o hormônio recombinante em *Oreochromis niloticus* (AIZEN *et al.*, 2007).

Conclusões

Considerando o objetivo principal proposto na dissertação, as metodologias aplicadas e os resultados obtidos, foi possível identificar a região codificante da cadeia β do agTSH e caracterizar *in silico* sua estrutura tridimensional. Além do mais, com as colaborações realizadas ao longo do percurso, desdobrou-se as possibilidades de investigar a qualidade do modelo e explorar propriedades físico-químicas, estruturais e de interação com receptor de *Scleropages formosus*, com proximidade de parâmetros similares aos obtidos com o modelo de TSH humano.

Sobre a sequência primária dos aminoácidos do ag- β TSH, as informações obtidas e compiladas demonstradas no percurso do trabalho foram submetidas ao banco de dados do NCBI com a metodologia descrita. Graças a essa informação foi possível realizar os ensaios relacionados à estrutura do heterodímero, utilizando metodologias atualizadas na construção de modelos tridimensionais.

Como mencionado durante a dissertação, o objetivo do trabalho não consistia na cristalização da proteína em si, mas entender seu comportamento com base na sequência dos aminoácidos. Isso foi possível e abre espaço para estudo de outras proteínas que venham a ser interessantes para desenvolvimento, conservação e preservação do *Arapaima gigas*.

Com relação à expressão transiente em células de mamífero, podemos considerar efetiva sua expressão - em um ensaio piloto - com base na caracterização e purificação por sistema de cromatografia de alta eficiência. A escolha da HEK293 foi de acordo com o planejado, e foi efetivo na expressão do agTSH. No entanto, é necessário realizar mais ensaios com diferentes proporções de plasmídeo ou mesmo utilizar outra linhagem celular para fins de comparação e otimização da obtenção do produto recombinante.

Dadas as limitações naturais no período da pandemia do covid-19 faz-se necessário a continuidade dos estudos com o agTSH: otimização da expressão, ensaios imunológicos, *in vitro* e *in vivo*.

Referências bibliográficas

- ACCORRONI, A.; CHIELLINI, G.; ORIGLIA, N. Effects of thyroid hormones and their metabolites on learning and memory in normal and pathological conditions. *Current Drug Metabolism*, v. 18, n. 3, p. 225-236, 2017.
- ALBERTS, B.; JOHNSON, A.; LEWIS, J.; MORGAN, D.; RAFF, M.; ROBERTS, K.; WALTER, P.; WILSON, J.; HUNT, T. *Biologia molecular da célula*. Artmed Editora, 2010.
- ANDERSON, R. J.; WENG, Z.; CAMPBELL, R. K.; JIANG, X. Main-chain conformational tendencies of amino acids. *Proteins*, v. 60, n. 4, p. 679-689, 2005.
- AIZEN, J.; KASUTO, H.; GOLAN, M. Tilapia Follicle-Stimulating Hormone (FSH): Immunochemistry, Stimulation by Gonadotropin-Releasing Hormone, and Effect of Biologically Active Recombinant FSH on Steroid Secretion. *Biology of Reproduction*, v. 76, n. 4, p. 692-700, 2007. Disponível em: <<https://dx.doi.org/10.1095/biolreprod.106.055822>>.
- AL-SAMARAI, F. R.; AL-KAZAZ, A. A. Molecular markers: An introduction and applications. *European Journal of Molecular Biotechnology*, v. 9, n. 3, p. 118-130, 2015.
- ALMEIDA, B. E.; DAMIANI, R.; OLIVEIRA, J. E.; DALMORA, S. L.; TORJESEN, P. A.; BARTOLINI, P.; RIBELA, M. T. C. P. Reversed-phase high performance liquid chromatography as an alternative to animal bioassay for human thyrotropin potency determination. *Analytical Methods*, v. 6, n. 17, p. 6688-6694, 2014.
- AQUINO, M. R. A. *O cotidiano do trabalho em Cruzeiro do Sul a partir da obra do pintor de murais* Ary, 2012.
- BARTOLINI, P.; CARVALHO, R. F.; SEVILHANO, T. C. A.; OLIVEIRA, J. E.; GARCEZ, R. Molecular cloning, characterization and phylogenetic analysis of pirarucu (Arapaima gigas) FSH and LH beta-subunits. *Journal of Biotechnology*, v. 208, p. s17, 2015. Sup. S. 2015.
- BERND, M. *Pinturas de paisagem amazônica e a construção de um imaginário da cultura popular*. 2011. 189 p. Dissertação (Mestrado em Design e Arquitetura) - Faculdade de Arquitetura e Urbanismo, Universidade de São Paulo, São Paulo.
- BLUM M.; CHANG H. Y.; CHUGURANSKY, S.; GREGO, T.; KANDASAAMY, S.; MITCHELL, A.; NUKA, G.; PAYSAN-LAFOSSE, T.; QURESHI, M.; RAJ, S.; RICHARDSON, L.; SALAZAR, G. A.; WILLIAMS, L.; BORK, P.; BRIDGE, A.; GOUGH, J.; HAFT, D. H.; LETUNIC, I.; MARCHLER-BAUER, A.; MI, H.; NATALE, D. A.; NECCI, M.; ORENGO, C. A.; PANDURANGAN, A. P.; RIVOIRE, C.; SIGRIST, C. J. A.; SILLITOE, I.; THANKI, N.; THOMAS, P. D.; TOSATTO, S. C. E.; WU, C. H.; BATEMAN, A.; FINN, R.D. The InterPro protein families and domains database: 20 years on. *Nucleic Acids Research*, v. 49, n. D1, p. D344-D354, 2021.
- BRYANT, E. M.; MILLICHAP, J. J.; SPINELLI, E. CALHOUN, J. D.; MILLER, C.; GIANNELLI, J.; WOLAK, J.; SANDERS, V.; CARVILL, G. L.; CHARROW, J. Oligosaccharyltransferase complex-congenital disorders of glycosylation: A novel congenital disorder of glycosylation. *American Journal of Medical Genetics, Part A*, v. 182, n. 6, p. 1460-1465, 2020. DOI: 10.1002/ajmg.a.61553.
- BUTLER, M. Optimization of the cellular metabolism of glycosylation for recombinant proteins produced by mammalian cell systems. *Cytotechnology*, v. 50, n. 1-3, p. 57-76, 2006. DOI: 10.1007/s10616-005-4537-x.
- CAMPOS-SILVA, J. V.; HAWES, J. E.; PERES, C. A. Population recovery, seasonal site fidelity, and daily activity of pirarucu (Arapaima spp.) in an Amazonian floodplain mosaic. *Freshwater Biology*, v. 64, n. 7, p. 1255-1264, 2019.
- CARVALHO, C. M.; OLIVEIRA, J. E.; ALMEIDA, B. E.; UEDA, E. K.; TORJESEN, P. A.; BARTOLINI, P.; RIBELA, M. T. Efficient isolation of the subunits of recombinant and pituitary glycoprotein hormones. *Journal of Chromatography A*, v. 1216, n. 9, p. 1431-1438, 2009.

- CHEN, Y.; YE, W.; ZHANG, Y.; XU, Y. High speed BLASTN: an accelerated MegaBLAST search tool. *Nucleic Acids Research*, v. 43, n. 16, p. 7762-7768, 2015.
- CHEN, C.; HUANG, H.; WU, C. H. Protein bioinformatics databases and resources. *Methods in Molecular Biology*, v. 1558, p. 3-39, 2017. DOI:10.1007/978-1-4939-6783-4_1
- COUSSEN, F.; AYON, A.; LE GOFF, A.; LEROY, J.; MASSOULIÉ, J.; BON, S. Addition of a glycoposphatidylinositol to acetylcholinesterase: Processing, degradation, and secretion. *Journal of Biological Chemistry*, v. 276, n. 30, p. 27881–27892, 2001. DOI: 10.1074/jbc.M010817200.
- COMEAU, S. R.; GATCHELL, D. W.; VAJDA, S.; CAMACHO, C. J. ClusPro: a fully automated algorithm for protein–protein docking. *Nucleic Acids Research*, v. 32, n. suppl_2, p. W96-W99, 2004.
- CYR, D. G.; EALES, J. G. Interrelationships between thyroidal and reproductive endocrine systems in fish. *Reviews in Fish Biology and Fisheries*, v. 6, n. 2, p. 165-200, 1996.
- DA SILVA, A. C. Cenas da vida amazônica: recepção crítica e representações de identidade na obra de José Veríssimo. In: Anais do Simpósio Linguagens e Identidades da/na Amazônia Sul-Occidental, n. 1, 2016. *Resumos ...*
- DAMIANI, R.; ALMEIDA, B. E.; OLIVEIRA, J. E.; BARTOLINI, P.; RIBELA, M. T. Enhancement of human thyrotropin synthesis by sodium butyrate addition to serum-free CHO cell culture. *Applied Biochemistry and Biotechnology*, v. 171, n. 7, p. 1658-1672, 2013.
- DEAL, C. K.; VOLKOFF, H. The role of the thyroid axis in fish. *Frontiers in Endocrinology*, v. 11, p. 596585, 2020. DOI: 10.3389/fendo.2020.596585.
- DE AZEVEDO-BEMVENUTI, M.; FISCHER, L. G. Peixes: morfologia e adaptações. *Cadernos de Ecologia Aquática*, v. 5, n. 2, p. 31-54, 2010.
- DE BEER, T.; Vliegenthart, J. F. G.; LÖFFLER, A.; HOFSTEENGE, J. The Hexopyranosyl Residue That Is C-Glycosidically Linked to the Side Chain of Tryptophan-7 in Human RNase Us is α -Mannopyranose. *Biochemistry*, v. 34, n. 37, p. 11785–11789, 1995. DOI: 10.1021/bi00037a016.
- DELANO, W.L. Pymol: An open-source molecular graphics tool. CCP4 Newsletter *Protein Crystallography*, v. 40, n. 1, p. 82-92, 2002.
- DU, K.; WUERTZ, S.; ADOLFI, M.; KNEITZ, S.; STÖCK, M.; OLIVEIRA, M.; NÓBREGA, R.; ORMANN, J.; KLOAS, W.; FERON, R.; KLOPP, C.; PARRINELLO, H.; JOURNOT, L.; HE, S.; POSTLETHWAIT, J.; MEYER, A.; GUIGUEN, Y.; SCHARTL, M. The genome of the arapaima (*Arapaima gigas*) provides insights into gigantism, fast growth, and chromosomal sex determination system. *Scientific Reports*, v. 9, n. 1, p. 1-11, 2019.
- DUVAUD, S.; GABELLA, C.; LISACEK, F.; STOCKINGER, H.; IOANNIDIS, V.; DURINX, C. Expasy, the Swiss Bioinformatics Resource Portal, as designed by its users. *Nucleic Acids Research*, v. 49, n. W1, p. W216-W227, 2021. DOI: 10.1093/nar/gkab225.
- FANG, Y. Gibbs free energy formula for protein folding. *Thermodynamics—Fundamentals and Its Application in Science*, Intech, Rijeka, p. 47-82, 2012.
- FARIA, M. T.; CARVALHO, R. F.; SEVILHANO, T. C.; OLIVEIRA, N. A.; SILVA, C. F.; OLIVEIRA, J. E.; SOARES, C. R.; GARCEZ, R.; SANTO, P. R.; BARTOLINI, P. Isolation of the pituitary gonadotrophic α -subunit hormone of the giant Amazonian fish: pirarucu (*Arapaima gigas*). *Fish Physiology and Biochemistry*, v. 39, n. 3, p. 683-693, 2013.
- FARIAS, I. P.; LEÃO, A.; ALMEIDA, Y. S.; VERBA, J. T.; CROSSA, M. C.; HONCZARYK, A.; HRBEK, T. Evidence of polygamy in the socially monogamous Amazonian fish *Arapaima gigas* (Schinz, 1822) (Osteoglossiformes, Arapaimidae). *Neotropical Ichthyology*, v. 13, p. 195-204, 2015.

