

ISIS SEBASTIÃO

**Estudos moleculares e caracterização biofísica da hemoglobina
de *Amyntas gracilis* (HbAg)**

Tese apresentada ao Instituto de Química,
Universidade Estadual Paulista, como parte
dos requisitos para obtenção do título de
Doutor em Biotecnologia.

Orientadora: Profa. Dra. Patrícia Soares
Santiago
Co-orientadora: Profa. Dra. Patrícia Gleydes
Morgante

Araraquara
2022

S443e	Sebastião, Isis Estudos moleculares e caracterização biofísica da hemoglobina de <i>Amyntas gracilis</i> (HbAg) / Isis Sebastião. -- Araraquara, 2022 125 f. : il., tabs., fotos Tese (doutorado) - Universidade Estadual Paulista (Unesp), Instituto de Química, Araraquara Orientadora: Patrícia Soares Santiago Coorientadora: Patrícia Gleydes Morgante 1. Hemoproteínas. 2. Marcadores bioquímicos. 3. Transcriptoma. 4. Minhocas. 5. Sequenciamento de nucleotídeo. I. Título.
-------	---

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Química, Araraquara. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.



UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Araraquara

CERTIFICADO DE APROVAÇÃO



TÍTULO DA TESE: "Estudos moleculares e caracterização biofísica da hemoglobina de *Amyntas gracilis* (HbAg)"

AUTORA: ISIS SEBASTIÃO

ORIENTADORA: PATRICIA SOARES SANTIAGO

COORIENTADORA: PATRÍCIA GLEYDES MORGANTE

Aprovada como parte das exigências para obtenção do Título de Doutora em BIOTECNOLOGIA, pela Comissão Examinadora:

Prof.^a Dr.^a PATRICIA SOARES SANTIAGO (Participação Virtual)
Universidade Estadual Paulista - UNESP - Registro

Dra. ELISA MARIA COSTA E SILVA DE PAIVA (Participação Virtual)
Departamento de Zoologia / Instituto de Biociências - USP - São Paulo

Prof. Dr. NORIVAL ALVES SANTOS FILHO (Participação Virtual)
Departamento de Bioquímica e Química Orgânica / Instituto de Química - UNESP - Araraquara

Dr.^a. FERNANDA ANGELICA SALA (Participação Virtual)
Departamento de Recursos Naturais e Proteção Ambiental / Universidade Federal de São Carlos - UFSCAR - Araras

Prof. Dr. FRANCISCO ADRIANO DE OLIVEIRA CARVALHO (Participação Virtual)
Instituto de Ciências Exatas - UNIFESSPA - Marabá

Araraquara, 24 de junho de 2022

DADOS CURRICULARES

IDENTIFICAÇÃO

Nome: Isis Sebastião

Nome em citações bibliográficas: SEBASTIAO, I.; SEBASTIÃO, ISIS; SEBASTIÃO, I.

ENDEREÇO PROFISSIONAL: Universidade Estadual Paulista Júlio de Mesquita Filho,
Av. Nelson Brihi Badur, 430 - Vila Tupy, Registro - SP, 11900-000

FORMAÇÃO ACADÊMICA/TITULAÇÃO:

Mestrado em Agronomia (Genética e Melhoramento de Plantas)

Graduação em Agronomia

PRODUÇÃO BIBLIOGRÁFICA

Artigos completos publicados em periódicos

1. OLIVEIRA, J. B. S.; RAMOS, L.; SOUZA, C. O.; **SEBASTIÃO, I.**; CARUSO, C.; CARVALHO, F. A. O.; CARVALHO, J. W. P.; MORGANTE, P. G.; SANTIAGO, P. S. Initial biophysical characterization of *Amyntas gracilis* giant extracellular hemoglobin (HbAg). **Eur. Biophys. J.**, v.49, p.473 - 484, 2020.
2. SOARES FIGUEIREDO, CAMILA; NUNES LEMES, ANA RITA; **SEBASTIÃO, ISIS**; DESIDÉRIO, JANETE APPARECIDA. Synergism of the *Bacillus thuringiensis* Cry1, Cry2, and Vip3 Proteins in *Spodoptera frugiperda* Control. **Appl. Biochem. Biotechnol.**, v.1, p.1 - 12, 2019.
3. LEMES, ANA RITA NUNES; FIGUEIREDO, CAMILA SOARES; **SEBASTIÃO, ISIS**; MARQUES DA SILVA, LILIANE; DA COSTA ALVES, REBEKA; DE SIQUEIRA, HERBERT ÁLVARO ABREU; LEMOS, MANOEL VICTOR FRANCO; FERNANDES, ODAIR APARECIDO; DESIDÉRIO, JANETE APPARECIDA. Cry1Ac and Vip3Aa proteins from *Bacillus thuringiensis* targeting Cry toxin resistance in *Diatraea flavipennella* and *Elasmopalpus lignosellus* from sugarcane. **PeerJ.**, v.5, p.e2866 - , 2017.
4. **SEBASTIAO, I.**; LEMES, A. R. N.; FIGUEIREDO, C. S.; POLANCZYK, R. A.; DESIDERIO, J. A.; LEMOS, M. V. F. Toxicidade e capacidade de ligação de proteínas Cry1 a receptores intestinais de *Helicoverpa armigera* (Lepidoptera: Noctuidae). **Pesquisa Agropecuária Brasileira** (Online). v.50, p.999 - 1005, 2015.
5. MORGANTE, P. G.; SEBASTIAO, I.; SILVEIRA, L. E. D.; Mori, G. M.; CONTE, M.; COFFANI-NUNES, J. V. Development of microsatellite markers for *Pimenta pseudocaryophyllus* (Myrtaceae), a wild South American species. **Am. J. Bot.** v.99, p.e434 - e436, 2012.

Livros publicados

1. LEMES, A. R. N.; FIGUEIREDO, C. S.; **SEBASTIAO, I.**; DESIDERIO, J. A.; LEMOS, M. V. F. Tópicos especiais em genética aplicada. Jaboticabal: FUNEP, 2015, v.2. p.109.

Capítulos de livros publicados

1. LEMES, A. R. N.; PEREIRA, F. D.; **SEBASTIAO, I.**; DESIDERIO, J. A.; LEMOS, M. V. F. *Bacillus thuringiensis*: histórico, importância e perspectivas In: Tópicos Especiais em Genética Aplicada. 1 ed. Jaboticabal: Funep, 2015, v.2, p. 01-109.

2. LEMES, A. R. N.; PEREIRA, F. D.; **SEBASTIAO, I.**; DESIDERIO, J. A.; LEMOS, M. V. F. DNA e Proteínas: processos biomoleculares que mudaram o contexto da ciência atual In: Tópicos Especiais em Genética Aplicada. 1 ed. Jaboticabal: FUNEP, 2014, v.1, p. 1-12.

Trabalhos publicados em anais de eventos (resumo)

1. SANTIAGO, P. S.; RAMOS, L.; IZUMI, K. S.; SOUZA, C. O.; **SEBASTIÃO, I.**; BERTINI, G. CONFORMATIONAL CHANGES OF FRESHWATER SHRIMP HEMOCYANIN INDUCED BY UREA. In: 20th IUPAB Congress, 50th Annual SBBq Meeting, 45th Annual SBBf Meeting, 2021

2. **SEBASTIÃO, I.**; SANTOS, E. S.; PEREIRA, L. M.; MORGANTE, P. G.; SANTIAGO, P. S. IDENTIFICATION AND PREDICTION OF HbAg D-CHAIN SEQUENCE. In: 20th IUPAB Congress, 50th Annual SBBq Meeting, 45th Annual SBBf Meeting, 2021

3. **SEBASTIAO, I.**; SOUZA, C. O.; RAMOS, L.; MORGANTE, P. G.; SANTIAGO, P. S. Interaction of monomer *d* of the extracellular hemoglobin of *Amyntas gracilis* (HbAg) with the anionic surfactant sodium dodecyl sulfate (SDS). In: XLIV Congress of the Brazilian Biophysical Society, 2019, Santos.

4. SOUZA, C. O.; RAMOS, L.; **SEBASTIAO, I.**; MORAES, M. L.; SANTIAGO, P. S. Use of Hemoglobins for Metal Biosensing. In: XLIV Congress of the Brazilian Biophysical Society, 2019, Santos.

5. MORAIS, C. M.; SOUZA, C. O.; ROZANE, D. E.; **SEBASTIAO, I.**; SANTIAGO, P. S.; MACHADO, L. P. ANÁLISES QUÍMICAS COMPARATIVAS DO PRODUTO BIOESTIMULANTE AGRÍCOLA DERIVADO DA MACROALGA *Ascophyllum nodosum* E DO EXTRATO DA MACROALGA NATIVA *Osmundaria obtusiloba*. In: SIMPESCA - III Simpósio Sul-Sudeste de Recursos Pesqueiros e Engenharia de Pesca, 2018, Registro.

6. MORAIS, C. M.; SOUZA, C. O.; ROZANE, D. E.; **SEBASTIAO, I.**; MACHADO, L. P. Análises químicas comparativas do produto bioestimulante agrícola derivado da macroalga *Ascophyllum nodosum* (Ochrophyta) e do extrato da macroalga nativa *Osmundaria obtusiloba* (Rhodophyta). In: 30º Congresso de Iniciação Científica da Unesp, 2018, Registro.

7. **SEBASTIAO, I.**; OLIVEIRA, J. B. S.; OHYA, M. C.; SOUZA, C. O.; RAMOS, L.; SANTIAGO, P. S. Purification and Estimation of the Isoelectric Point (pI) of the Monomer of the Extracellular Hemoglobin of *Amyntas gracilis* (HbAg). In: XLIII Congresso da Sociedade Brasileira de Biofísica, 2018, Santos.
8. SOUZA, C. O.; RAMOS, L.; **SEBASTIAO, I.**; MORAES, M. L.; SANTIAGO, P. S. SENSITIVITY OF DETECTION OF COPPER AND CADMIUM USING ANNELID HEMOGLOBIN. In: XLIII Congresso Brasileiro da Sociedade de Biofísica, 2018, Santos.
9. LEMES, A. R. N.; FIGUEIREDO, C. S.; **SEBASTIAO, I.**; LEMOS, M. V. F.; FERNANDES, O. A.; DESIDERIO, J. A. Combinação de genes cry1Aa e vip3Aa na construção de cana-de-açúcar geneticamente modificada para controle de *Elasmopalpus lignosellus*. In: 8º Congresso Brasileiro de Melhoramento de Plantas, 2015, Goiânia.
10. LEMES, A. R. N.; **SEBASTIAO, I.**; SILVA, L. M.; FIGUEIREDO, C. S.; ALVES, R. C.; SIQUEIRA, H. A. A.; LEMOS, M. V. F.; FERNANDES, O. A.; DESIDERIO, J. A. Controle de *Diatraea flavipennella* com proteínas Vip3 de *Bacillus thuringiensis*. In: 14º Simpósio de Controle Biológico - Siconbiol, 2015, Teresópolis.
11. FIGUEIREDO, C. S.; **SEBASTIAO, I.**; LEMES, A. R. N.; DESIDERIO, J. A. Interação das proteínas Cry2Aa, Cry2Ab e Vip3Aa no controle de lagartas neonatas de *Anticarsia gemmatilis*. In: 14º Simpósio de Controle Biológico-Siconbiol, 2015, Teresópolis.
12. FIGUEIREDO, C. S.; LEMES, A. R. N.; **SEBASTIAO, I.**; DESIDERIO, J. A. Proteínas Cry1Ab e Cry1Ac como estratégia de manejo de resistência de *Anticarsia gemmatilis* plantas transgênicas de soja (*Glicine max*). In: 8º Congresso Brasileiro de Melhoramento de Plantas, 2015, Goiânia.
13. LEMOS, M. V. F.; **SEBASTIAO, I.**; FIGUEIREDO, C. S.; LEMES, A. R. N.; POLANCZYK, R. A.; DESIDERIO, J. A. Toxicity and interaction of Cry1 proteins from *Bacillus thuringiensis* in *Helicoverpa armigera* (Hübner) (Lepidoptera: Noctuidae). In: International Congress on Invertebrate Pathology and Microbial Control and the 48th Annual Meeting of the Society for Invertebrate Pathology, 2015, Vancouver.
14. FIGUEIREDO, C. S.; **SEBASTIAO, I.**; LEMES, A. R. N.; POLANCZYK, R. A.; DESIDERIO, J. A. Atividade inseticida das proteínas Vip3Aa50 e Cry2Ab de *Bacillus thuringiensis* para *Helicoverpa armigera* (Hübner) (Lepidoptera: Noctuidae) In: XXV Congresso Brasileiro de Entomologia (XXV CBE), 2014, Goiânia.
15. **SEBASTIAO, I.**; FIGUEIREDO, C. S.; LEMES, A. R. N.; DESIDERIO, J. A.; POLANCZYK, R. A.; LEMOS, M. V. F. Suscetibilidade de *Helicoverpa armigera* (Hübner) (Lepidoptera: Noctuidae) às proteínas Cry1 de *Bacillus thuringiensis* In: XXV Congresso Brasileiro de Entomologia (XXV CBE), 2014, Goiânia.
16. FIGUEIREDO, C. S.; **SEBASTIAO, I.**; LEMES, A. R. N.; DESIDERIO, J. A. Suscetibilidade de *Spodoptera frugiperda* e *Anticarsia gemmatilis* à proteína Cry2Aa de *Bacillus thuringiensis* T01-326. In: XXII Congresso Latinoamericano de Microbiología - ALAM 2014, 2014, Cartagena.