FERREIRA, G.; MARCOVITCH, J.; VAL, A. L. A systematic review of the production chain of the Arapaima gigas, the giant fish of the Amazon. *Management of Environmental Quality: An International Journal*, 2020.

FORTIN, J.; LAMBA, P.; WANG, Y.; BERNARD, D. J. Conservation of mechanisms mediating gonadotrophin-releasing hormone 1 stimulation of human luteinizing hormone β subunit transcription. *Molecular Human Reproduction*, v. 15, n. 2, p. 77-87, 2009.

GISMENE, C.; HERNÁNDEZ GONZÁLEZ, J.E.; SANTISTEBAN, A.R.N.; ZIEM NASCIMENTO, A.F.; DOS SANTOS CUNHA, L.; DE MORAES, F.R.; DE OLIVEIRA, C.L.P.; OLIVEIRA, C.C.; JOCELAN SCARIN PROVAZZI, P.; PASCUTTI, P.G.; ARNI, R.K.; BARROS MARIUTTI, R. Staphylococcus aureus Exfoliative Toxin E, Oligomeric State and Flip of P186: Implications for Its Action Mechanism. *International Journal of Molecular Sciences*, v. 23, p. 9857, 2022.

GREENWOOD, M. P.; FLIK, G.; WAGNER, G. F.; BALMENT, R. J. The corpuscles of Stannius, calcium-sensing receptor, and stanniocalcin: responses to calcimimetics and physiological challenges. *Endocrinology*, v. 150, n. 7, p. 3002-3010, 2009.

GURDAK, D. J.; STEWART, D. J.; CASTELLO, L.; ARANTES, C. C. Diversity in reproductive traits of arapaima (Arapaima spp., Müller, 1843) in Amazonian várzea floodplains: conservation implications. *Aquatic Conservation: Marine and Freshwater Ecosystems*, v. 29, n. 2, p. 245-257, 2019.

HABIBI, H. R.; NELSON, E. R.; ALLAN, E. R. O. New insights into thyroid hormone function and modulation of reproduction in goldfish. *General and comparative endocrinology*, v. 175, n. 1, p. 19-26, 2012.

HALL, T. BioEdit: an important software for molecular biology. *GERF Bulletin of Biosciences*, v. 2, n. 1, p. 60-61, 2011.

HAGEMANS, D.; BELZEN, I.A.E.M.V.; LUENGO, T.M.; RÜDIGER, S.G.D. A script to highlight hydrophobicity and charge on protein surfaces. *Frontiers in Molecular Biosciences*, v. 2, 13 out. 2015.

HAYNES, P. A. Phosphoglycosylation: A new structural class of glycosylation? *Glycobiology*, v. 8, n. 1, p. 1-5, 1998. DOI: 10.1093/glycob/8.1.1.

HELLER, S. R.; RODRIGUES-GOULART, H.; ARTHUSO, F. S.; OLIVEIRA, T. L.; BARTOLINI, P.; SOARES, C. R. Synthesis, purification and characterization of recombinant glycosylated human prolactin (G-hPRL) secreted by cycloheximide-treated CHO cells. *Journal of Biotechnology*, v. 145, n. 4, p. 334-340, 2010.

HOGEWEG, P. The roots of bioinformatics in theoretical biology. *PLoS Computational Biology*, v. 7, n. 3, p. e1002021, 2011.

HORIE, Y.; NOMURA, M.; OKAMOTO, K.; TAKAHASHI, C.; SATO, T.; MIYAGAWA, S.; OKAMURA, H.; IGUCHI, T. Effect of thyroid hormone-disrupting chemicals on swim bladder inflation and thyroid hormone-related gene expression in Japanese medaka and zebrafish. *Journal of Applied Toxicology*, v.42, n. 8, p. 1385-1395,2022.

JAOKEN, J. Congenital disorders of glycosylation. *Annals of the New York Academy of Sciences*, v. 1214, n. 1, p. 190-198, 2010. DOI: 10.1111/j.1749-6632.2010.05840.x.

JAHAGIRDAR, V.; MCNAY, E. C. Thyroid hormone's role in regulating brain glucose metabolism and potentially modulating hippocampal cognitive processes. *Metabolic Brain Disease*, v. 27, n. 2, p. 101-111, 2012.

KENDREW, J.C.; DICKERSON, R. E.; STRANDBERG, B. E.; HART, R. G.; DAVIES, D. R.; PHILLIPS, D. C.; SHORE, V. C. Structure of myoglobin: A three-dimensional Fourier synthesis at 2 Å. resolution. *Nature*, v. 185, n. 4711, p. 422-427, 1960.

JONES, D. T.; TAYLOR, W. R.; THORNTON, J. M. The rapid generation of mutation data matrices from protein sequences. *Bioinformatics*, v. 8, n. 3, p. 275-282, 1992.

- KAMEI, H.; OHIRA, T.; YOSHIURA, Y.; UCHIDA, N.; NAGASAWA, H.; AIDA, K. Expression of a biologically active recombinant follicle stimulating hormone of Japanese Eel *Anguilla japonica* using methylotropic yeast, *Pichia pastoris*. *General and Comparative Endocrinology*, v. 134, n. 3, p. 244–254, 2003. DOI: 10.1016/S0016-6480(03)00259-4.
- KANNAN, L.; WHEELER, W. C. Maximum parsimony on phylogenetic networks. *Algorithms for Molecular Biology*, v. 7, n. 1, p. 9, 2012. Disponível em: <<https://dx.doi.org/10.1186/1748-7188-7-9>>.
- KAPUSTIN, Y.; SOUVOROV, A.; TATUSOVA, T.; LIPMAN, D. Splign: algorithms for computing spliced alignments with identification of paralogs. *Biology Direct*, v. 3, n. 1, p. 1-13, 2008. DOI: 10.1186/1745-6150-3-20.
- KASUTO, H.; LEVAVI-SIVAN, B. Production of biologically active tethered tilapia LH β by the methylotropic yeast *Pichia pastoris*. *General and Comparative Endocrinology*, v. 140, n. 3, p. 222–232, 2005. DOI: 10.1016/j.ygcen.2004.10.016.
- KAZETO, Y.; KOHARA, M.; MIURA, T.; MIURA, C.; YAMAGUCHI, S.; TRANT, J. M.; ADACHI, S.; YAMAUCHI, K. Japanese Eel Follicle-Stimulating Hormone (FSH) and Luteinizing Hormone (LH): Production of Biologically Active Recombinant Fsh and Lh by Drosophila S2 Cells and Their Differential Actions on the Reproductive Biology1. *Biology of Reproduction*, v. 79, n. 5, p. 938–946, 2008. DOI: 10.1095/biolreprod.108.070052.
- KEARSE, M.; MOIR, R.; WILSON, A.; STONES-HAVAS, S.; CHEUNG, M.; STURROCK, S.; BUXTON, S.; COOPER, A.; MARKOWITZ, S.; DURAN, C.; THIERER, T.; ASHTON, B.; MEINTJES, P.; DRUMMOND, A. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, v. 28, n. 12, p. 1647-1649, 2012.
- KIRYUTIN, B.; SOUVOROV, A.; TATUSOVA, T. ProSplign—Protein to Genomic Alignment Tool. In: Proc. 11th Annual International Conference in Research in Computational Molecular Biology, San Francisco, USA. 2007. *Proceedings...*
- KNUDSEN, M.; WIUF, C. The CATH database. *Human Genomics*, v. 4, n. 3, p. 207, 2010.
- KUMAR, S.; STECHER, G.; LI, M.; KNYAZ, C.; TAMURA, K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, v. 35, n. 6, p. 1547, 2018.
- LASKOWSKI, R. A.; JABŁOŃSKA, J.; PRAVDA, L.; VAŘEKOVÁ, R. S.; THORNTON, J. M. PDBsum: Structural summaries of PDB entries. *Protein Science*, v. 27, n. 1, p. 129–134, 2018.
- LATRUBESSE, E. M.; ARIMA, E. Y.; DUNNE, T.; PARK, E.; BAKER, V. R.; D'HORTA, F. M.; WIGHT, C.; WITTMANN, F.; ZUANON, J.; BAKER, P. A.; RIBAS, C. C.; NORGAARD, R. B.; FILIZOLA, N.; ANSAR, A.; FLYVBJERG, B.; STEVAUX, J. C.. Damming the rivers of the amazon basin. *Nature*, v. 546, n. 7658, p. 363–369, 2017. Disponível em: <<https://dx.doi.org/10.1038/nature22333>>.
- LI, G.; FANG, C.; LI, Y.; WANG, Z.; SUN, S.; HE, S.; QI, W.; BAO, C.; MA, H.; FAN, Y.; FENG, Y.; LIU, X.. Global impacts of future urban expansion on terrestrial vertebrate diversity. *Nature Communications*, v. 13, n. 1, 2022.
- LIMA, A. F. The influence of sex ratio on the reproduction of pirarucu, *Arapaima gigas*, in captivity. *Acta Amazonica*, v. 48, p. 38-41, 2018.
- LIS, H.; SHARON, N. Protein glycosylation: Structural and functional aspects. *European Journal of Biochemistry*, v. 218, n. 1, p. 1–27, 1993. DOI: 10.1111/j.1432-1033.1993.tb18347.x.
- MA, Y.; LADISA, C.; CHANG, J. P.; HABIBI, H. R. Seasonal related multifactorial control of pituitary gonadotropin and growth hormone in female goldfish: influences of neuropeptides and thyroid hormone. *Frontiers in Endocrinology*, v. 11, p. 175, 2020.
- MANN, M.; JENSEN, O. N. Proteomic analysis of post-translational modifications. *Nature Biotechnology*, v. 21, n. 3, p. 255–261, 2003. DOI: 10.1038/nbt0303-255.