17. FIGUEIREDO, C. S.; OLIVEIRA, F. M.; LEMES, A. R. N.; **SEBASTIAO, I.**; DESIDERIO, J. A. Toxicidade e interação das proteínas Vip3Aa50 e Cry2Ab de *Bacillus thuringiensis* no controle de *Spodoptera frugiperda* In: XXV Congresso Brasileiro de Entomologia (XXV CBE), 2014, Goiânia.
18. LEMES, A. R. N.; FIGUEIREDO, C. S.; **SEBASTIAO, I.**; LEMOS, M. V. F.; FERNANDES, O. A.; DESIDERIO, J. A. Uso de proteínas Vip3 de *Bacillus thuringiensis* com potencial no controle de *Elasmopalpus lignosellus* na cultura da cana-de-açúcar. In: XXV Congresso Brasileiro de Entomologia (XXV CBE), 2014, Goiânia.
19. MORGANTE, P. G.; **SEBASTIAO, I.**; SILVEIRA, L. E.; Mori, G. M.; COMTE, M.; COFFANI-NUNES, J. V. Development of eleven polymorphic microsatellite markers for a Myrtaceae species. In: Congress IUBMB/FEBS, 2012, Sevilha.
20. MORGANTE, P. G.; **SEBASTIAO, I.**; SILVEIRA, L. E.; Mori, G. M. Development of Polymorphic Microsatellite Markers for *Pimenta pseudocaryophyllus* (Gomes) Landrum, Myrtaceae. In: XVIII International Botanical Congress, 2011, Melbourne.
21. **SEBASTIAO, I.**; SILVEIRA, L. E.; HIRATA, D. M.; COFFANI-NUNES, J. V.; MORGANTE, P. G. Busca de Locos SSR Polimórficos em *Pimenta pseudocaryophyllus* (Gomes) Landrum, Myrtaceae. In: X Congresso Latinoamericano de Botânica, 2010, La Serena.
22. SILVEIRA, L. E.; **SEBASTIAO, I.**; WEISSENBERG, E. W.; COFFANI-NUNES, J. V.; MORGANTE, P. G. Estabelecimento de Condições de Armazenamento e Extrações de DNA de Convolvulaceae. In: X Congresso Latinoamericano de Botânica, 2010, La Serena.
23. MORGANTE, P. G.; **SEBASTIAO, I.**; SILVEIRA, L. E.; WEISSENBERG, E. W.; COFFANI-NUNES, J. V. Uso de PCR Touchdown (TD - PCR) para Obtenção de Condições Ótimas de Amplificação de Locos Microsatélites. In: X Congresso Latinoamericano de Botânica, 2010, La Serena

APRESENTAÇÕES DE TRABALHO

1. **SEBASTIÃO, I.**; SANTOS, E. S.; PEREIRA, L. M.; MORGANTE, P. G.; SANTIAGO, P. S. IDENTIFICATION AND PREDICTION OF HbAg D-CHAIN SEQUENCE. 2021. (Apresentação de Trabalho/Congresso).
2. **SEBASTIAO, I.**; SOUZA, C. O.; RAMOS, L.; MORGANTE, P. G.; SANTIAGO, P. S. Interaction of monomer *d* of the extracellular hemoglobin of *Amyntas gracilis* (HbAg) with the anionic surfactant sodium dodecyl sulfate (SDS). 2019. (Apresentação de Trabalho/Congresso).
3. **SEBASTIAO, I.**; OLIVEIRA, J. B. S.; OHYA, M. C.; SOUZA, C. O.; RAMOS, L.; SANTIAGO, P. S. Purification and Estimation of the Isoelectric Point (pI) of the Monomer-*d* of the Extracellular Hemoglobin of *Amyntas gracilis* (HbAg). 2018. (Apresentação de Trabalho/Congresso)
4. **SEBASTIAO, I.** As proteínas Cry. 2016. (Apresentação de Trabalho/Conferência ou palestra).

5. SEBASTIAO, I.; FIGUEIREDO, C. S.; LEMES, A. R. N.; DESIDERIO, J. A.; POLANCZYK, R. A.; LEMOS, M. V. F. Suscetibilidade de *Helicoverpa armigera* (Hübner) (Lepidoptera: Noctuidae) às proteínas Cry1 de *Bacillus thuringiensis*. 2014. (Apresentação de Trabalho/Congresso).

6. SEBASTIAO, I.; MORGANTE, P. G.; SILVEIRA, L. E.; Mori, G. M. Marcadores Moleculares Microssatélites Isolados de *Pimenta pseudocaryophyllus* (Gomes) Landrum, Myrtaceae. 2011. (Apresentação de Trabalho/Congresso).

7. SEBASTIAO, I.; SILVEIRA, L. E.; HIRATA, D. M.; COFFANI-NUNES, J. V.; MORGANTE, P. G. Busca de Locos SSR Polimórficos em *Pimenta pseudocaryophyllus* (Gomes) Landrum, Myrtaceae. 2010. (Apresentação de Trabalho/Congresso).

8. SEBASTIAO, I.; MORGANTE, P. G.; SILVEIRA, L. E.; HIRATA, D. M.; COFFANI-NUNES, J. V. Estabelecimento de Condições de Amplificação por PCR de Locos SSR em *Pimenta pseudocaryophyllus* (Gomes) Landrum, Myrtaceae. 2010. (Apresentação de Trabalho/Congresso).

Produção técnica

1. LEMES, A. R. N.; SEBASTIAO, I.; PEREIRA, F. D.; DESIDERIO, J. A.; LEMOS, M. V. F. Técnicas e manipulação de biomoléculas (DNA e proteína), 2014 e 2015. (Extensão, Curso de curta duração ministrado)

PARTICIPAÇÃO EM BANCA DE TRABALHOS DE CONCLUSÃO

Graduação

1. SANTIAGO, P. S.; BERTINI, G.; SEBASTIAO, I. Participação em banca de Marcia Cristina Ohya. Separação e caracterização inicial do monômero-d da hemoglobina extracelular de *Amyntas gracilis* (HbAg), 2017. (Agronomia) Universidade Estadual Paulista- Campus Experimental de Registro

PARTICIPAÇÃO EM BANCA DE COMISSÕES JULGADORAS

1. 1ª fase do XXXI Congresso de Iniciação Científica da Unesp, 2019 -UNESP Campus Experimental de Registro

2. IV Congresso Brasileiro de Fitossanidade - IV Conbraf - Avaliadora de resumos submetidos no congresso, 2017

3. 1ª fase do XXIX Congresso de Iniciação Científica da Unesp, 2017- UNESP Campus Experimental de Registro

PARTICIPAÇÃO EM EVENTOS CIENTÍFICOS

1. 20th IUPAB Congress, 50th Annual SBBq Meeting, 45th Annual SBBf Meeting, 1. IDENTIFICATION AND PREDICTION OF HBAG D-CHAIN SEQUENCE. 2021. (Congresso).

2. XLIII Congresso da Sociedade Brasileira de Biofísica. Purification and Estimation of the Isoelectric Point (pI) of the Monomer-d of the Extracellular Hemoglobin of *Amyntas gracilis* (HbAg).2018. (Congresso).

3. I Aula Magna. 2014. (Outra).

4. Inova Jab - 1ª Feira de Inovação em Ciências Agrárias. 2014. (Feira).
5. XXV Congresso Brasileiro de Entomologia (XXV CBE). Suscetibilidade de *Helicoverpa armigera* (Hübner) (Lepidoptera: Noctuidae) às proteínas Cry1 de *Bacillus thuringiensis*. 2014. (Congresso).
6. 1º TECNOBIO Pragas. 2013. (Outra).
7. IX Curso de Inverno de Genética. 2013. (Outra).
8. III WORKSHOP SOBRE NUTRIÇÃO E ADUBAÇÃO DA BANANA. 2011. (Outra).
9. VII Semana de Atualização Agronômica UNESP-Registro. 2011. (Outra).
10. XXIII Congresso de Iniciação Científica da Unesp- fase I. Marcadores Moleculares Microssatélites Isolados de *Pimenta pseudocaryophyllus* (Gomes) Landrum, Myrtaceae. 2011. (Congresso).
11. XXIII Congresso de Iniciação Científica da Unesp- fase II. Marcadores Moleculares Microssatélites Isolados de *Pimenta pseudocaryophyllus* (Gomes) Landrum. 2011. (Congresso).
12. VI - Semana Agronômica Campus Experimental de Registro. 2010. (Outra).
13. X Congresso Latinoamericano de Botânica. Busca de Loci SSR Polimórficos em *P. pseudocaryophyllus* (Gomes) Landrum, Myrtaceae; Uso de PCR -TD para obtenção de condições ótimas de amplificação de locos microssatélites; Estabelecimento de condições de armazenamento e extrações de DNA de Convolvulaceae.. 2010. (Congresso).
14. XXII Congresso de Iniciação Científica da UNESP. "Estabelecimento de Condições de Amplificação por PCR de Loci SSR em *Pimenta pseudocaryophyllus* (Gomes) Landrum, Myrtaceae"; "Condições Adequadas para o Armazenamento e Extração de DNA de Convolvulaceae. 2010. (Congresso).
15. I Dia de Campo " A cultura da Lichia: Potencial para o Vale do Ribeira". 2009. (Outra).
16. II Simpósio de Fitossanidade no Vale do Ribeira. 2009. (Outra).
17. UcorSA/ VR. 2009. (Outra).
18. V Semana de Atualização Agronômica/ Transgênicos: Situação Atual e Perspectivas. 2009. (Outra).
19. Workshop sobre Arborização Urbana no Vale do Ribeira. 2009. (Outra).
20. IV Semana Agronômica. 2008. (Outra).
21. III Semana Agronômica. 2007. (Outra).
22. I Simpósio de Fitossanidade no Vale do Ribeira. 2007. (Simpósio).

Dedico este trabalho aos meus pais que sempre me ajudaram nos momentos em que precisei.

Dedico também a todas as minhas figuras internas, ao início da construção do arquétipo da minha mãe interna, que me maternou nos momentos em que eu precisei de colo, cuidado e acolhimento.

AGRADECIMENTOS

Agradeço aos meus pais Renato e Wilma e toda família pelo apoio e acolhimento que recebi durante esta fase.

À minha orientadora Profa. Dra. Patrícia Soares Santiago pelo convite e por me apresentar uma nova área de estudo e aperfeiçoamento, pela orientação e todas as oportunidades e parcerias que ela buscou para realização desse projeto de Doutorado.

À minha co-orientadora Profa. Dra. Patrícia Gleydes Morgante pela orientação e participação ativa no projeto, além disso, pela amizade, escuta e apoio nos meus momentos de vulnerabilidade durante esse período.

Ao Prof. Dr. Richard Charles Garrat do IFSC-USP-São Carlos pela parceria e oportunidade de realizar parte do projeto em seu laboratório.

Ao pesquisador Dr. Leandro de Mattos Pereira pela participação ativa neste trabalho nas análises da anotação do transcriptoma.

Ao pesquisador Dr. José Fernando Ruggiero Bachega também pela participação ativa neste trabalho nas análises do modelo comparativo da proteína HbAg.

Ao professor Sergio Akira Uyemura da Faculdade de Ciências Farmacêuticas da USP-Ribeirão Preto, por me receber em seu laboratório para etapa do sequenciamento do material genético. Ao técnico de laboratório Emerson Santos que me acompanhou em toda essa etapa.

À minha amiga Luciana que me recebeu em sua casa em Ribeirão Preto durante o período de experimentos na USP.

Aos professores e funcionários da UNESP Registro, em especial a Liz e Luís Carlos pela amizade e conversas.

Aos meus colegas de trabalho do Biomoléculas: Claudemir, Lierge, Jonathan, Raphael, Camilinha pelos ensinamentos, trabalhos que realizamos juntos, por compartilharmos de perrengues e momentos de dificuldade, além da grande amizade que conseguimos construir neste período.

À Maria, a qual tive a oportunidade de reencontrar e partilhar muito dessa jornada.

A todos os amigos que a academia me trouxe nestes anos de trabalho, em especial à Camila que conheci na época do Mestrado e tivemos a oportunidade de se reencontrar neste período de Doutorado. À Ana Rita que pelas conversas diárias, conselho e por me fazer enxergar a realidade em momentos confusos.

Ao Laboratório de cristalografia em São Carlos, em especial à Su, Fer e Murilo que me deram suporte técnico e pela amizade.

A todas as pessoas queridas que encontrei em São Carlos, em especial ao Leo pela conexão, apoio, parceria e partilha de grandes momentos, momentos felizes e difíceis dessa fase.

Aos meus queridos amigos da vida que mesmo distantes sempre estiveram presentes comigo.

As funcionárias da STPG de Araraquara pelos serviços e apoio prestados.

Ao programa de pós-graduação em Biotecnologia do IQ-UNESP Araraquara.

Ao CNPq pela bolsa concedida (processo nº 141022/2018-3) e a FAPESP (processo 2015/11447-1) pelo apoio financeiro na execução deste projeto.

**Ninguém se torna iluminado por imaginar figuras de luz, mas sim por tornar
consciente a escuridão**

Carl G. Jung

RESUMO

Espécies de minhocas são comumente utilizadas em pesquisas de ecotoxicologia por possuírem características que auxiliam na identificação de poluentes no meio ambiente. Com o avanço da biologia molecular foi possível rastrear proteínas que podem ser utilizadas como marcadores de contaminação ambiental, é o caso da Hemoglobina extraída da espécie de *Amyntas gracilis* (HbAg). Conhecida como Hb extracelular gigante, a HbAg é um complexo proteico com massa molecular em torno de 3,6 MDa, com capacidade elevada de carregamento de moléculas de O₂, estabilidade térmica e resistência ao estresse oxidativo. Todas essas propriedades fazem com que essas Hbs sejam alvo de pesquisas dentro da área de biotecnologia. O objetivo deste trabalho foi anotar o transcriptoma de *A. gracilis*, permitindo dessa forma, identificar sequências de proteínas com potencial biotecnológico, assim como as sequências codificadoras para as cadeias que compõem a HbAg. Além disso, foi realizado um estudo inicial de caracterização de uma das cadeias dessa hemoproteína, o monômero-*d*. A montagem do transcriptoma foi realizada utilizando estratégia *de novo* com o programa Trinity e a anotação funcional foi realizada com o programa OmicsBox que inclui a ferramenta do Blast2GO. As sequências das cadeias de HbAg foram identificadas através da análise de BLAST contra o banco de dados UniprotKB. O modelo para a unidade biológica HbAg foi construído por modelagem comparativa com auxílio do SWISS-model e os esquemas foram desenhados com o programa PyMOL. A caracterização inicial do monômero-*d* de HbAg foi feita através das técnicas de espectroscopia: absorção UV-Vis e fluorescência, além da técnica de espalhamento de luz (LSI) variando as condições experimentais em função do pH e interação com o surfactante SDS. Foram identificadas 260.616 ORFs das quais 187.866 sequências foram não redundantes e, a partir destas foram anotadas 92.939 sequências de proteínas. Proteínas com potencial biotecnológico assim como classes de enzimas envolvidas no processo de vermicompostagem foram identificadas. Todas as sequências das cadeias de HbAg foram preditas e partir delas construímos um modelo comparativo com base na estrutura de HbGp, Hb homóloga a HbAg. Todos os domínios proteicos mantiveram-se conservados em HbAg. Para o conjunto de dados do monômero-*d* em função do pH, notou-se auto-oxidação do grupo Heme em pH 9, aumento da intensidade de fluorescência em pH mais básico e possível auto agregação no pH mais ácido (5,0) e pH mais básico (9,0). Já para a interação com o SDS, notamos uma oxidação mais pronunciada a partir da concentração de 4 mmol L⁻¹. O aumento gradativo da intensidade de fluorescência foi observado a partir da concentração de 1,2 mmol L⁻¹ de surfactante. Os dados de LSI levantam a hipótese da formação de agregados entre o monômero-*d* e o surfactante assim como a presença de auto agregados de monômero em solução. Os dados apresentados neste trabalho trazem conhecimento básico acerca do potencial de *A. gracilis* no campo da biotecnologia e poderão fomentar pesquisas futuras nas áreas de genômica, com enfoque para o desenvolvimento e caracterização de biomarcadores moleculares na investigação de contaminação ambiental.

Palavras-chave: hemoproteína; biomarcador; transcriptoma; minhoca; RNA-Seq.

ABSTRACT

Earthworm species have been used in the field of ecotoxicology research due to their characteristics that support the identification of environmental pollution. Advances in molecular biology have made it possible to detect proteins with potential application of biomarkers, such as the Hemoglobin from *Amyntas gracilis* (HbAg). Known as an extracellular giant Hb, the HbAg is a protein complex with molecular mass of 3,6 MDa, high capacity of O₂ carrier, thermal stability, and oxidative stress resistance. These properties lead to giant Hbs becoming a target of research in the biotechnological field. The objective of this study was to annotate the *A. gracilis* transcriptome, allowing the sequence identification of proteins with biotechnological potential, as well as the identification of chains that compound HbAg. Besides that, the initial biophysical characterization of monomer-*d* from HbAg was carried out. *De novo* transcriptome assembly was carried out using the Trinity program and the functional annotation was realized with OmicsBox program that includes the Blast2GO tool. The chain sequences of HbAg were identified through BLAST analysis against to UniprotKB database. The comparative model of HbAg biological unit was predicted using SWISS model and the schemes were drawn with the PyMOL program. The initial biophysical characterization of the monomer-*d* of HbAg was carried out using absorption spectroscopy: UV-Vis and fluorescence, in addition to the light scattering intensity technique, varying the experimental conditions as a function of pH and interaction with the surfactant SDS. In the transcriptome were identified 260,616 ORFs, of which 187,866 sequences were non-redundant. The OmicsBox program annotated 92,939 protein sequences. Proteins with biotechnological potential and enzymes involved in the vermicomposting process were identified. All sequences of the HbAg chains were predicted and the comparative model was constructed using HbGp structure as a template. The conservation of protein domains in HbAg model was preserved. The evaluation of the pH effect on monomer-*d* revealed the Heme oxidation in pH 9.0, the increase of fluorescence intensity in basic pH and the possible autoaggregation was observed in acid pH (5.0) and basic pH (9.0). The interaction of monomer-*d* and SDS showed a more pronounced Heme oxidation from the concentration of 4 mmol L⁻¹ of the surfactant. The increase of fluorescence intensity was observed from 1.2 mmol L⁻¹ of surfactant concentration. The LSI data support the hypothesis of aggregation between monomer-*d* and SDS, as well as, the presence of dimer of monomer-*d* in solution. The data presented in this work bring basic knowledge about the potential of *A. gracilis* in the field of biotechnology and may encourage future research in the area of genomics, focusing on the development and characterization of molecular biomarkers in the investigation of environmental contamination.

Key words: hemoprotein; biomarker; transcriptome; earthworm; RNA-Seq.

LISTA DE FIGURAS

Figura 1: Representação gráfica da estrutura da Hb humana e grupo Heme.	27
Figura 2: Representação da estrutura de HbGp, Hb homóloga à HbAg.	29
Figura 3: Representação esquemática das etapas envolvidas nos sequenciadores da plataforma Illumina.	36
Figura 4: Purificação do RNA total.	38
Figura 5: Confeção das moléculas de cDNA dupla fita.	40
Figura 6: Placa de utilizada na quantificação da biblioteca de cDNA, por meio de PCR em tempo real.	42
Figura 7: Gel de agarose 1% em corrida de eletroforese realizada em condições desnaturantes para verificação da extração de RNA total de <i>A. gracilis</i>	45
Figura 8: Gel de agarose 1% corado com brometo de etídio para verificação do tamanho da biblioteca de cDNA de <i>A. gracilis</i>	47
Figura 9: Representação gráfica dos ciclos de amplificação x intensidade de fluorescência emitida durante o processamento das reações de PCR em tempo real	49
Figura 10: Gráficos gerados pela ferramenta do FastQC, que representam a qualidade do sequenciamento por base.	51
Figura 11: Representação gráfica da relação entre o número de hits dentro do transcriptoma de <i>A. gracilis</i> que tiveram correspondência com sequências depositadas no banco de dados NR do NCBI através da análise de BLAST.	54
Figura 12: Anotação obtida pelo Gene Ontology (GO).	56
Figura 13: Distribuição das principais proteínas anotadas através da ferramenta InterproScan.	64
Figura 14: Distribuição das enzimas encontrada no transcriptoma de <i>A. gracilis</i>	65
Figura 15: Alinhamento das sequências das globinas <i>a</i> , <i>b</i> , <i>c</i> e <i>d</i> de HbAg contra as Hbs homólogas HbGp e HbLt.	75
Figura 16: Árvore filogenética construída a partir de análise bayesiana entre sequências de diferentes globinas extracelulares.	77
Figura 17: Alinhamento das sequências dos <i>linkers</i> de HbAg e os <i>linkers</i> das Hbs homólogas HbGp e HbLt.	79
Figura 18: Esquema geral da organização do complexo proteico HbAg.	80
Figura 19: Representação das quatro globinas de HbAg [<i>abcd</i>].	81
Figura 20: Detalhes da região coiled coil ou super hélice formada pela interação dos <i>linkers</i> L1, L2 e L3.	83
Figura 21: Esquema detalhando os contatos realizados entre os <i>linkers</i> na formação da bicamada hexagonal de HbAg.	84
Figura 22: Sítios de ligação aos íons Ca^{2+} em L1, L2 e L3.	85
Figura 23: Representação da localização dos sítios de ligação aos íons de Zn^{2+} em HbAg.	86
Figura 24: Representação esquemática do Diagrama de Jablonski.	93
Figura 25: Purificação do monômero- <i>d</i> de HbAg: Cromatografia e Gel SDS-PAGE.	98
Figura 26: Espectro de absorção de luz UV-VIS para oxi-HbAg em pH 7,0.	99
Figura 27: Representação gráfica dos espectros de absorção de luz para o monômero- <i>d</i> sob diferentes faixas de pH.	101
Figura 28: Espectros de emissão de fluorescência do monômero- <i>d</i> de HbAg em função da variação de pH.	104