MENDONÇA, F.; OLIVEIRA, J. E.; BARTOLINI, P.; RIBELA, M. T. Two-step chromatographic purification of recombinant human thyrotrophin and its immunological, biological, physico-chemical and mass spectral characterization. *Journal of Chromatography A*, v. 1062, n. 1, p. 103-112, 2005.

MERONI, S.B.; GALARDO, M.N.; RINDONE, G.; GORGA, A.; RIERA, M.F.; CIGORRAGA, S.B. Molecular mechanisms and signaling pathways involved in sertoli cell proliferation. *Frontiers in endocrinology*, v. 10, p. 224, 2019.

MIRDITA, M.; SCHÜTZE, K.; MORIWAKI, Y.; HEO, L.; OVCHINNIKOV, S.; STEINEGGER, M. Colabfold: making protein folding accessible to all. *Nature Methods*, v. 19, n. 6, p. 679–682, 2022.

MONTEIRO, L. B. B.; SOARES, M. C.; CATANHO, M. T. J.; HONCZARYK, A. Aspectos reprodutivos e perfil hormonal dos esteróides sexuais do pirarucu, Arapaima gigas (SCHINZ, 1822), em condições de cativeiro. *Acta Amazônica*, v. 40, p. 435-449, 2010.

NATHANS, D.; SMITH, H. O. Restriction endonucleases in the analysis and restructuring of DNA molecules. *Annual Review of Biochemistry*, v. 44, n. 1, p. 273-293, 1975.

NEI, M.; KUMAR, S. *Molecular evolution and phylogenetics*. Oxford University Press, USA, 2000.

NELSON, D. L.; COX, M. M. *Princípios de Bioquímica de Lehninger*, 7 ed. Artmed Editora, 2018.

NOGUEIRA, F.; REGO, P. S.; QUEIROZ, H.; VENERE, P.; VARELA, E. S.; SAMPAIO, I.; SCHNEIDER, H.; ARARIPE, J. Genetic diversity and structuring in the arapaima (Osteoglossiformes, Osteoglossidae) population reveal differences between the Amazon and the Tocantins-Araguaia basins. *Anais da Academia Brasileira de Ciências*, v. 92, 2020. DOI.org/10.1590/0001-3765202020180496.

NUNES, M.T. Hormônios tiroideanos: mecanismo de ação e importância biológica. *Arquivos Brasileiros de Endocrinologia & Metabologia*, v. 47, n. 6, p. 639-643, 2003.

NÚÑEZ, J.; CHU-KOO, F.; BERLAND, M.; ARÉVALO, L.; RIBERYRO, O.; DUPONCHELLE, F.; RENNO, J.F. Reproductive success and fry production of the paiche or pirarucu, Arapaima gigas (Schinz), in the region of Iquitos, Perú. *Aquaculture Research*, v. 42, n. 6, p. 815-822, 2011.

OKONECHNIKOV, K.; GOLOSOVA, O.; FURSOV, M.; UGENE TEAM. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*, v. 28, n. 8, p. 1166-1167, 2012.

OLIVEIRA, J. E.; MENDONÇA, F.; PERONI, C. N.; BARTOLINI, P.; RIBELA, M. T. Determination of Chinese hamster ovary cell-derived recombinant thyrotropin by reversed-phase liquid chromatography. *Journal of Chromatography B*, v. 787, n. 2, p. 345-355, 2003.

OLIVEIRA, J. E.; DAMIANI, R.; VORAUER-UHL, K.; BARTOLINI, P.; RIBELA, M.T. Influence of a reduced CO₂ environment on the secretion yield, potency and N-glycan structures of recombinant thyrotropin from CHO cells. *Molecular Biotechnology*, v. 39, n. 2, p. 159-166, 2008.

PAIT, A. S.; NELSON, J. O. *Endocrine disruption in fish: an assessment of recent research and results*, NOAA Tech. Memo. NOS NCCOS CCMA 149. Silver Spring, MD: NOAA, NOS, Center for Coastal Monitoring and Assessment, 2002.

PARK, J. I.; SEMYONOV, J.; CHANG, C. L.; HSU, S. Y. Conservation of the heterodimeric glycoprotein hormone subunit family proteins and the LGR signaling system from nematodes to humans. *Endocrine*, v. 26, n. 3, p. 267-276, 2005.

PERONI, C. N.; SOARES, C. R.; GIMBO, E.; MORGANTI, L.; RIBELA, M. T.; BARTOLINI, P. High-level expression of human thyroid-stimulating hormone in Chinese hamster ovary cells by co-transfection of dicistronic expression vectors followed by a dual-marker amplification strategy. *Biotechnology and Applied Biochemistry*, v. 35, n. 1, p. 19-26, 2002.

POWER, D. M.; LLEWELLYN, L.; FAUSTINO, M.; NOWELL, M. A.; BJÖRNSSON, B. T.; EINARSDOTTIR, I. E.; CANARIO, A. V.; SWEENEY, G. E. Thyroid hormones in growth and development of fish. *Comparative Biochemistry and Physiology Part C: Toxicology and Pharmacology: CBP*, v. 130, n. 4, p. 447-459, 2001.

RAJPAL, D. K. Understanding biology through bioinformatics. *International Journal of Toxicology*, v. 24, n. 3, p. 147-152, 2005.

RIBELA, M. T.; DAMIANI, R.; SILVA, F. D.; LIMA, E. R.; OLIVEIRA, J. E.; PERONI, C. N.; TORJESEN, P. A.; SOARES, C. R.; BARTOLINI, P. N-Glycoprofiling analysis for carbohydrate composition and site-occupancy determination in a poly-glycosylated protein: human thyrotropin of different origins. *International Journal of Molecular Sciences*, v. 18, n. 2, p. 131, 2017.

SANDERS, J.; CHIRGADZE, D. Y.; SANDERS, P.; BAKER, S.; SULLIVAN, A.; BHARDWAJA, A.; BOLTON, J.; REEVE, M.; NAKATAKE, N.; EVANS, M.; RICHARDS, T.; POWELL, M.; MIGUEL, R. N.; BLUNDELL, T. L.; FURMANIAK, J.; SMITH, B. R. Crystal structure of the TSH receptor in complex with a thyroid-stimulating autoantibody. *Thyroid*, v. 17, n. 5, p. 395-410, 2007.

SANT'ANA, P. M.; OLIVEIRA, J. E.; LIMA, E. R.; SOARES, C. R. J.; PERONI, C. N.; BARTOLINI, P.; RIBELA, M. T. C. P. Human thyroid-stimulating hormone synthesis in human embryonic kidney cells and related N-glycoprofiling analysis for carbohydrate composition determination. *Applied Microbiology and Biotechnology*, v. 102, n. 3, p. 1215-1228, 2018.

SIEVERS, F.; WILM, A.; DINEEN, D.; GIBSON, T. J.; KARPLUS, K.; LI, W.; LOPEZ, R.; MCWILLIAM, H.; REMMERT, M.; SÖDING, J.; THOMPSON, J. D.; HIGGINS, D. G. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, v. 7, n. 1, p. 539, 2011.

SILVA, J. F.; OCARINO, N. M.; SERAKIDES, R. Thyroid hormones and female reproduction. *Biology of reproduction*, v. 99, n. 5, p. 907-921, 2018.

SCHWARZ, F.; AEBI, M. Mechanisms and principles of N-linked protein glycosylation. *Current Opinion in Structural Biology*, [S. l.], v. 21, n. 5, p. 576-582, 2011. DOI: 10.1016/j.sbi.2011.08.005. Disponível em: <http://dx.doi.org/10.1016/j.sbi.2011.08.005>.

SENDAK, R. A.; WANG, F.; GEAGAN, L. B.; ARMSTRONG, L. A.; THYNE, C. D.; COLE, E. S.; MATTALIANO, R. J. Comparison of two in vitro methods for the measurement of recombinant human TSH bioactivity. *Biologicals*, v. 30, n. 3, p. 245-254, 2002.