Figura 29: Área total de emissão de fluorescência do monômero- <i>d</i> de HbAg normalizada pela emissão do monômero-d em pH 7,0.....	105
Figura 30: Gráficos relacionando pH e espalhamento de luz <i>light scattering intensity</i> monômero- <i>d</i>	107
Figura 31: Interação entre o monômero- <i>d</i> e o surfactante SDS monitorada por absorção ótica.....	109
Figura 32: Interação entre o monômero- <i>d</i> e o surfactante SDS monitorada por fluorescência.....	111
Figura 33: Curva das medidas de espalhamento de luz (LSI) para a interação entre o monômero- <i>d</i> de HbAg com o surfactante aniônico SDS.....	112

LISTA DE TABELAS

Tabela 1: Resultados estatísticos para a montagem do transcriptoma de <i>A. gracilis</i> realizada através do Software Trinity.	52
Tabela 2: Lista dos 20 genes mais expressos no transcriptoma de <i>A. gracilis</i>	59
Tabela 3: Tabela descritiva correlacionando as sequências identificadas para as subunidades de HbAg com os parâmetros obtidos através de análise de BLAST.....	72

LISTA DE ABREVIATURAS E SÍMBOLOS

2,3-DPG 2,3 difosfatoglicerato

Arg Arginina

Asn Asparagina

Asp Ácido aspártico

CC Componentes Celulares

Cd cádmio

cDNA ácido desoxirribonucleico complementar

CT *threshold cycle*

Cu cobre

DNA ácido desoxirribonucleico

FM Funções Moleculares

Gln Glutamina

Glu Ácido glutâmico

GO Gene Ontology

Hb Hemoglobina

HbAg Hemoglobina gigante extraída da espécie *Amyntas gracilis*

HbAm Hemoglobina gigante extraída da espécie *Arenicola marina*

HbGp Hemoglobina gigante extraída da espécie *Glossoscolex paulistus*

HBL-Hb hexagonal bilayer hemoglobin

HbLt Hemoglobina gigante extraída da espécie *Lumbricus terrestris*

Hc Hemocianina

Hr Hemeritina

His Histidina

Ile Isoleucina

kDa kilo Dalton

LegHb leghemoglobina

Leu Leucina

LMCT Transferência de carga ligante-metal

MALDI TOF-MS *Matrix Assisted Lazer Desorption Ionization Time-of-Flight Mass Spectrometry*

MM massa molecular

mRNA RNA mensageiro

N Nitrogênio

NCBI *National Center of Biotechnology Information*

NGS *Next Generation Sequencing*

NR banco de dados *non-redundant* do NCBI

ORF *open reading frame*

PB Processos Biológicos

Pb par de bases

PDB *Protein Data Bank*

Phe Fenilalanina

RNA Ácido ribonucleico

rRNA RNA ribossômico

RSEM *RNA-Seq by Expectation Maximization*

SBS Sequenciamento por Síntese

SDS Dodecilsulfato de sódio

Ser Serina

Thr Treonina

Tyr Tirosina

Val Valina

Zn Zinco

λ – Comprimento de onda

$\lambda_{\text{máx}}$ – Comprimento de onda de absorção máxima

Sumário

CAPITULO 1 – CONSIDERAÇÕES GERAIS	24
1 REFERENCIAL TEÓRICO	24
1.1 <i>Amynthas gracilis</i>	24
1.2 Pigmentos respiratórios	25
1.3 Hemoglobinas.....	26
1.4 Hemoglobinas extracelulares gigantes	28
2 OBJETIVOS	31
2.1 Objetivo Geral	31
2.2 Objetivos específicos.....	31
CAPITULO 2 – ANOTAÇÃO FUNCIONAL DO TRANSCRIPTOMA DE <i>Amynthas gracilis</i>	32
RESUMO.....	32
1 INTRODUÇÃO	33
2 MATERIAIS E MÉTODOS	34
2.1 Sequenciamento de nova geração/ plataforma Illumina	34
2.2 Extração de RNA total de <i>A. gracilis</i>	37
2.3 Quantificação do RNA total	37
2.4 Purificação das amostras de RNA total.....	38
2.5 Quantificação da amostra de RNA mensageiro (mRNA)	39
2.6 Confecção das moléculas de cDNA dupla fita.....	39
2.7 Purificação das moléculas de cDNA dupla-fita	40
2.8 Preparo da biblioteca de cDNA.....	41
2.9 Quantificação da biblioteca de cDNA por PCR em tempo real	42
2.10 Desnaturação e diluição da biblioteca de cDNA	43
2.11 Análise das sequências	44
2.12 Anotação funcional do transcriptoma de <i>A. gracilis</i>	44
3 RESULTADOS E DISCUSSÃO	45

3.1	Extração e quantificação do RNA total de <i>A. gracilis</i>	45
3.2	Purificação e quantificação do mRNA.....	46
3.3	Quantificação do cDNA	46
3.4	Quantificação da biblioteca de cDNA.....	47
3.5	Análise das sequências e montagem do transcriptoma	49
3.6	Anotação funcional	53
4	CONCLUSÕES	66
CAPITULO 3 – IDENTIFICAÇÃO DAS SEQUÊNCIAS DE HbAg E CONSTRUÇÃO DO SEU MODELO MOLECULAR		67
RESUMO.....		67
1	INTRODUÇÃO	68
2	MATERIAIS E MÉTODOS	69
2.1	Identificação das sequências de HbAg.....	69
2.2	Construção do complexo proteico HbAg.....	70
3	RESULTADOS E DISCUSSÃO.....	70
4	CONCLUSÕES	87
CAPITULO 4 CARACTERIZAÇÃO BIOFÍSICA INICIAL DO MONÔMERO-D NATIVO DE HbAg.		89
Resumo		89
1	INTRODUÇÃO	90
2	MATERIAIS E MÉTODOS	90
2.1	Análises espectroscópicas	90
2.1.1	Espectroscopia de absorção UV-VIS.....	90
2.1.2	Espectroscopia de Fluorescência	92
2.2	Extração e purificação da HbAg	94
2.3	Dissociação e purificação do monômero- <i>d</i> de HbAg.....	95
2.4	Análises de Espectroscopia UV-VIS da subunidade- <i>d</i> de HbAg em função do pH e diferentes concentrações de SDS.	96

2.5	Análises de Fluorescência e Espalhamento de Luz (LSI) da subunidade- <i>d</i> de HbAg	96
3	RESULTADOS E DISCUSSÃO	97
3.1	Dissociação e purificação do monômero- <i>d</i> de HbAg.....	97
3.2	Análises espectroscópicas do monômero- <i>d</i> de HbAg em função do pH.	99
3.3	Interação do monômero- <i>d</i> de HbAg com o surfactante aniônico Dodecilsulfato de sódio (SDS) através de análises de Espectroscopia UV-VIS, Fluorescência e Espalhamento de Luz (LSI)	107
4	CONCLUSÕES	112
	CONCLUSÃO GERAL E PERSPECTIVAS FUTURAS	114
	REFERÊNCIAS.....	116

CAPITULO 1 – CONSIDERAÇÕES GERAIS

1 REFERENCIAL TEÓRICO

1.1 *Amyntas gracilis*

As minhocas são anelídeos muito utilizados em estudos de ecotoxicologia. Consideradas como organismos modelo, são utilizadas como bioindicadores de contaminação do solo. Elas transitam pelas diferentes fases do solo (líquida, gasosa e sólida) e modificam a dinâmica dos nutrientes, agindo na distribuição e viabilidade desses nutrientes para outros organismos que habitam o mesmo ambiente, além de afetar outros elementos estruturais e físicos do solo. De forma geral, a capacidade de suportar níveis de contaminação confere uma característica de grande plasticidade nesses indivíduos (PARELHO et al., 2018).

A espécie de minhoca *Amyntas gracilis* (classe Oligochaeta, família Megascolecidae), também conhecida como minhoca puladeira, é uma das espécies mais comuns em diversas áreas do território brasileiro (PARELHO et al., 2018). Encontra-se em regiões mais superficiais do solo, como na interfase entre a liteira e solo presente em zonas de florestas ou nas primeiras camadas do solo. Possui origem asiática e em regiões do Sul e Sudeste do Brasil pode ser encontrada em áreas frescas e úmidas. São encontradas preferencialmente em zonas agrícolas, mas também em florestas nativas, podendo causar alterações nas propriedades físicas do solo e ciclagem de nutrientes (DE SILVA et al., 2010). No Vale do Ribeira, SP, é muito comum encontrá-la entre as culturas de bananas e pupunha (CREMONESI et al., 2021).

A exposição das minhocas a solos contaminados com metais pesados pode causar mudanças morfológicas nestes indivíduos. Em exemplares de *A. gracilis* que foram expostos a metais pesados observaram-se mudanças conformacionais na epiderme das minhocas, tanto na sua espessura quanto no número de células secretoras. Essas mudanças conformacionais não são constitutivas e sim condicionadas as modificações do ambiente em que esses indivíduos se encontram (CUNHA et al., 2011).

Certamente alterações em outros níveis podem ser detectados nestes anelídeos, como por exemplo, alterações a nível molecular que podem levar a mudanças em moléculas como a HbAg. Informações a respeito de aspectos biológicos ou morfológicos de *A.*

gracilis são escassos na literatura; pouco se sabe sobre a caracterização de sua hemoglobina e diversas pesquisas têm sido realizadas com o intuito de caracterizá-la.

1.2 Pigmentos respiratórios

O aumento da concentração de O₂ atmosférico ao longo do tempo agiu com um fator de pressão de seleção que levou ao surgimento dos pigmentos respiratórios e acredita-se que esse evento ocorreu durante a fase Pré-cambriana. (ALVAREZ-CARREÑO; BECERRA; LAZCANO, 2016). O desenvolvimento de um sistema circulatório com proteínas carreadoras de oxigênio, estando essas proteínas dentro de componentes celulares ou dispersas em fluídos sanguíneos, o aumento de tamanho dos corpos dos animais, a resistência à auto oxidação dessas proteínas carreadoras de oxigênio e sua capacidade de resistir ao danos causados pelas espécies de oxigênio reativas contribuíram para evolução dos metazoários (KURTZ, 1999; MANWELL, 1960).

Os pigmentos respiratórios podem ser classificados em três categorias: as hemoglobinas (Hbs), hemocianinas (Hcs) e as hemeritrinas (Hrs), sendo que as Hcs são subdivididas em Hc de moluscos e Hc de artrópodes (COATES; DECKER, 2016; TERWILLIGER; TERWILLIGER; SCHABTACH, 1976).

Começando pelas Hbs é sabido que as diversas Hbs descritas variam desde formas monoméricas até complexos gigantes, como o caso da proteína alvo deste trabalho HbAg. Em humanos elas se encontram abrigadas nas hemácias que garantem uma estabilidade para a molécula contra a proteólise, baixa pressão osmótica e também previne perdas durante o processo de filtração nos rins (COATES; DECKER, 2016). Sua estrutura consiste em tetrâmeros formados pela união de dímeros das cadeias α e β totalizando um tamanho em torno de 64 kDa (FERMI et al., 1984). Cada uma das cadeias α e β possui um grupo Heme capaz de ligar-se às moléculas de O₂, o que leva a uma capacidade de ligação de 4 moléculas de O₂ por molécula de Hb (FERMI et al., 1984). Já os complexos de Hb gigantes extracelulares encontrados em anelídeos tem uma elevada capacidade de ligação ao O₂ totalizando um número de até 144 moléculas de O₂ por molécula de Hb gigante (COATES; DECKER, 2016; WEBER; VINOGRADOV, 2001).