SEVILHANO, T.; CARVALHO, R. F.; OLIVEIRA, N. A. J.; OLIVEIRA, J. E.; MALTAROLLO, V. G.; TROSSINI, G.; GARCEZ, R.; BARTOLINI, P. Molecular cloning and characterization of pirarucu (*Arapaima gigas*) follicle-stimulating hormone and luteinizing hormone β -subunit cDNAs. *PloS One*, v. 12, n. 8, p. e0183545, 2017.

SHARMA, K.; MISHRA, A. K.; MEHRAJ, V.; DURAISAMY, G. S. Advances and applications of molecular cloning in clinical microbiology. *Biotechnology and Genetic Engineering Reviews*, v. 30, n. 1, p. 65-78, 2014.

SILVA, F. D.; OLIVEIRA, J. E.; FREIRE, R. P.; SUZUKI, M. F.; SOARES, C. R.; BARTOLINI, P. Expression of glycosylated human prolactin in HEK293 cells and related N-glycan composition analysis. *AMB Express*, v. 9, n. 1, p. 1-11, 2019.

SILVA, J. F.; OCARINO, N. M.; SERAKIDES, R. Thyroid hormones and female reproduction. *Biology of Reproduction*, v. 99, n. 5, p. 907-921, 2018.

SOARES, C. R.; MORGANTI, L.; MILOUX, B.; LUPKER, J. H.; FERRARA, P.; BARTOLINI, P. High-level synthesis of human prolactin in Chinese-hamster ovary cells. *Biotechnology and Applied Biochemistry*, v. 32, n. 2, p. 127-135, 2000.

- SOUVOROV, A.; KAPUSTIN, Y.; KIRYUTIN, B.; CHETVERNIN, V.; TATUSOVA, T.; LIPMAN, D. Gnomon–NCBI eukaryotic gene prediction tool. *National Center for Biotechnology Information*, p. 1-24, 2010.
- STANLEY, P.; TANIGUCHI, N.; AEBI, M. N-glycans. In: *Essentials of Glycobiology [Internet]*. 3 ed. Cold Spring Harbor Laboratory Press, 2017.
- SZKUDLINSKI, M. W.; THE, N. G.; GROSSMANN, M.; TROPEA, J. E.; WEINTRAUB, B. D. Engineering human glycoprotein hormone superactive analogues. *Nature Biotechnology*, v. 14, n. 10, p. 1257-1263, 1996.
- SZKUDLINSKI, M. W.; FREMONT, V.; RONIN, C.; WEINTRAUB, B. D. Thyroid-stimulating hormone and thyroid-stimulating hormone receptor structure-function relationships. *Physiological Reviews*, v. 82, n. 2, p. 473-502, 2002.
- TAMURA, K.; STECHER, G.; KUMAR, S. MEGA11: molecular evolutionary genetics analysis version 11. *Molecular Biology and Evolution*, v. 38, n. 7, p. 3022-3027, 2021.
- TEUFEL, F.; ALMAGRO-ARMENTEROS, J. J.; JOHANSEN, A. R.; GÍSLASON, M. H.; PIHL, S. I.; TSIRIGOS, K. D.; WINTHER, O.; BRUNAK, S.; VON HEIJNE, G.; NIELSEN, H. SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nature Biotechnology*, p. 1-3, 2022.
- TIRABASSI, R.; BIO, B. *Foundations of Molecular Cloning-Past, Present and Future*. 2014.
- TONUS, J. L. Samuel Rawet e o teatro da transgressão: uma leitura das peças inéditas Os amantes e A farsa da pesca do pirarucu e da caçada do jacu. *Estudos de Literatura Brasileira Contemporânea*, n. 23, p. 89-113, 2004.
- TORATI, L. S.; TAGGART, J. B.; VARELA, E. S.; ARARIPE, J.; WEHNER, S.; MIGAUD, H. Genetic diversity and structure in Arapaima gigas populations from Amazon and Araguaia-Tocantins river basins. *BMC Genetics*, v. 20, n. 1, p. 1-13, 2019.
- TOSHIMA, K.; TATSUTA, K. Recent Progress in O-Glycosylation Methods and Its Application to Natural Products Synthesis. *Chemical Reviews*, [S. l.], v. 93, n. 4, p. 1503–1531, 1993. DOI: 10.1021/cr00020a006.
- TOVO-NETO, A.; SILVA-RODRIGUES, M.; HABIBI, H. R.; NÓBREGA, R. H. Thyroid hormone actions on male reproductive system of teleost fish. *General and Comparative Endocrinology*, v. 265, p. 230-236, 2018.
- UNNI, S.; HUANG, Y.; HANSON, R. M.; TOBIAS, M.; KRISHNAN, S.; LI, W. W.; NIELSEN, J. E.; BAKER, N. A. Web servers and services for electrostatics calculations with APBS and PDB2PQR. *Journal of Computational Chemistry*, v. 32, n. 7, p. 1488-1491, 2011.
- VALGÔDE, F. G. S.; DA SILVA, M. A.; VIEIRA, D. P.; RIBELA, M. T. C. P.; BARTOLINI, P.; OKAZAKI, K. Cytotoxic and genotoxic effects of ¹³¹I and ⁶⁰Co in follicular thyroid cancer cell (WRO) with and without recombinant human thyroid-stimulating hormone treatment. *Environmental and Molecular Mutagenesis*, v. 58, n. 6, p. 451-461, 2017.
- VARKI, A. Biological roles of oligosaccharides: all the theories are correct. *Glycobiology*, v. 3, n. 2, p. 97-130, 1993.
- VASWANI, A.; SHAZEER, N.; PARMAR, N.; USZKOREIT, J.; JONES, L.; GOMEZ, A. N.; KAISER, Ł.; POLOSUKHIN, I. Attention is all you need. *Advances in Neural Information Processing Systems*, v. 30, 2017.
- VENTURIERI, R.; BERNARDINO, G. Pirarucu, espécie ameaçada, pode ser salvo através do cultivo. *Revista Panorama da Aquicultura*, v. 9, 1999.
- VERÍSSIMO, J. Cenas da vida amazônica. Com um estudo sobre as populações indígenas e mestiças da Amazônia. *Primeiro livro (único publicado)*. Lisboa, Tavares Cardoso, 1886.

VIALLE, R. A.; DE SOUZA, J. E. S.; LOPES, K. P.; TEIXEIRA, D. G.; ALVES-SOBRINHO, P. A.; RIBEIRO-DOS-SANTOS, A. M.; FURTADO, C.; SAKAMOTO, T.; OLIVEIRA-SILVA, F. A.; DE OLIVEIRA, E. H. C.; HAMOY, I. G.; ASSUMPCÃO, P. P.; RIBEIRO-DOS-SANTOS, Â.; SANTOS-LIMA, J. P. M.; SEUÁNEZ, H. N.; DE SOUZA, S. J.; SANTOS, S. Whole genome sequencing of the pirarucu (*Arapaima gigas*) supports independent emergence of major teleost clades. *Genome Biology and Evolution*, v. 10, n. 9, p. 2366-2379, 2018.

WALKER-NASIR, E.; AHMAD, I.; SALEEM, M.; HOESSLI, C. D. Glycosyltransferase and glypiation inhibitors. *Current Organic Chemistry*, v. 11, n. 7, p. 591-607, 2007.

WALTER, K. M.; MILLER, G. W.; CHEN, X.; YAGHOUBI, B.; PUSCHNER, B.; LEIN, P. J. Effects of thyroid hormone disruption on the ontogenetic expression of thyroid hormone signaling genes in developing zebrafish (*Danio rerio*). *General and Comparative Endocrinology*, v. 272, p. 20-32, 2019.

WATANABE, L.; GOMES, F.; VIANEZ, J.; NUNES, M.; CARDOSO, J.; LIMA, C.; SCHNEIDER, H.; SAMPAIO, I. De novo transcriptome based on next-generation sequencing reveals candidate genes with sex-specific expression in *Arapaima gigas* (Schinz, 1822), an ancient Amazonian freshwater fish. *PloS One*, v. 13, n. 10, p. e0206379, 2018.

WHELAN, S.; GOLDMAN, N. A General Empirical Model of Protein Evolution Derived from Multiple Protein Families Using a Maximum-Likelihood Approach. *Molecular Biology and Evolution*, v. 18, n. 5, p. 691-699, 2001.

WILKINSON, K. A.; HENLEY, J. M. Mechanisms, regulation and consequences of protein SUMOylation. *Biochemical Journal*, v. 428, n. 2, p. 133-145, 2010. DOI: 10.1042/BJ20100158.

YU, U.; LEE, S. H.; KIM, Y. J.; KIM, S. Bioinformatics in the post-genome era. *BMB Reports*, v. 37, n. 1, p. 75-82, 2004.

APÊNDICE A - Ranqueamento dos 56 modelos de substituição de aminoácidos para construção da árvore de máxima verossimilhança

Model	#Param	BIC	AICc	lnL	Invariant	Gamma	Freq A	Freq R	Freq N	Freq D	Freq C	Freq Q	Freq E
JTT+G	74	9823.231236	9333.919322	-4691.983997	n/a	1.20787499	0.0769	0.0611	0.0425	0.0513	0.0203	0.0411	0.0618
JTT+G+I	75	9827.411683	9331.523823	-4589.738523	0.064159282	1.88503915	0.0769	0.0511	0.0425	0.0513	0.0203	0.0411	0.0618
WAG+G	74	9848.05661	9456.743936	-4654.376154	n/a	1.22748454	0.06803729	0.043972	0.0390984	0.0570451	0.0193078	0.0367281	0.0590589
WAG+G+I	75	9855.344161	9456.447202	-4652.200612	0.055336849	1.6254325	0.06803729	0.043972	0.0390984	0.0570451	0.0193078	0.0367281	0.0590589
JTT+I	74	9971.850412	9482.278488	-4686.143555	0.105043335	n/a	0.0769	0.0611	0.0425	0.0513	0.0203	0.0411	0.0618
Dayhoff+G	74	9890.232833	9490.910319	-4670.46245	n/a	1.22447893	0.0769	0.0611	0.0425	0.0513	0.0203	0.0411	0.0618
Dayhoff+G+I	75	9895.949992	9490.063032	-4669.003728	0.039868632	1.51906493	0.0769	0.0611	0.0425	0.0513	0.0203	0.0411	0.0618
JTT+G+I	75	10016.47683	9402.175222	-4608.514043	n/a	1.21868121	0.06803768	0.0527527	0.033457249	0.042485396	0.088209649	0.027438485	0.04496371
Dayhoff+G+I	74	9895.949992	9490.063032	-4669.003728	0.039868632	1.51906493	0.0769	0.0611	0.0425	0.0513	0.0203	0.0411	0.0618
JTT+G+I+H	84	10016.47683	9402.175222	-4608.514043	n/a	1.21868121	0.06803768	0.0527527	0.033457249	0.042485396	0.088209649	0.027438485	0.04496371
Dayhoff+G+I+H	74	9895.949992	9490.063032	-4669.003728	0.039868632	1.51906493	0.0769	0.0611	0.0425	0.0513	0.0203	0.0411	0.0618
LG+G	75	10022.52885	9526.631986	-4687.443159	0.096602637	1.880176728	0.0769	0.056941	0.041977	0.033052	0.012937	0.040787	0.071589
LG+G+I	74	10023.27112	9533.959202	-4681.883606	n/a	1.058304169	0.0769	0.056941	0.041977	0.033052	0.012937	0.040787	0.071589
JTT	73	10078.19079	9598.47386	-4724.287861	n/a	n/a	0.0769	0.0511	0.0425	0.0513	0.0203	0.0411	0.0618
Dayhoff+G+I	83	10057.70864	9473.462327	-4642.172446	n/a	1.24479658	0.06803768	0.0527527	0.033457249	0.042485396	0.088209649	0.027438485	0.04496371
Dayhoff+G+I+H	84	10093.12331	9472.251028	-4640.517654	0.040250727	1.540920723	0.06803768	0.0527527	0.033457249	0.042485396	0.088209649	0.027438485	0.04496371
WAG+G+I	83	10100.94081	9486.639497	-4648.746031	n/a	1.273892543	0.06803768	0.0527527	0.033457249	0.042485396	0.088209649	0.027438485	0.04496371
WAG+I	74	10104.33514	9516.232329	-4732.81562	0.104656881	n/a	0.06803768	0.0527527	0.033457249	0.042485396	0.088209649	0.027438485	0.04496371
WAG+G+I+H	84	10105.44513	9484.872846	-4648.878574	0.054076319	1.877654576	0.06803768	0.0527527	0.033457249	0.042485396	0.088209649	0.027438485	0.04496371
JTT+I+H	83	10159.26885	9545.087336	-4677.975051	0.110024237	n/a	0.06803768	0.0527527	0.033457249	0.042485396	0.088209649	0.027438485	0.04496371
Dayhoff+I	74	10176.69743	9587.885513	-4788.847052	0.076427623	n/a	0.06803768	0.0527527	0.033457249	0.042485396	0.088209649	0.027438485	0.04496371
LG+I	74	10177.81312	9488.501203	-4749.254027	0.113297165	n/a	0.0769	0.056941	0.041977	0.033052	0.012937	0.040787	0.071589
mREV+G	74	10197.11905	9707.807138	-4778.907875	n/a	0.748739759	0.0769	0.0631	0.041	0.0756	0.0091	0.0382	0.0495
mREV+G+I	75	10198.13515	9703.268191	-4775.598397	0.09552708	1.779200108	0.0646	0.0463	0.0376	0.0422	0.0114	0.0806	0.0507
WAG	73	10201.14069	9718.414069	-4785.236066	n/a	0.8862979	0.06803768	0.049972	0.0390984	0.0670451	0.0193078	0.0367281	0.0590589
mREV+G	74	10203.59223	9714.287317	-4782.147954	n/a	1.007844596	0.0646	0.0463	0.0376	0.0422	0.0114	0.0806	0.0507
LG+G+I+H	84	10206.46888	9486.594597	-4687.188448	0.087121278	1.700063316	0.06803768	0.0527527	0.033457249	0.042485396	0.088209649	0.027438485	0.04496371
LG+G+I	83	10207.23424	9592.932927	-4701.892746	n/a	1.057218919	0.06803768	0.0527527	0.033457249	0.042485396	0.088209649	0.027438485	0.04496371
mREV+G+I	75	10238.17672	9742.27976	-4785.117092	0.115384615	0.838454866	0.06803768	0.0527527	0.033457249	0.042485396	0.088209649	0.027438485	0.04496371
WAG+I+H	83	10248.31607	9594.014758	-4722.433601	0.102627878	n/a	0.06803768	0.0527527	0.033457249	0.042485396	0.088209649	0.027438485	0.04496371
Dayhoff	72	10252.33883	9779.612895	-4815.837379	n/a	0.0871	0.0409	0.0409	0.0404	0.0335	0.0335	0.0383	0.0495
JTT+I+H	82	10267.86504	9680.265438	-4736.592762	n/a	0.8862979	0.06803768	0.0527527	0.033457249	0.042485396	0.088209649	0.027438485	0.04496371
Dayhoff+I+H	83	10270.97447	9686.873167	-4733.765861	0.081756865	n/a	0.06803768	0.0527527	0.033457249	0.042485396	0.088209649	0.027438485	0.04496371
mREV+I+G+I+H	83	10328.65465	9714.65937	-4762.753101	n/a	0.855461405	0.06803768	0.0527527	0.033457249	0.042485396	0.088209649	0.027438485	0.04496371
mREV+I+G+I+H+I	84	10328.61942	9708.647143	-4768.715721	0.07687059	1.281486838	0.06803768	0.0527527	0.033457249	0.042485396	0.088209649	0.027438485	0.04496371
LG	73	10338.15203	9455.425904	-4853.743663	n/a	n/a	0.0769	0.056941	0.041977	0.033052	0.012937	0.040787	0.071589
WAG+I	82	10338.76329	9731.063891	-4771.991889	n/a	n/a	0.06803768	0.0527527	0.033457249	0.042485396	0.088209649	0.027438485	0.04496371
mREV+I+G+I	74	10346.63009	9657.327176	-4853.687694	n/a	0.801585209	0.072	0.019	0.039	0.019	0.008	0.025	0.024
mREV+I+G+I+H	84	10347.89812	9727.11684	-4767.95007	0.064168868	1.688661292	0.06803768	0.0527527	0.033457249	0.042485396	0.088209649	0.027438485	0.04496371
mREV+I+G+I+H+I	75	10349.77032	9653.873384	-4850.813894	0.066351549	1.082351902	0.072	0.019	0.039	0.019	0.008	0.025	0.024
Dayhoff+G+I	83	10351.60781	9737.206501	-4774.028533	n/a	0.878875497	0.06803768	0.0527527	0.033457249	0.042485396	0.088209649	0.027438485	0.04496371
Dayhoff+I	82	10353.36364	9745.633946	-4779.27016	n/a	n/a	0.06803768	0.0527527	0.033457249	0.042485396	0.088209649	0.027438485	0.04496371
mREV+I+H	74	10354.83897	9686.528058	-4857.786335	0.112888851	n/a	0.0646	0.0463	0.0376	0.0422	0.0114	0.0806	0.0507
LG+I+H	83	10362.85464	9748.053326	-4779.452346	0.112842585	n/a	0.06803768	0.0527527	0.033457249	0.042485396	0.088209649	0.027438485	0.04496371
mREV+I+G+I+H	83	10403.66839	9788.787078	-4819.808621	0.73321278	n/a	0.06803768	0.0527527	0.033457249	0.042485396	0.088209649	0.027438485	0.04496371
mREV+I+G+I+H+I	84	10434.47161	9813.599236	-4811.191758	0.115384615	n/a	0.06803768	0.0527527	0.033457249	0.042485396	0.088209649	0.027438485	0.04496371
cpREV+I+G+I+H	74	10449.13368	9698.821268	-4804.915619	0.882121896	n/a	0.06803768	0.0527527	0.033457249	0.042485396	0.088209649	0.027438485	0.04496371
mREV+I+H	83	10512.74397	9698.442663	-4854.547614	0.112754513	n/a	0.06803768	0.0527527	0.033457249	0.042485396	0.088209649	0.027438485	0.04496371
LG+I	82	10522.05034	9914.320747	-4893.820417	0.112754513	n/a	0.06803768	0.0527527	0.033457249	0.042485396	0.088209649	0.027438485	0.04496371
mREV	73	10549.75228	10087.02815	-4959.544106	n/a	n/a	0.06803768	0.0527527	0.033457249	0.042485396	0.088209649	0.027438485	0.04496371
mREV+I+G+I+H	83	10573.0801	9658.787894	-4884.81568	0.111770784	n/a	0.06803768	0.0527527	0.033457249	0.042485396	0.088209649	0.027438485	0.04496371
cpREV	73	10581.9003	10099.17717	-4975.812615	n/a	n/a	0.06803768	0.0527527	0.033457249	0.042485396	0.088209649	0.027438485	0.04496371
mREV+I+H	83	10650.68728	10016.29686	-4913.874258	0.115384615	n/a	0.06803768	0.0527527	0.033457249	0.042485396	0.088209649	0.027438485	0.04496371
mREV+I+H+I	84	10646.24754	10158.93663	-5003.47212	0.11127731	n/a	0.072	0.019	0.039	0.019	0.008	0.025	0.024
mREV+I+H+I+H	82	10708.85629	10100.6657	-4956.887891	n/a	n/a	0.06803768	0.0527527	0.033457249	0.042485396	0.088209649	0.027438485	0.04496371
mREV+I+H+I+H+I	82	10750.64118	10142.91158	-4977.915834	n/a	n/a	0.06803768	0.0527527	0.033457249	0.042485396	0.088209649	0.027438485	0.04496371
cpREV+I+H	82	10756.90369	10156.26400	-4986.003037	n/a	n/a	0.06803768	0.0527527	0.033457249	0.042485396	0.088209649	0.027438485	0.04496371
mREV+I+H+I+H+I+H	73	10815.95327	10332.68914	-5022.365603	n/a	n/a	0.072	0.019	0.039	0.019	0.008	0.025	0.024