Diferentemente das Hbs as Hcs possuem um centro dicobre coordenado por uma histidina proximal que se liga às moléculas de O₂ (COATES; DECKER, 2016). As Hcs de artrópodes são formadas por hexâmeros com 72 kDa, onde as subunidades contêm três domínios proteicos distintos. Já as subunidades funcionais das Hcs dos moluscos são

caracterizadas pela presença de dois domínios estruturais diferentes. A estrutura dessas moléculas é formada por decâmeros cilíndricos que vem do arranjo de conjuntos formados pela união de sete a oito subunidades funcionais (COATES; DECKER, 2016). As Hcs além das funções básicas de transporte de oxigênio possuem um papel importante no sistema imunológico atuando no combate de infecções causadas por microrganismos (COATES; DECKER, 2016).

Por último as Hrs são proteínas que não possuem o grupo prostético Heme como um ligante nas suas cadeias polipeptídicas, ao invés disso existe a presença de um centro diferido (COATES; DECKER, 2016). Esse centro contendo 2 átomos de Ferro é circundado por 5 histidinas, um ácido aspártico e um resíduo de ácido glutâmico e o oxigênio irá ligar-se entre os dois átomos de Fe (COATES; DECKER, 2016). As Hrs subdividem-se em cHrs e mioHrs, na qual as cHr localizam-se no sistema circulatório dentro de eritrócitos e as mioHrs encontram-se nos tecidos musculares (COSTA-PAIVA et al., 2017; MANWELL, 1960). Especificamente as cHrs são formadas estruturalmente por octâmeros de massa molecular de 108 kDa, enquanto as mioHrs são estruturas monoméricas (COSTA-PAIVA; COATES, 2020).

Neste trabalho foram conduzidos estudos com uma hemoglobina de uma espécie de minhoca chamada *A. gracilis* e dados recentes mostraram que os anelídeos possuem a ocorrência das três categorias de pigmentos respiratórios (COSTA-PAIVA; COATES, 2020).

1.3 Hemoglobinas

Hemoglobinas são proteínas de importância vital amplamente encontradas na maioria dos filos pertencentes ao reino animal. O número de cadeias que as compõem divergem sendo encontradas proteínas monoméricas (mioglobina presente em tecidos musculares) até os complexos proteicos gigantes encontrados em espécies de anelídeos (WEBER; VINOGRADOV, 2001). As Hbs estão distribuídas em diferentes locais nos organismos, desde citoplasmas até tecidos ou alojadas dentro de corpos celulares, podendo estar ou não dispersas nos fluidos sanguíneos (WEBER; VINOGRADOV, 2001).

As Hbs também são encontradas em associação com plantas leguminosas, conhecidas como LegHbs (BECANA et al., 2020); em associação com tecido nervoso onde são denominadas como neuroglobinas (BRACKE; HOOGEWIJS; DEWILDE, 2018) e até mesmo em cianobactérias, as cianoglobinas (TRENT et al., 2004).

Tomando como exemplo a hemoglobina (Hb) humana, estruturalmente essa proteína é um tetrâmero com massa molecular (MM) de 64 kDa, composto por 2 cadeias α e β que se unem formando dímeros $\alpha_1\beta_1$ e $\alpha_2\beta_2$ (**Figura 1A**) (VOET; VOET, 2013). É comum à todas as estruturas das Hbs um enovelamento que resulta na presença de seis a oito α -hélices conectadas por *loops* curtos (BOLOGNESI et al., 1997). O grupo Heme encontra-se numa região altamente hidrofóbica. O grupo Heme é uma protoporfirina com um átomo de Ferro que possui seis ligações coordenadas, quatro destas estão ligadas a própria porfirina e as outras duas se ligam a um resíduo de Histidina (His) presente nas globinas e a outra coordenação fica livre para ligar-se à molécula de O_2 (**Figura 1B**) (NELSON; COX, 2014). Quando o átomo de Fe está em seu estado ferroso (Fe^{2+}) a ligação com o O_2 ocorre de forma reversível, já quando o grupo Heme sofre oxidação o Fe passa para o estado férrico (Fe^{3+}) e nesse estado a ligação com o O_2 não ocorre (NELSON; COX, 2014).

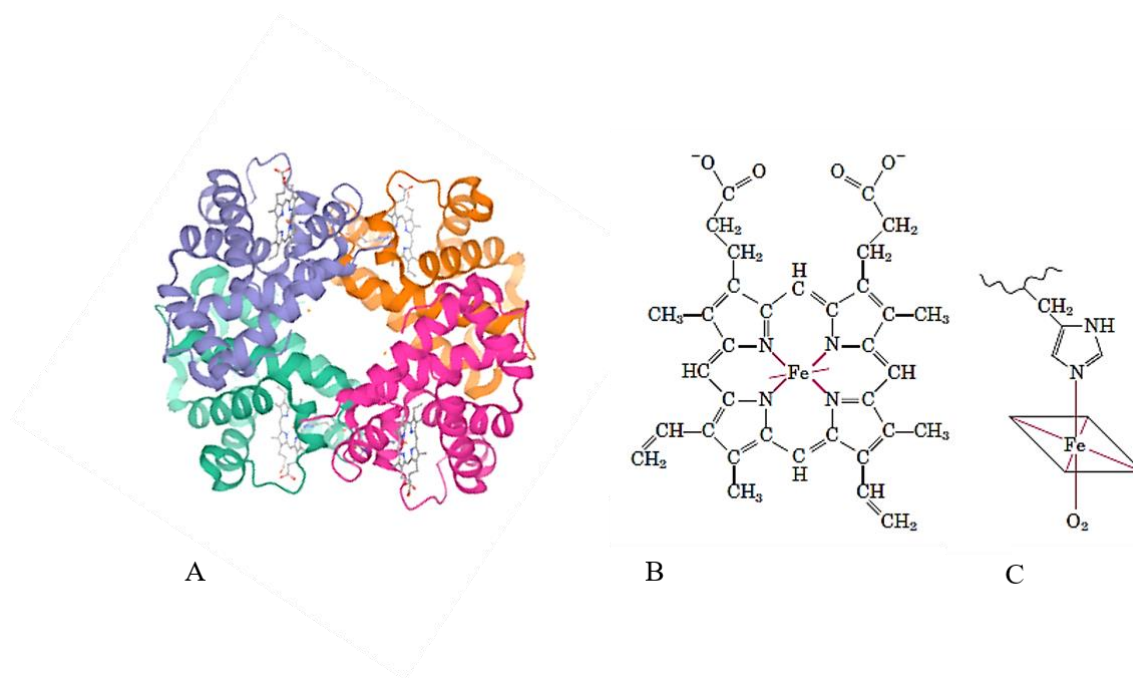


Figura 1: Representação gráfica da estrutura da Hb humana e grupo Heme. (A) Estrutura da Hb humana (código PDB: 2HHB). Tetrâmero formado por 2 cadeias α e duas cadeias β ; em cada uma das cadeias há a presença de um anel porfirínico (grupo Heme). (B) Grupo Heme: porfirina com um átomo de Fe no centro que possui seis pontos de coordenação; essas ligações estão representadas na figura por traços vermelhos, nota-se que quatro coordenações são realizadas na própria porfirina e duas delas (C) correspondem respectivamente à ligação com o N do resíduo His presente em uma das α -hélices da Hb e outra coordenação fica livre para ligar-se à molécula de O_2 .

Fonte: A: FERMI et al., (1984); B e C: NELSON e COX (2014).

A ligação do O₂ ao grupo Heme confere uma mudança conformacional na estrutura das Hbs. As Hbs apresentam dois estados conformacionais diferentes relacionados à ligação com o O₂. No estado T a Hb encontra-se na forma deoxi e no estado R a Hb encontra-se na sua forma oxi, a ligação com o O₂ é favorecida quando a Hb passa do estado T para o R, sendo que a ligação estabiliza o estado R. A transição entre esses dois estados conformacionais levam a rotação e translação dos dímeros $\alpha\beta$ um em relação ao outro. No estado T a porfirina encontra-se levemente pregueada, dessa forma o Fe é projetado para o lado da His proximal, enquanto no estado R ao ocorrer a ligação com o O₂ o grupo Heme adquire uma posição planar deslocando a posição da His proximal e da hélice F onde está ligada (PERUTZ, 1970; PERUTZ et al., 1998).

O efeito alostérico é definido como a modulação funcional de um sítio ativo ocasionado pela ligação de um ligante de um outro sítio ativo na mesma molécula (RONDA; BRUNO; BETTATI, 2013). Esse efeito é encontrado nas Hbs e explica a cooperatividade encontrada nessas moléculas, ou seja, quando uma molécula de O₂ se liga no grupo Heme de uma das subunidades, a ligação do O₂ à outras porfirinas presentes em outras subunidades é favorecida, existe uma tendência das cadeias de globinas passarem do estado T para R favorecendo dessa forma a entrada do ligante (VOET; VOET, 2013).

1.4 Hemoglobinas extracelulares gigantes

As Hbs extracelulares gigantes são complexos proteicos com MM variando de 3000 a 4000 kDa, possuem 180 cadeias polipeptídicas, sendo 144 caracterizadas como globinas as quais possuem o grupo prostético Heme em sua estrutura e, outras 36 cadeias correspondem aos *linkers* (BACHEGA et al., 2015; ROYER et al., 2006; ZAL; ROUSSELOT, 2014). As funções atribuídas a essa classe de Hbs estão relacionadas à estrutura das globinas, enquanto que os *linkers* desempenham um papel estrutural na montagem do complexo proteico (CHABASSE et al., 2006a; NEGRISOLO et al., 2001; ROYER et al., 2006; ROYER; OMARTIAN; KNAPP, 2007).

As Hbs gigantes recebem denominações distintas na literatura, podendo encontrá-las como eritrocruorinas (VOET; VOET, 2013) até proteínas HBL-Hb, uma abreviatura da expressão em inglês *Hexagonal bilayer hemoglobin* (WEBER; VINOGRADOV, 2001). A classificação HBL-Hb remete à maneira como o complexo proteico é formado. As globinas estão dispostas em estruturas dodecaméricas, chamadas de *cap*, com estequiometria $[abc]_3[d]_3$, onde o trímero $[abc]$ é conectado por ligações dissulfeto e a cadeia *d* interage a

partir de interações eletrostáticas, as quais podem ser consideradas mais fracas quando comparadas às ligações dissulfetos (BACHEGA et al., 2015; ROYER et al., 2006). A união dos *linkers* formam heterotrímeros (L1L2L3) e se unem ao *cap* formando uma estrutura conhecida como protômero com estequiometria de $[abc]_3[d]_3(L1L2L3)$. O complexo proteico inteiro é então formado por uma dupla camada hexagonal, onde cada uma dessas camadas é composta por 6 protômeros (BACHEGA et al., 2015; ROYER et al., 2006; WEBER; VINOGRADOV, 2001). A **Figura 2** abaixo representa a formação das HBL-Hb como foram descritas acima.

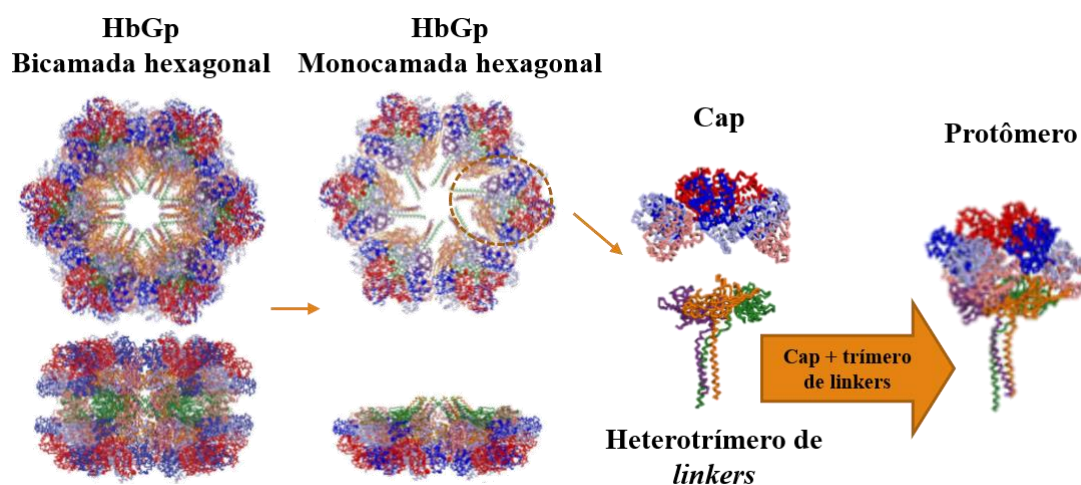


Figura 2: Representação da estrutura de HbGp, Hb homologa à HbAg. Da esquerda para a direita está representada a vista frontal da estrutura da unidade biológica formando uma bicamada hexagonal com 12 protômeros; seguindo temos a monocamada hexagonal de HbAg vista superior e lateral e o protômero. A monocamada hexagonal é composta por 6 protômeros. O protômero é a união entre o *cap* formado pelas globinas com estequiometria $[abc]_3[d]_3$ e um heterotrímero de *linkers* (L1L2L3). As globinas foram representadas pelas cores azul marinho (globina *a*); rosa claro (globina *b*); cinza (globina *c*) e vermelho (globina *d*). Os *linkers* foram representados pelas cores verde (L1), roxo (L2) e laranja (L3).

Fonte: BACHEGA (2013). Adaptado pela autora (2022).

Diferentes espécies de anelídeos apresentam essas proteínas, e como exemplos podem ser citadas: a HbLt extraída do oligoqueta terrestre *Lumbricus terrestris*, a HbGp presente na espécie de minhocoço *Glossoscolex paulistus* e HbAm que advém do poliqueta marinho *Arenicola marina*. As Hbs dessas três espécies de anelídeos têm sido estudadas visando sua caracterização estrutural e biofísica (MARTIN et al., 1996; SANTIAGO et al.,

2008). O alvo de estudo desta tese de Doutorado foi HbAg, Hb proveniente da espécie de minhoca *Amyntas gracilis*.

Ao longo dos anos, as Hbs gigantes vem sendo alvo de estudos em grupos de pesquisa pelo mundo (ROUSSELOT et al., 2006; ZAL; ROUSSELOT, 2014; ZIMMERMAN et al., 2017) visando elucidar sua estrutura assim como sua utilização para fins biotecnológicos, como por exemplo, substituto sanguíneo, alvo de grande destaque para HbAm (CARVALHO et al., 2011a; ROUSSELOT et al., 2006; VARNEY et al., 2021).

Algumas propriedades presentes nestas moléculas fazem com que elas sejam ótimas candidatas para o desenvolvimento de carreadores de O₂/substitutos sanguíneos. São proteínas que formam seus oligômeros em meio extracelular; as características de ligação e liberação ao O₂ são similares a Hb humana porém com peso molecular em torno de 50 vezes maior (maior carreamento de O₂, capacidade de ligação a até 156 moléculas de O₂ por unidade biológica de Hb gigante); sua funcionalidade é independente de moléculas secundárias, tais como 2,3-DPG; sua função também não depende de tratamentos adicionais ou químicos; alta estabilidade oxidativa e térmica, mantendo-se funcional em faixas de temperatura que variam de 4°C a 30°C; ausência de efeito vasoconstritor; propriedades antioxidantes (VARNEY et al., 2021; ZAL; ROUSSELOT, 2014).

A empresa de origem francesa Hemarina possui produtos comerciais a partir de HbAm que visam a utilização desta proteína como um substituto sanguíneo. Em seu portfólio existem produtos de fins terapêuticos voltados para área médica, cujo objetivo é o suprimento de O₂ para sistemas que necessitam de oxigenação (ZAL; ROUSSELOT, 2014). Alguns exemplos desses produtos são: o carreador de oxigênio universal HEMOXYCarrier® e a solução para preservação de órgãos pré-transplante HEMO₂life® que encontram-se disponíveis no mercado (LE MEUR et al., 2020; LE PAPE et al., 2015). Alguns outros produtos da empresa estão em fase de desenvolvimento e são destinados para diferentes fins: como o HEMOXCell® desenvolvido para promover a melhoria da oxigenação em meios de cultura; o HEMBoost® utilizado no processo de fermentação nas indústrias farmacêuticas e de alimentos; a solução terapêutica HEMHealing® destinada ao tratamento de feridas e o HEMDental-Care® destinado a área de tratamento dentário.

Além da notória contribuição que as Hbs gigantes trazem para a pesquisa na área da saúde existem outras áreas com potencial de exploração biotecnológica. Estudos preliminares envolvendo HbAg como matriz de um biossensor eletroquímico mostraram que a molécula é sensível quantitativamente a Cu, Cd e Zn (dados não publicados). Além

disso a detecção do metal Cu em solução também pode ser monitorada por absorção óptica, que coloca essa biomolécula como promissora em pesquisas para o desenvolvimento de um biossensor óptico para detecção de metais pesados (SOUZA, 2018) .

2 OBJETIVOS

2.1 Objetivo Geral

Estudar o complexo proteico HbAg a partir do sequenciamento de suas cadeias, assim como realizar uma anotação geral das proteínas identificadas no transcriptoma da minhoca *A. gracilis*.

2.2 Objetivos específicos

- Realizar a anotação funcional do transcriptoma de *A. gracilis*;
- Identificar as sequências das cadeias que compõe a HbAg;
- Construir um modelo do complexo proteico HbAg a partir das sequências preditas das cadeias de HbAg;
- Caracterizar biofisicamente o monômero-*d* de HbAg a partir de análises espectroscópicas.

CAPITULO 2 – ANOTAÇÃO FUNCIONAL DO TRANSCRIPTOMA DE *Amyntas gracilis*

RESUMO

O avanço nas técnicas de sequenciamento de nova-geração assim como o desenvolvimento de novas metodologias para análise desses dados permitiu a descoberta de novas informações que auxiliam em estudos genéticos para as espécies cujo genoma ainda não foi sequenciado. Dados a respeito da espécie de minhoca *Amyntas gracilis* são escassos e este capítulo teve como objetivo sequenciar e realizar uma anotação geral do transcriptoma desta espécie. Para tanto foi construída uma biblioteca de cDNA e o sequenciamento foi realizado em um equipamento MiSeq da Illumina a partir de duas corridas, cada uma contendo 600 ciclos de sequenciamento através da técnica *pair-end*. Na ausência de um genoma de referência, a montagem do transcriptoma foi realizada partir da técnica *de novo* com o auxílio do programa Trinity 2.0.6 e, a anotação funcional foi feita utilizando o programa OmicsBox. Foram obtidos 477.632 transcritos, das quais 260.016 sequências foram preditas como ORF e 187.866 são sequências não redundantes. O maior número de *hits* encontrados no transcriptoma de *A. gracilis* teve correspondência com as sequências das espécies *C. teleta* e *H. robusta*, muito possivelmente devido ao grande volume de dados destas espécies disponibilizados nos bancos de dados. Dentro dos genes mais expressos encontram-se sequências de proteínas com função estrutural e destacamos a expressão concomitante dos pigmentos respiratórios Hemeritrina e HB gigante. Foram identificadas algumas proteases, tais como, a lumbrokinase cuja a presença ocorre apenas em anelídeos. Uma série de enzimas foram anotadas no transcriptoma aqui reportado, enzimas estas que podem estar relacionadas com as atividades biológicas executadas pelas minhocas no meio ambiente.

1 INTRODUÇÃO

Dentro da classe dos anelídeos, as minhocas desempenham papéis importantes em diferentes áreas de atuação. Junto com a microfauna que habita os solos agem nos processos de fertilidade e estruturação dos solos, sendo utilizadas como bioindicadores de qualidade e contaminação ambiental (ORDOÑEZ-ARÉVALO et al., 2018). Além disso, as minhocas apresentam algumas classes de proteínas que possuem potencial de exploração biotecnológica não apenas na área ambiental como também na área médica. Segundo a medicina Chinesa as minhocas eram prescritas por possuir ação antiinflamatória e diurética, e com o avanço nas pesquisas, algumas proteases presentes nestes organismos foram caracterizadas com atividade antitrombótica, sendo assim, com potencial uso no tratamento de doenças cardiovasculares (WANG et al., 2019).

Amyntas gracilis (classe Oligochaeta, família Megascolecidae) é uma espécie exótica no Brasil, de origem asiática, bastante encontrada em áreas frescas e/ou úmidas das regiões Sul e Sudeste do país. No Vale do Ribeira, SP, é muito comum encontrá-la entre as culturas de banana e pupunha (DE SILVA et al., 2010). Apresenta tamanho pequeno e é classificada como espécie epi-endogeica pois se alimenta da camada superficial do solo, liteira (DEMETRIO et al., 2019).

Para aprofundar-se nos estudos de caracterização de proteínas, um fator importante a ser considerado é conhecer sua estrutura primária, ou seja, identificar a sequência de aminoácidos que codifica determinada molécula. Neste sentido, uma das metodologias difundidas atualmente é a obtenção de sequências de proteínas preditas a partir de dados de sequenciamento, montagem e anotação de transcriptomas.

Com o avanço das técnicas de sequenciamento de nova geração, um grande número de dados referentes a anotação de genomas completos ou parciais e de transcriptomas foram gerados. Especificamente relacionado às espécies de minhocas é encontrado um volume de dados maior e mais consistente para as espécies do gênero *Eisenia* (LIU et al., 2020; MIKAMI et al., 2015; PONESAKKI et al., 2017; ZWARYCZ et al., 2016). Esses dados podem ser explorados na elucidação de identificação das sequências de proteínas homólogas. Até o momento são encontradas algumas sequências depositadas em banco de dados pertencentes a espécie *A. gracilis* (BELATO et al., 2019), porém ainda não foi registrada uma anotação completa de seu genoma ou transcriptoma. Este trabalho traz a primeira anotação de um transcriptoma de *A. gracilis*, além de identificar as sequências que codificam as subunidades que compõem a HbAg.

Sendo assim neste capítulo foi apresentada e discutida uma anotação geral das proteínas identificadas no transcriptoma de *A. gracilis*, incluindo algumas classes de proteínas que apresentam potencial de utilização biotecnológica.

2 MATERIAIS E MÉTODOS

2.1 Sequenciamento de nova geração/ plataforma Illumina

O surgimento das técnicas de sequenciamento de DNA remontam a década de 70 com metodologias desenvolvidas por Sanger (SANGER; NICKLEN; COULSON, 1977). Alguns marcos na história da biologia molecular contribuíram para o avanço das tecnologias de sequenciamento, como a descoberta da PCR e o projeto do sequenciamento do genoma humano. A execução e concretização final do sequenciamento do genoma humano foi possível devido ao investimento de recursos no desenvolvimento de novos sequenciadores, o que permitiu a diminuição de custos e tempo na obtenção de dados deste projeto (VAN DIJK et al., 2014). Os sequenciadores de nova geração (NGS) revolucionaram a forma de obter sequências de DNA e RNA. A terminologia *next generation sequencing* descrevem os métodos de sequenciamento de alto rendimento capaz de obter milhões de sequências por corrida (LEVY; BOONE, 2019).

Diferentemente do sequenciamento de Sanger nos métodos NGS, os fragmentos de DNA ou cDNA são sequenciados a partir de bibliotecas depositadas em uma lâmina de fluxo, não havendo a necessidade de clonagem bacteriana dos fragmentos. Milhares de reações são produzidas paralelamente com a produção dos sinais que identificam as bases, sem a etapa de verificação em gel de eletroforese (VAN DIJK et al., 2014).