APÊNDICE B – Ordem dos comandos em Python utilizados no AlphaFold

```
@title Install software
@markdown Please execute this cell by pressing the _Play_ button
@markdown on the left.

# setup device

import warnings

warnings.simplefilter(action='ignore', category=FutureWarning)

import os

import sys

import tensorflow as tf

import jax

try:

    # check if TPU is available

    import jax.tools.colab_tpu

    jax.tools.colab_tpu.setup_tpu()

    print('Running on TPU')

    DEVICE = "tpu"

except:

    if jax.local_devices()[0].platform == 'cpu':

        print("WARNING: no GPU detected, will be using CPU")

        DEVICE = "cpu"

    else:

        print('Running on GPU')

        DEVICE = "gpu"

        # disable GPU on tensorflow

        tf.config.set_visible_devices([], 'GPU')

from IPython.utils import io
```

```

import subprocess

import tqdm.notebook

install_jackhmmmer = True

GIT_REPO = 'https://github.com/deepmind/alphafold'

SOURCE_URL =
'https://storage.googleapis.com/alphafold/alphafold_params_2021-07-14.tar'

PARAMS_DIR = './alphafold/data/params'

PARAMS_PATH = os.path.join(PARAMS_DIR, os.path.basename(SOURCE_URL))

TQDM_BAR_FORMAT = '{l_bar}{bar}| {n_fmt}/{total_fmt} [elapsed: {elapsed} remaining: {remaining}]'

TMP_DIR = "tmp"

os.makedirs(TMP_DIR, exist_ok=True)

# if not already installed

total = 55

with tqdm.notebook.tqdm(total=total, bar_format=TQDM_BAR_FORMAT) as pbar:

    if not os.path.isdir("alphafold"):

        # download alphafold code

        os.system(f"git clone {GIT_REPO} alphafold; cd alphafold; git checkout 1d43aaff941c84dc56311076b58795797e49107b")

        os.system(f"git clone https://github.com/sokrypton/ColabFold.git")

        # apply patches

        os.system(f"patch -u alphafold/alphafold/model/model.py -i ColabFold/beta/model.patch")

        os.system(f"patch -u alphafold/alphafold/model/modules.py -i ColabFold/beta/modules.patch")

        os.system(f"patch -u alphafold/alphafold/model/folding.py -i ColabFold/beta/folding.patch")

        os.system(f"patch -u alphafold/alphafold/model/config.py -i ColabFold/beta/config.patch")

        # apply multi-chain patch from Lim Heo @huhlim

        os.system(f"patch -u alphafold/alphafold/common/protein.py -i ColabFold/beta/protein.patch")

```

```

pbar.update(4)

os.system(f"pip install biopython dm-haiku==0.0.5 ml-collections
py3Dmol")

pbar.update(6)

# download model params (speedup from kaczmarj)

os.system(f"mkdir --parents {PARAMS_DIR}")

os.system(f"curl -fsSL {SOURCE_URL} | tar x -C {PARAMS_DIR}")

pbar.update(14+27)

# install hhsuite

os.system(f"curl -fsSL https://github.com/soedinglab/hh-
suite/releases/download/v3.3.0/hhsuite-3.3.0-SSE2-Linux.tar.gz | tar
xz -C {TMP_DIR}/")

# install jackhmmer

if install_jackhmmer:

    os.system(f"sudo apt install --quiet --yes hmmer")

    pbar.update(3)

    # create a ramdisk to store a database chunk to make Jackhmmer
    run fast.

    os.system(f"sudo mkdir -m 777 --parents /tmp/ramdisk")

    os.system(f"sudo mount -t tmpfs -o size=9G ramdisk /tmp/ramdisk")

    pbar.update(1)

else:

    pbar.update(4)

else:

    pbar.update(55)

#####
#####

# --- Python imports ---

if 'alphafold' not in sys.path:

    sys.path.append('alphafold')

if 'ColabFold/beta' not in sys.path:

```

```

sys.path.append('ColabFold/beta')

if f"{TMP_DIR}/bin" not in os.environ['PATH']:
    os.environ['PATH'] += f":{TMP_DIR}/bin:{TMP_DIR}/scripts"

import colabfold as cf
import colabfold_alpha as cf_af

import json

import matplotlib.pyplot as plt

import numpy as np

try:
    from google.colab import files

    IN_COLAB = True
except:
    IN_COLAB = False

#@title Enter the amino acid sequence to fold 

import re

# define sequence

sequence = 'NFNVGCEECKLKENKYFSRLGAPIFQCMGCCFSRAYPTPLRSKKTMLVPGKNTSEATCCVAKEVKRLI
TLNNVRLNHTDCHCNTCYHKS:FCALTDYTLVEKAGCDFCVAINTTICMGFCYSRDTNIVGLAGKR
LLLQRTCTYRSVEYHTTALPGCHQHADPFFSYVALDCYCSTCDTGSHDCTHKGDDNSAQCAKPLLHI
YPYPGQSNHI' #@param {type:"string"}

jobname = "agtsh w/out sp" #@param {type:"string"}

homooligomer = "1" #@param {type:"string"}

#@markdown - `sequence` Specify protein sequence to be modelled.

#@markdown - Use `/` to specify intra-protein chainbreaks (for trimming
regions within protein).

#@markdown - Use `:` to specify inter-protein chainbreaks (for modeling
protein-protein hetero-complexes).

#@markdown - For example, sequence `AC/DE:FGH` will be modelled as
polypeptides: `AC`, `DE` and `FGH`. A separate MSA will be generated
for `ACDE` and `FGH`.

#@markdown - If `pair_msa` is enabled, `ACDE`'s MSA will be paired
with `FGH`'s MSA.

```

#@markdown - `homooligomer` Define number of copies in a homooligomeric assembly.

#@markdown - Use `:` to specify different homooligomeric state (copy number) for each component of the complex.

#@markdown - For example, `**sequence:**`ABC:DEF``, `**homooligomer:**`2:1``, the first protein `ABC` will be modeled as a homodimer (2 copies) and second `DEF` a monomer (1 copy).

```
I = cf_af.prep_inputs(sequence, jobname, homooligomer, clean=IN_COLAB)
```

```
#@title Search against genetic databases
```

#@markdown Once this cell has been executed, you will see

#@markdown statistics about the multiple sequence alignment

#@markdown (MSA) that will be used by AlphaFold. In particular,

#@markdown you'll see how well each residue is covered by similar

#@markdown sequences in the MSA.

#@markdown (Note that the search against databases and the actual prediction can take some time, from minutes to hours, depending on the length of the protein and what type of GPU you are allocated by Colab.)

```
#@markdown ---
```

```
msa_method = "jackhmmer" #@param
["mmseqs2","jackhmmer","single_sequence","precomputed"]
```

#@markdown - `mmseqs2` - FAST method from [ColabFold] (<https://github.com/sokrypton/ColabFold>)

#@markdown - `jackhmmer` - default method from Deepmind (SLOW, but may find more/less sequences).

#@markdown - `single_sequence` - use single sequence input

#@markdown - `precomputed` If you have previously run this notebook and saved the results,

#@markdown you can skip this step by uploading

#@markdown the previously generated `prediction_????/msa.pickle`

```
#@markdown ---
```

```
#@markdown **custom msa options**
```

```
add_custom_msa = False #@param {type:"boolean"}
```

```
msa_format = "fas" #@param ["fas","a2m","a3m","sto","psi","clu"]
```

#@markdown - `add_custom_msa` - If enabled, you'll get an option to upload your custom MSA in the specified `msa_format`. Note: Your MSA

will be supplemented with those from 'mmseqs2' or 'jackhmmer', unless `msa_method` is set to 'single_sequence'.

```

#@markdown ---

#@markdown pair msa options

#@markdown Experimental option for protein complexes. Pairing currently
only supported for proteins in same operon (prokaryotic genomes).

pair_mode          =          "unpaired+paired"          #@param
["unpaired","unpaired+paired","paired"] {type:"string"}

#@markdown - `unpaired` - generate separate MSA for each protein.

#@markdown - `unpaired+paired` - attempt to pair sequences from the
same operon within the genome.

#@markdown - `paired` - only use sequences that were successfully paired.

#@markdown Options to prefilter each MSA before pairing. (It might help
if there are any paralogs in the complex.)

pair_cov = 50 #@param [0,25,50,75,90] {type:"raw"}

pair_qid = 30 #@param [0,15,20,30,40,50] {type:"raw"}

#@markdown - `pair_cov` prefilter each MSA to minimum coverage with
query (%) before pairing.

#@markdown - `pair_qid` prefilter each MSA to minimum sequence identity
with query (%) before pairing.

# --- Search against genetic databases ---

I = cf_af.prep_msa(I, msa_method, add_custom_msa, msa_format,
                  pair_mode, pair_cov, pair_qid, TMP_DIR=TMP_DIR)

mod_I = I

if len(I["msas"][0]) > 1:

    plt = cf.plot_msas(I["msas"], I["ori_sequence"])

    plt.savefig(os.path.join(I["output_dir"],"msa_coverage.png"),
bbox_inches = 'tight', dpi=200)

    plt.show()

#@title Filter options (optional)

trim = "" #@param {type:"string"}

trim_inverse = False #@param {type:"boolean"}

#@markdown - Use `trim` to specify regions to trim. For example:
`trim:5-9,20` will remove positions 5,6,7,8,9 and 20.

```

```

#@markdown - For complexes, you can use `trim:A1-A3,B5-B7` to remove
positions 1,2,3 in 1st protein and positions 5,6,7 in 2nd protein.

#@markdown - Note: This function is 1-indexed, meaning the first
position is 1, not 0.

#@markdown - To specify regions to keep instead of trim, enable
`trim_inverse`

cov = 0 #@param [0,25,50,75,90,95] {type:"raw"}

qid = 0 #@param [0,15,20,25,30,40,50] {type:"raw"}

#@markdown - `cov` minimum coverage with query (%)

#@markdown - `qid` minimum sequence identity with query (%)

mod_I = cf_af.prep_filter(I, trim, trim_inverse, cov, qid)

if I["msas"] != mod_I["msas"]:

    plt.figure(figsize=(16,5),dpi=100)

    plt.subplot(1,2,1)

    plt.title("Sequence coverage (Before)")

    cf.plot_msas(I["msas"], I["ori_sequence"], return_plt=False)

    plt.subplot(1,2,2)

    plt.title("Sequence coverage (After)")

    cf.plot_msas(mod_I["msas"], mod_I["ori_sequence"], return_plt=False)

plt.savefig(os.path.join(I["output_dir"],"msa_coverage.filtered.png")
, bbox_inches = 'tight', dpi=200)

plt.show()

#@title Run alphafold

num_relax = "None"

rank_by = "pLDDT" #@param ["pLDDT","pTMscore"]

use_turbo = True #@param {type:"boolean"}

max_msa = "512:1024" #@param ["512:1024", "256:512", "128:256",
"64:128", "32:64"]

max_msa_clusters, max_extra_msa = [int(x) for x in max_msa.split(":")]

#@markdown - `rank_by` specify metric to use for ranking models (For
protein-protein complexes, we recommend pTMscore)

```

#@markdown - `use_turbo` introduces a few modifications (compile once, swap params, adjust max_msa) to speedup and reduce memory requirements. Disable for default behavior.

#@markdown - `max_msa` defines: `max_msa_clusters:max_extra_msa` number of sequences to use. When adjusting after GPU crash, be sure to `Runtime` → `Restart runtime`. (Lowering will reduce GPU requirements, but may result in poor model quality. This option ignored if `use_turbo` is disabled)

```
show_images = True #@param {type:"boolean"}
```

#@markdown - `show_images` To make things more exciting we show images of the predicted structures as they are being generated. (WARNING: the order of images displayed does not reflect any ranking).

#@markdown ---

#@markdown ****sampling options****

#@markdown There are two stochastic parts of the pipeline. Within the feature generation (choice of cluster centers) and within the model (dropout).

#@markdown To get structure diversity, you can iterate through a fixed number of random_seeds (using `num_samples`) and/or enable dropout (using `is_training`).

```
num_models = 5 #@param [1,2,3,4,5] {type:"raw"}
```

```
use_ptm = True #@param {type:"boolean"}
```

```
num_ensemble = 1 #@param [1,2,4,8] {type:"raw"}
```

```
max_recycles = 3 #@param [0,1,3,6,12,24,48] {type:"raw"}
```

```
is_training = False #@param {type:"boolean"}
```

```
num_samples = 1 #@param [1,2,4,8,16,32] {type:"raw"}
```

#@markdown - `num_models` specify how many model params to try. (5 recommended)

#@markdown - `use_ptm` uses Deepmind's `ptm` finetuned model parameters to get PAE per structure. Disable to use the original model params. (Disabling may give alternative structures.)

#@markdown - `num_ensemble` the trunk of the network is run multiple times with different random choices for the MSA cluster centers. (`1`=`default`, `8`=`caspl4 setting`)

#@markdown - `max_recycles` controls the maximum number of times the structure is fed back into the neural network for refinement. (3 recommended)

#@markdown - `is_training` enables the stochastic part of the model (dropout), when coupled with `num_samples` can be used to "sample" a diverse set of structures.

```

#@markdown - `num_samples` number of random_seeds to try.

subsample_msa = True #@param {type:"boolean"}

#@markdown - `subsample_msa` subsample large MSA to `3E7/length`
sequences to avoid crashing the preprocessing protocol. (This option
ignored if `use_turbo` is disabled.)

if not use_ptm and rank_by == "pTMscore":

    print("WARNING: models will be ranked by pLDDT, 'use_ptm' is needed
to compute pTMscore")

    rank_by = "pLDDT"

# prep input features

feature_dict = cf_af.prep_feats(mod_I, clean=IN_COLAB)

Ls_plot = feature_dict["Ls"]

# prep model options

opt = {"N":len(feature_dict["msa"]),

       "L":len(feature_dict["residue_index"]),

       "use_ptm":use_ptm,

       "use_turbo":use_turbo,

       "max_recycles":max_recycles,

       "tol":0.0,

       "num_ensemble":num_ensemble,

       "max_msa_clusters":max_msa_clusters,

       "max_extra_msa":max_extra_msa,

       "is_training":is_training}

if use_turbo:

    if "runner" in dir():

        # only recompile if options changed

        runner = cf_af.prep_model_runner(opt, old_runner=runner)

    else:

        runner = cf_af.prep_model_runner(opt)

else:

```

```

runner = None

#####

# run alphafold

#####

outs, model_rank = cf_af.run_alphafold(feature_dict, opt, runner,
num_models, num_samples, subsample_msa,

                                     rank_by=rank_by,
show_images=show_images)

#@title Refine structures with Amber-Relax (optional)

#@markdown If side-chain bond geometry is important to you, enable
Amber-Relax by specifying how many top ranked structures you want
relaxed. By default, we disable Amber-Relax since it barely moves the
main-chain (backbone) structure and can overall double the runtime.

num_relax = "Top1"  #@param ["None", "Top1", "Top5", "All"]
{type:"string"}

if num_relax == "None":

    num_relax = 0

elif num_relax == "Top1":

    num_relax = 1

elif num_relax == "Top5":

    num_relax = 5

else:

    num_relax = num_models * num_samples

#@markdown - `num_relax` specify how many of the top ranked structures
to relax

if num_relax > 0 and not os.path.isfile("stereo_chemical_props.txt"):

    try:

        total = 45

        with tqdm.notebook.tqdm(total=total, bar_format=TQDM_BAR_FORMAT)
as pbar:

            pbar.set_description(f'INSTALL AMBER')

            with io.capture_output() as captured:

```

```

# Install OpenMM and pdbfixer.

%shell rm -rf /opt/conda

%shell wget -q -P /tmp \

    https://repo.anaconda.com/miniconda/Miniconda3-latest-
Linux-x86_64.sh \

    && bash /tmp/Miniconda3-latest-Linux-x86_64.sh -b -p
/opt/conda \

    && rm /tmp/Miniconda3-latest-Linux-x86_64.sh

pbar.update(4)

PATH=%env PATH

%env PATH=/opt/conda/bin:{PATH}

%shell conda update -qy conda \

    && conda install -qy -c conda-forge \

        python=3.7 \

        openmm=7.5.1 \

        pdbfixer

pbar.update(40)

%shell wget -q -P /content \

    https://git.scicore.unibas.ch/schwede/openstructure/-
/raw/7102c63615b64735c4941278d92b554ec94415f8/modules/mol/alg/src/st
ereo_chemical_props.txt

pbar.update(1)

%shell mkdir -p /content/alphafold/common

%shell cp -f /content/stereo_chemical_props.txt
/content/alphafold/common

# Apply OpenMM patch.

%shell pushd /opt/conda/lib/python3.7/site-packages/ && \

    patch -p0 < /content/alphafold/docker/openmm.patch && \

    popd

except subprocess.CalledProcessError:

```

```

print(captured)

raise

if num_relax > 0:

    if "relax" not in dir():

        # add conda environment to path

        sys.path.append('/opt/conda/lib/python3.7/site-packages')

        if 'LC_ALL' in os.environ:

            del os.environ['LC_ALL']

            # import libraries

            from alphafold.relax import relax

            from alphafold.relax import utils

        with tqdm.notebook.tqdm(total=num_relax,
bar_format=TQDM_BAR_FORMAT) as pbar:

            pbar.set_description(f'AMBER relaxation')

            for n, key in enumerate(model_rank):

                if n < num_relax:

                    prefix = f"rank_{n+1}_{key}"

                    pred_output_path =
os.path.join(I["output_dir"], f'{prefix}_relaxed.pdb')

                    if not os.path.isfile(pred_output_path):

                        amber_relaxer = relax.AmberRelaxation(

                            max_iterations=0,

                            tolerance=2.39,

                            stiffness=10.0,

                            exclude_residues=[],

                            max_outer_iterations=20)

                            relaxed_pdb_lines, _ =
amber_relaxer.process(prot=outs[key] ["unrelaxed_protein"])

                            with open(pred_output_path, 'w') as f:

                                f.write(relaxed_pdb_lines)

                    pbar.update(n=1)

```

```

#@title Display 3D structure (optional) {run: "auto"}
rank_num = 1 #@param ["1", "2", "3", "4", "5"] {type:"raw"}
color = "lDDT" #@param ["chain", "lDDT", "rainbow"]
show_sidechains = False #@param {type:"boolean"}
show_mainchains = False #@param {type:"boolean"}
key = model_rank[rank_num-1]
prefix = f"rank_{rank_num}_{key}"

pred_output_path =
os.path.join(I["output_dir"],f'{prefix}_relaxed.pdb')

if not os.path.isfile(pred_output_path):

    pred_output_path =
os.path.join(I["output_dir"],f'{prefix}_unrelaxed.pdb')

cf.show_pdb(pred_output_path, show_sidechains, show_mainchains, color,
Ls=Ls_plot).show()

if color == "lDDT": cf.plot_plddt_legend().show()

if use_ptm:

    cf.plot_confidence(outs[key]["plddt"], outs[key]["pae"],
Ls=Ls_plot).show()

else:

    cf.plot_confidence(outs[key]["plddt"], Ls=Ls_plot).show()

```


APÊNDICE C - Relatório dos valores de pKa dos aminoácidos para *fold*ing

2022-09-15 12:50:39,796 INFO:main.py:240:print_splash_screen:PDB2PQR v3.5.2:
biomolecular structure conversion software.