O sequenciador 454 da Roche foi o primeiro aparelho de sequenciamento NGS lançado em 2005. Com método de detecção por piro-sequenciamento utilizando amplificação por *beads* seguido por PCR em emulsão, quando foi lançado tinha capacidade de geração de *reads* com comprimento de 110 pares de bases (pb), passando por atualizações em 2008 nas quais acarretaram a capacidade de gerar fragmentos de 700 pb (LIU et al., 2012; MARGULIES et al., 2005). Em 2007 foram lançadas as plataformas Solexa/Illumina e o sequenciador SOLiD (*Sequencing by Oligo Ligation Detection*) pela antiga Applied Biosystems (atual Life Technologies). Inicialmente o SOLiD gerava *reads* curtos de 35 pb e maior rendimento comparado ao 454, cerca a 100 milhões de *reads* (VALOUEV et al., 2008); neste método os fragmentos a serem sequenciados também são

hibridizados em *beads* e amplificados via PCR e a detecção das bases é feita por sondas que emitem sinais de fluorescência, uma cor diferente para cada tipo de base (LIU et al., 2012). Em 2010 foi lançado o Ion Torrent, o seu sistema foi desenvolvido pelo mesmo desenvolvedor do sistema 454 e sua principal diferença está na detecção das bases que se baseia na variação de pH que ocorre durante o sequenciamento (VAN DIJK et al., 2014).

A plataforma utilizada neste trabalho foi a Illumina e, portanto, encontra-se abaixo uma descrição mais detalhada do funcionamento deste método de sequenciamento. O princípio do sequenciamento Illumina é o sequenciamento por síntese (SBS) (LIU et al., 2012). A **Figura 3** ilustra as etapas envolvidas no sequenciamento através da plataforma Illumina: (1) preparação da biblioteca; (2) formação dos *clusters* através de reações de PCR em ponte; (3) sequenciamento da biblioteca de cDNA/DNA e (4) alinhamento das sequências e análise dos dados. Na etapa 3, que consiste no sequenciamento propriamente dito, ocorre a formação dos *clusters*, obtidos a partir de reações de PCR em ponte. São utilizados dois tipos de adaptadores que se ligam nas extremidades dos fragmentos das bibliotecas e atuam na inserção desses fragmentos na lâmina de fluxo do sequenciador através da complementaridade das sequências dos fragmentos com os oligonucleotídeos presentes na lâmina. A lâmina de fluxo é uma pequena placa de vidro onde cada uma das linhas contém trechos de oligonucleotídeos de dois tipos distintos. A hibridização ocorre primeiramente com uns dos pares dos oligonucleotídeos que estão depositados na superfície. Através da ação da DNA polimerase é criada uma cópia desse fragmento acoplado; essa molécula é então desnaturada e a original é lavada da lâmina de fluxo. A outra ponta da fita se dobra e se liga por complementariedade no segundo tipo de adaptador, formando uma ponte de amplificação. Novamente, pela ação da DNA polimerase é gerada uma fita complementar e esse processo é repetido em sucessivas vezes, levando à formação dos *clusters*. Nucleotídeos marcados com os fluoróforos vão sendo adicionados à fita, os *clusters* são excitados por uma fonte de luz e um sinal fluorescente é emitido. O número de ciclos determina o tamanho dos *reads*. Milhares de *clusters* são lidos massivamente ao mesmo tempo. Inicialmente são lidas as bases no sentido *forward* (*read1*) e posteriormente as bases no sentido *reverse* (*read 2*) (ILLUMINA INC., 2018).

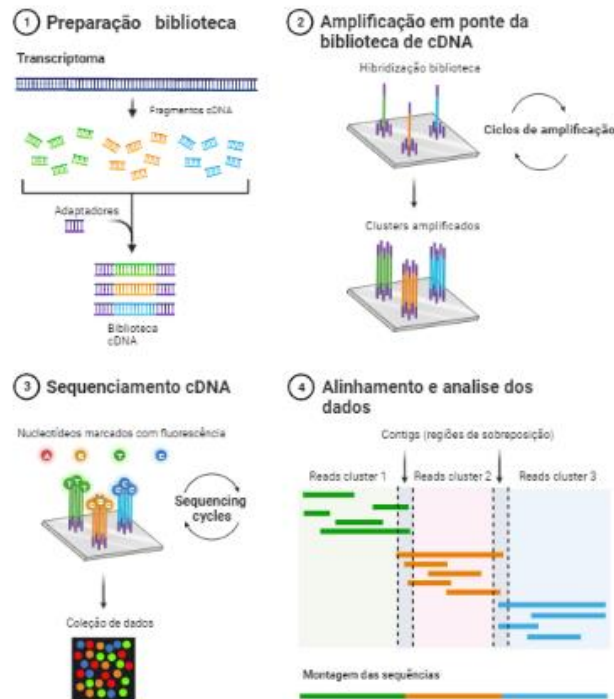


Figura 3: Representação esquemática das etapas envolvidas nos sequenciadores da plataforma Illumina. (1) Preparação da biblioteca: o material genético é fragmentado em pedaços menores; na figura pequenos fragmentos dupla-fita de cDNA estão representados nas cores verde, laranja e azul. Esses fragmentos ligam-se à adaptadores que terão a função de ligar os fragmentos à lâmina de fluxo. (2) Após hibridização da biblioteca na lâmina de fluxo ocorre a formação dos *clusters* por reações sucessivas de PCR em ponte. (3) Os nucleotídeos de adenina (a), guanina (g), timina (t) e citosina (c) são marcados com fluoróforos que emitem sinais de fluorescência com colorações distintas; a adenina está representada na cor vermelha, a guanina na cor amarela, a timina na cor verde e a citosina na cor azul. Conforme os ciclos de sequenciamento ocorrem, sinais luminosos vão sendo emitidos gerando imagens conforme descrito na figura. (4) De forma simplificada, os *reads* gerados no sequenciamento, representados por traços nas cores verde, laranja e azul serão sobrepostos através de programas de alinhamento para formação dos *contigs* e posterior montagem do genoma ou transcriptoma.

Fonte: Adaptado do “Next Generation Sequencing (Illumina)”, por BioRender.com (2022). Retirado de <https://app.biorender.com/biorender-templates>.

Esse sistema de sequenciamento utiliza dNTPs com terminadores reversíveis fluorescentes, onde outro nucleotídeo não pode se ligar imediatamente pois o fluoróforo ocupa a posição 3' OH⁻, e este sítio deve ser clivado antes da polimerização continuar, permitindo assim que o processo de sequenciamento ocorra de forma sincrônica (HEATHER; CHAIN, 2016; RUPAREL et al., 2005; VAN DIJK et al., 2018). Como consequência dessas características esse sistema de sequenciamento é considerado um

sistema com taxas baixas de erro bruto em comparação às outras tecnologias, tornando-o uma técnica altamente precisa.

2.2 Extração de RNA total de *A. gracilis*

Os indivíduos de *A. gracilis* utilizados para os procedimentos de obtenção de ácido ribonucleico (RNA) total foram criados sob condições normais, na ausência de luz e ambiente úmido, sem qualquer tipo de tratamento diferenciado; apenas foram mantidos por 24 horas sem alimentação, antes do isolamento de RNA, com a finalidade de não ter material contaminante vindo do tubo digestivo. A extração de RNA total foi realizada utilizando o Kit PureLink® RNA Mini Kit segundo as instruções do fabricante. Para tanto foram macerados dois exemplares da espécie em nitrogênio líquido. A extração foi realizada com o preparo de 10 colunas contendo cada uma delas 50 mg de tecido macerado. Após adicionar o tampão de lise contendo 1% de mercaptoetanol se processou a etapa de homogeneização. A etapa de homogeneização se caracterizou por passar a amostra em um vortéx e, em uma seringa com agulha de calibre 22 G cerca de 10 vezes. Após centrifugar a amostra a 12000 xg por 2 minutos o sobrenadante foi transferido pra um novo tubo. Foi realizado tratamento com DNase seguido das etapas de lavagem. O volume de eluição para cada coluna foi de 30 µL de água livre de RNase.

As amostras provenientes das 10 colunas foram agrupadas em único tubo para o próximo passo: o da concentração das amostras. Para tanto foi realizado o procedimento de *Clean up* do Kit PureLink® RNA Mini Kit. Dessa forma, foram adicionados à amostra unificada um volume de tampão de lise com 1% de mercaptoetanol, acrescido de um volume de etanol absoluto. Após a homogeneização, uma alíquota de 700 µL foi transferida para um tubo com filtro para ser processada com as etapas de lavagem recomendadas segundo o fabricante. O volume de eluição final foi de 50 µL de água livre de RNase.

2.3 Quantificação do RNA total

A quantificação do RNA total foi feita por NanoDrop, utilizando 1 µL da amostra no aparelho. O NanoDrop realiza a quantificação da amostra por espectrofotometria com varredura em 260 nm, pois as moléculas de RNA absorvem luz neste comprimento de onda. A integridade da amostra de RNA foi verificada em gel de agarose 1% sob condições desnaturantes.

2.4 Purificação das amostras de RNA total

Para a etapa de purificação do RNA total foi utilizado o kit RiboMinus™ Eukaryote Kit (Ambion). Foram utilizadas 4 colunas contendo 10 µg de amostra em cada uma delas. O procedimento realizado nesta etapa foi a eliminação do conteúdo de RNA ribossômico. Para tanto, foi utilizada uma sonda que se liga às moléculas de RNA ribossômico (rRNA) e, o rRNA junto a sonda se ligam a uma suspensão de microesferas (*beads*). O conjunto sonda-rRNA+ *beads* é colocado em uma estante magnética, para separação da fase aquosa e do conjunto (**Figura 4**). O sobrenadante contém a amostra de RNA livre de rRNA. Etapas de lavagens foram realizadas seguindo as instruções do fabricante e, após esse processo a amostra de RNA purificada foi concentrada utilizando o Concentrating RiboMinus™ (Ambion). O volume de eluição foi de 10 µL de água livre de RNase.

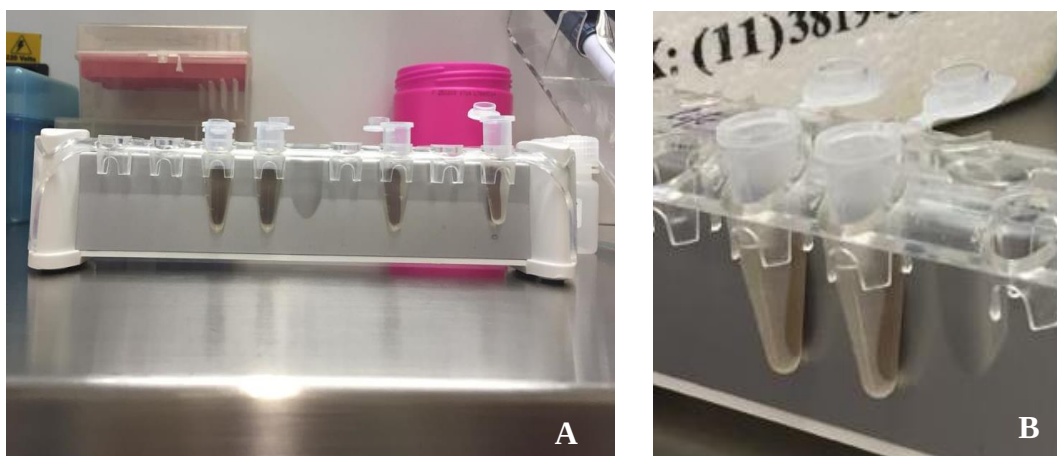


Figura 4: Purificação do RNA total. Imagens de uma das etapas envolvidas no processo de purificação das amostras de RNA total, ilustrando a formação do complexo contendo amostra+sonda RiboMinus™ + suspensão de beads e a separação deste complexo da fase aquosa do conjunto. (A) Visão frontal das colunas, contendo a amostra+sonda RiboMinus™ + suspensão de beads, posicionadas na estante magnética. (B) Detalhe da migração do complexo sonda-rRNA+ beads em direção à estante magnética.

Fonte: Elaborado pela Autora (2022).