2022-09-15 12:50:39,796 INFO:main.py:242:print_splash_screen>Please cite: Jurrus E, et al. Improvements to the APBS biomolecular solvation software suite. *Protein Sci* 27 112-128 (2018).

2022-09-15 12:50:39,796 INFO:main.py:242:print_splash_screen>Please cite: Dolinsky TJ, et al. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res* 35 W522-W525 (2007).

2022-09-15 12:50:39,796 INFO:main.py:752:main_driver:Checking and transforming input arguments.

2022-09-15 12:50:39,797 INFO:main.py:756:main_driver>Loading topology files.

2022-09-15 12:50:39,865 INFO:main.py:758:main_driver>Loading molecule:
rank_1_model_4_ptm_seed_0_relaxed.pdb

2022-09-15 12:50:39,884 INFO:main.py:761:main_driver:Dropping water from structure.

2022-09-15 12:50:39,886 INFO:main.py:763:main_driver:Setting up molecule.

2022-09-15 12:50:39,918 INFO:main.py:408:setup_molecule:Created biomolecule object with 215 residues and 3282 atoms.

2022-09-15 12:50:39,918 INFO:main.py:767:main_driver:Setting termini states for biomolecule chains.

2022-09-15 12:50:39,928 INFO:main.py:603:non_trivial>Loading forcefield.

2022-09-15 12:50:39,937 INFO:main.py:607:non_trivial>Loading hydrogen topology definitions.

2022-09-15 12:50:39,940 INFO:main.py:459:is_repairable:This biomolecule is clean. No repair needed.

2022-09-15 12:50:39,940 INFO:main.py:622:non_trivial:Updating disulfide bridges.

2022-09-15 12:50:39,946 INFO:main.py:625:non_trivial:Debumping biomolecule.

2022-09-15 12:50:40,040 INFO:main.py:632:non_trivial:Assigning titration states with PROPKA.

2022-09-15 12:50:40,057 INFO:output.py:60:print_header:

propka3.4.0

2022-09-15

2022-09-15 12:50:40,369 INFO:output.py:209:print_pka_section:pKa summary:

SUMMARY OF THIS PREDICTION

Group	pKa	model-pKa
ASP 80 A	2.46	3.80
ASP 6 B	3.99	3.80
ASP 17 B	3.70	3.80
ASP 35 B	4.55	3.80
ASP 72 B	2.22	3.80
ASP 82 B	2.93	3.80
ASP 89 B	2.27	3.80
ASP 94 B	4.52	3.80

ASP 101 B	2.98	3.80
ASP 102 B	3.77	3.80
GLU 7 A	2.37	4.50
GLU 8 A	4.55	4.50
GLU 13 A	4.56	4.50
GLU 55 A	4.05	4.50
GLU 63 A	4.56	4.50
GLU 76 A	4.12	4.50
GLU 12 B	4.03	4.50
GLU 58 B	4.45	4.50
C- 91 A	3.35	3.20
HIS 78 A	5.46	6.50
HIS 82 A	5.84	6.50
HIS 89 A	4.35	6.50
HIS 60 B	6.44	6.50
HIS 68 B	6.51	6.50
HIS 70 B	6.38	6.50
HIS 93 B	5.23	6.50
HIS 97 B	5.67	6.50
HIS 113 B	5.99	6.50
CYS 6 A	99.99	9.00
CYS 9 A	99.99	9.00
CYS 27 A	99.99	9.00
CYS 30 A	99.99	9.00
CYS 31 A	99.99	9.00
CYS 58 A	99.99	9.00
CYS 59 A	99.99	9.00
CYS 81 A	99.99	9.00
CYS 83 A	99.99	9.00
CYS 86 A	99.99	9.00
CYS 2 B	99.99	9.00
CYS 16 B	99.99	9.00
CYS 19 B	99.99	9.00
CYS 27 B	99.99	9.00
CYS 31 B	99.99	9.00
CYS 52 B	99.99	9.00
CYS 67 B	99.99	9.00
CYS 83 B	99.99	9.00
CYS 85 B	99.99	9.00
CYS 88 B	99.99	9.00
CYS 95 B	99.99	9.00
CYS 107 B	99.99	9.00
TYR 16 A	10.10	10.00
TYR 36 A	12.41	10.00

TYR 87 A	12.00	10.00
TYR 88 A	10.39	10.00
TYR 7 B	11.25	10.00
TYR 10 B	10.33	10.00
TYR 32 B	9.99	10.00
TYR 54 B	13.67	10.00
TYR 59 B	10.16	10.00
TYR 77 B	11.09	10.00
TYR 84 B	9.54	10.00
TYR 115 B	10.25	10.00
LYS 10 A	11.49	10.50
LYS 12 A	11.01	10.50
LYS 15 A	10.55	10.50
LYS 43 A	8.90	10.50
LYS 44 A	10.14	10.50
LYS 50 A	10.18	10.50
LYS 62 A	11.52	10.50
LYS 65 A	10.39	10.50
LYS 90 A	9.86	10.50
LYS 13 B	10.90	10.50
LYS 44 B	10.34	10.50
LYS 98 B	10.41	10.50
LYS 109 B	10.31	10.50
ARG 19 A	12.41	12.50
ARG 34 A	13.76	12.50
ARG 41 A	12.46	12.50
ARG 66 A	13.27	12.50
ARG 74 A	12.37	12.50
ARG 34 B	13.18	12.50
ARG 45 B	12.40	12.50
ARG 50 B	12.26	12.50
ARG 55 B	12.99	12.50
N+ 1 A	7.89	8.00
N+ 1 B	7.37	8.00

2022-09-15 12:50:40,387 INFO: main.py:635: non_trivial: PROPKA information:

Free energy of folding (kcal/mol) as a function of pH (using neutral reference)

0.00	16.87
1.00	16.77
2.00	15.90
3.00	12.43
4.00	7.92
5.00	4.98

6.00	1.32
7.00	-2.01
8.00	-3.31
9.00	-3.89
10.00	-3.18
11.00	1.64
12.00	7.18
13.00	11.17
14.00	13.63

The pH of optimum stability is 9.3 for which the free energy is -3.9 kcal/mol at 298K

The free energy is within 80 % of maximum at pH 7.9 to 10.0

The free energy is negative in the range 6.4 - 10.7

Protein charge of folded and unfolded state as a function of pH

pH	unfolded	folded
0.00	33.00	32.98
1.00	32.98	32.78
2.00	32.76	31.38
3.00	31.00	27.48
4.00	24.05	21.42
5.00	16.25	14.12
6.00	12.13	9.02
7.00	7.00	5.49
8.00	4.12	3.51
9.00	1.72	1.47
10.00	-6.13	-4.25
11.00	-18.06	-13.47
12.00	-23.64	-20.33
13.00	-28.78	-26.18
14.00	-30.72	-29.78

The pI is 9.40 (folded) and 9.40 (unfolded)

Anexo 2 - Confirmação da sequência depositada no GenBank.

02/09/2022 14:40

Universidade de São Paulo Mail - GenBank OP125854



Renan Passos Freire <renan.passos.freire@usp.br>

GenBank OP125854

gb-admin@ncbi.nlm.nih.gov <gb-admin@ncbi.nlm.nih.gov>
 To: renan.passos.freire@usp.br, repassosfreire@gmail.com

Tue, Aug 2, 2022 at 2:47 PM

Dear GenBank Submitter:

Thank you for your submission of sequence data to GenBank, a contribution which will benefit the scientific community.

We have provided a GenBank accession number for your nucleotide sequence:

BankIt2602407 Seq1 OP125854

The GenBank accession number should appear in any publication that reports or discusses these data, as it gives the community a unique label with which they may retrieve your data from our on-line servers. You may prepare and submit your manuscript before your accession is released in GenBank.

Submissions are not automatically deposited into GenBank after being accessioned. Each sequence record is individually examined and processed by the GenBank annotation staff to ensure that it is free of errors or problems.

You have not requested a specific release date for your sequence data. Therefore, your record(s) will be released to the public database once they are processed. If this is not what you intended, please contact us as soon as possible with the correct release date.

Since the flatfile record is a display format only and is not an editable format of the data, do not make changes directly to a flatfile. For complete information about different methods to update a sequence record, see: <https://www.ncbi.nlm.nih.gov/Genbank/update.html>

Any inquiries about your submission should be sent to gb-admin@ncbi.nlm.nih.gov

For more information about the submission process or the available submission tools, please contact GenBank User Support at info@ncbi.nlm.nih.gov.

Please reply using the current Subject line.

Sincerely,

Rich McVeigh Ph.D.
 Contractor

The GenBank Direct Submission Staff
 Bethesda, Maryland USA

gb-admin@ncbi.nlm.nih.gov (for updates/replies to GenBank entries)
info@ncbi.nlm.nih.gov (for general questions regarding GenBank)

INSTITUTO DE PESQUISAS ENERGÉTICAS E NUCLEARES
Diretoria de Pesquisa, Desenvolvimento e Ensino
Av. Prof. Lineu Prestes, 2242 – Cidade Universitária CEP: 05508-000
Fone/Fax(0XX11) 3133-8908
SÃO PAULO – São Paulo – Brasil
<http://www.ipen.br>

O IPEN é uma Autarquia vinculada à Secretaria de Desenvolvimento, associada à Universidade de São Paulo e gerida técnica e administrativamente pela Comissão Nacional de Energia Nuclear, órgão do Ministério da Ciência, Tecnologia, Inovações e Comunicações.