2.5 Quantificação da amostra de RNA mensageiro (mRNA)

A quantificação foi realizada através do NanoDrop conforme descrito para quantificação das moléculas de RNA total no item 2.3. Uma segunda quantificação utilizando o kit QuantiFluor® RNA System (Promega) foi realizada pois esta análise é mais precisa comparada a quantificação por NanoDrop. O princípio da segunda quantificação se baseia na utilização de um fluoróforo que se liga somente às moléculas de RNA. A concentração da amostra foi obtida mediante comparação do conteúdo da amostra com uma curva de concentração de RNA padrão fornecida pelo kit. As concentrações de RNA padrão utilizadas foram 5000 pg μL^{-1} ; 2000 pg μL^{-1} , 200 pg μL^{-1} ; 20 pg μL^{-1} e 0 pg μL^{-1} . A amostra foi diluída 1000 x. Os procedimentos foram feitos em triplicata.

2.6 Confecção das moléculas de cDNA dupla fita

A confecção das moléculas de cDNA dupla fita foi feita utilizando-se o Superscript® Double-Stranded cDNA Synthesis kit (Invitrogen). As duas fitas foram sintetizadas separadamente em etapas sequenciais. A síntese da primeira fita de cDNA consiste na extensão da mesma a partir da molécula de mRNA molde. Para tanto foram adicionados em um microtubo 1 μL de *primers* (100 pmol μL^{-1}) e água DEPC, sendo esta mistura incubada a 70 °C por 10 minutos. Posteriormente foram adicionados o tampão de síntese da primeira fita 5X, 0.1 M de DTT, 10 mM de uma mistura de dNTPs e o mRNA molde. Nesta etapa a reação permaneceu a 45 °C por 2 minutos. Após o equilíbrio da reação o último componente adicionado foi 1 μL da enzima transcriptase reversa, deixando a mistura a 45 °C por uma hora. O *primer* escolhido para realização desta fase foi um *primer* oligo-dT, pois ele se acopla na cauda de poli-A dos mRNAs e, dessa forma, consegue capturar somente as moléculas de interesse. As quantidades descritas aqui foram as recomendadas pelo kit de síntese de cDNA. Para síntese da segunda fita foram adicionados água DEPC, tampão de síntese da segunda fita 5X, 10 mM de dNTP e as enzimas DNA ligase (10 U μL^{-1}), DNA polimerase I (10 U μL^{-1}) e RNase H (2 U μL^{-1}), todas provenientes de *Escherichia coli*. As enzimas foram os últimos componentes colocados na reação para prevenir sua degradação e assegurar sua atividade na reação.

Para a síntese da segunda fita de cDNA é necessário inserir uma sequência de nucleotídeos conhecida que se liga a ponta 5', para então, ocorrer a extensão da segunda fita. Assim, foi fornecida à reação a enzima DNA ligase que insere esse trecho de sequência

à outra ponta da molécula. A DNA polimerase adiciona os nucleotídeos à segunda fita por complementariedade e a enzima RNase é responsável por degradar a molécula de mRNA molde, dessa forma ao final do processo foram obtidas moléculas de cDNA correspondentes ao mRNA extraído. A **Figura 5** é uma representação esquemática do processo global de síntese de cDNA.

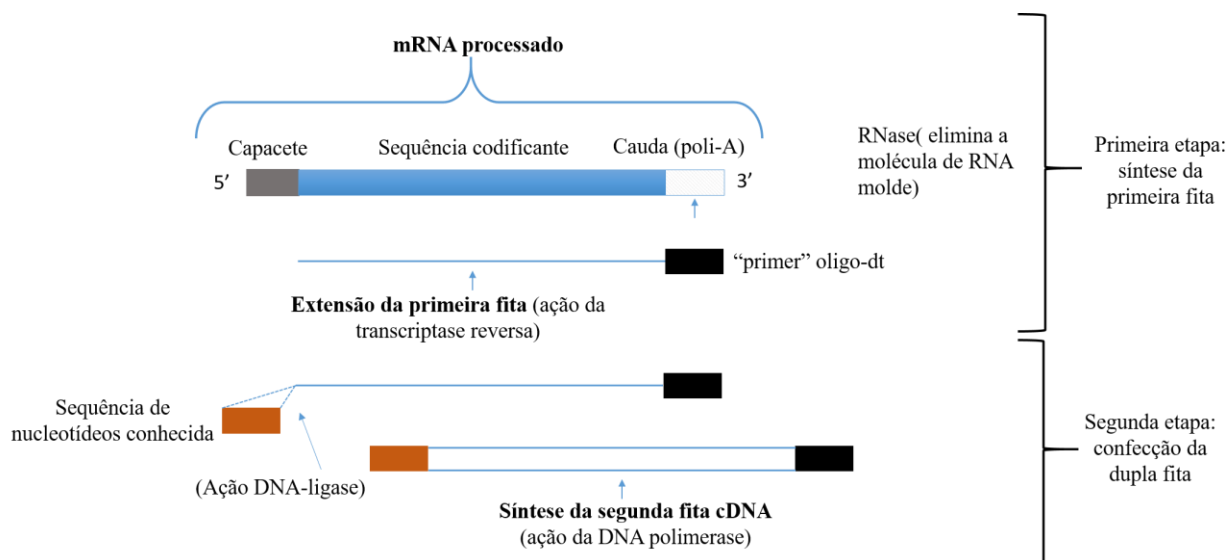


Figura 5: Confecção das moléculas de cDNA dupla fita. Representação esquemática unificada das duas etapas de síntese das moléculas de cDNA dupla fita a partir do mRNA purificado, onde o mesmo foi representado com o nome de mRNA processado. O nome mRNA processado faz uma alusão ao processamento das moléculas de RNA realizado pelos eucariotos entre as etapas de transcrição e tradução em proteínas. Na figura foram ilustrados os elementos que participam da confecção do cDNA dupla fita assim como enzimas utilizadas neste processo.

Fonte: Elaborado pela Autora (2022).

2.7 Purificação das moléculas de cDNA dupla-fita

A purificação das moléculas de cDNA foi realizada com o kit AMPure XP beads (Agencourt). Para um volume de 150 µL de reação contendo o cDNA foram adicionados 270 µL de uma suspensão de microesferas (*beads*). Esta mistura foi incubada por 5 minutos a temperatura ambiente e depois colocada em uma estante magnética. O DNA alvo, cDNA neste caso, fica aderido às *beads*, seguiu-se então para as etapas de limpeza utilizando etanol 80%. A cada etapa de limpeza o sobrenadante foi descartado. Ao final das etapas de limpeza o complexo *beads* + cDNA foi mantido em repouso para secagem do etanol. A amostra foi eluída em 60 µL de água livre de nucleases e, 55,5 µL dessa mistura foi

transferida para um novo tubo. Após passar pela etapa de purificação as amostras foram quantificadas utilizando o do Kit Quantifluor™ dsDNA System (Promega). A quantificação foi realizada de forma análoga às quantificações feitas por fluorimetria para as amostras de RNA total e mRNA. As concentrações de DNA padrão utilizadas foram 2000 pg μL^{-1} , 200 pg μL^{-1} , 20 pg μL^{-1} , 2 pg μL^{-1} e 0 pg μL^{-1} .

2.8 Preparo da biblioteca de cDNA

Para o preparo da biblioteca foi utilizado o kit Nextera XT DNA Library Prep Kit (Illumina). A primeira fase do preparo da biblioteca consiste na geração dos fragmentos que serão anexados à placa de sequenciamento. Essa fragmentação se dá pela ação de *transposomes* que se ligam a uma distância aproximada de 300 pb e clivam as moléculas de cDNA. Para o preparo desta primeira reação foram adicionados 5 μL de uma mistura contendo os *transposomes*, 10 μL de tampão e 5 μL de cDNA, essa mistura foi levada a um termociclador e mantida a 55 °C por 5 minutos. Quando a reação atingiu a temperatura de 10 °C foi adicionado imediatamente 5 μL de tampão de neutralização, para que o processo de fragmentação fosse cessado.

A etapa seguinte foi a inserção dos adaptadores às extremidades dos fragmentos gerados na fase anterior. Esses adaptadores são sequências de nucleotídeos conhecidas que irão anexar os fragmentos à placa de sequenciamento, além de promover a formação de novas moléculas através de reação de PCR em ponte, formando vários *clusters*, estes *clusters* são um conjunto de clones provenientes de um único fragmento.

Os adaptadores foram adicionados aos fragmentos por meio de uma reação de PCR, sendo assim, foram adicionados os adaptadores N701 e S517 à mistura contendo os componentes necessários para o processamento da reação, fornecido pelo kit. A reação passou por 12 ciclos de amplificação com as seguintes etapas: 95 °C por 10 segundos; 55 °C por 30 segundos; 72 °C por 30 segundos e 72 °C por 5 minutos.

A finalização da construção da biblioteca de sequenciamento de cDNA foi a etapa de limpeza utilizando uma suspensão de *beads* e etanol 80%. Foram adicionados 50 μL de produto de PCR em um microtubo e 30 μL da suspensão de *beads*. A amostra foi agitada a 1800 rpm por 2 minutos e incubada à temperatura ambiente por 5 minutos. Transcorrido o tempo de incubação, o tubo foi colocado em uma estante magnética e o sobrenadante descartado, portanto, a biblioteca ficou aderida às *beads*. Processou-se duas lavagens com

etanol 80% e ao final do processo a biblioteca foi ressuspensa em 52,5 µL de tampão fornecido pelo kit.

2.9 Quantificação da biblioteca de cDNA por PCR em tempo real

As reações de PCR em tempo real foram processadas de acordo com as instruções do kit QIAseq Library Quant Assay (Qiagen). Assim, foi realizada uma diluição seriada 10 X com DNA padrão, oferecido pelo kit de quantificação. A biblioteca foi diluída em duas concentrações, sendo uma de 1:20 e a outra 1:2000. Os componentes da reação foram: água livre de RNase, uma mix contendo o fluoróforo SYBR Green, *primers* e a diluição de DNA padrão ou da amostra da biblioteca. Além das amostras de DNA padrão e da biblioteca foi utilizado um controle, com ausência de moléculas de DNA. As reações foram aplicadas em uma placa de PCR de 96 poços em triplicata seguindo o esquema apresentado na **Figura 6**.

	1	2	3	4	5	6	7	8	9	10	11	12
A	Padrão D*. 1	Padrão D*. 1	Padrão D*. 1									
B	Padrão D*. 2	Padrão D*. 2	Padrão D*. 2									
C	Padrão D*. 3	Padrão D*. 3	Padrão D*. 3									
D	Padrão D*. 4	Padrão D*. 4	Padrão D*. 4									
E	Padrão D*. 5	Padrão D*. 5	Padrão D*. 5									
F	Controle	Controle	Controle									
G	Amostra (1:20)	Amostra (1:20)	Amostra (1:20)									
H	Amostra (1:2000)	Amostra (1:2000)	Amostra (1:2000)									

Figura 6: Placa de utilizada na quantificação da biblioteca de cDNA, por meio de PCR em tempo real. O esquema visa ilustrar a ordem de aplicação das amostras na placa. No conjunto de poços formados pelas linhas de A à linha E e colunas de 1 a 3 foram aplicadas as amostras de DNA padrão fornecidas pelo kit com diferentes diluições (D* representa diluição), onde na linha A foram aplicadas as diluições número 1, na linha B as diluições número 2, seguindo essa ordem sucessivamente até a linha E onde foram aplicadas as diluições número 5. Na linha F foram aplicadas as alíquotas contendo o controle (solução tampão). Na linha G foram aplicadas as amostras da biblioteca a ser quantificada com a diluição de 1:20 e na linha H foram aplicadas as amostras da biblioteca a ser quantificada com a diluição de 1:2000. Os poços coloridos em tons de cinza não foram utilizados, e, portanto, não possuem a presença de nenhuma amostra.

Fonte: Elaborado pela Autora (2022).

Ao final dos ciclos de amplificação, a quantificação foi feita através de uma planilha disponibilizada pela empresa fabricante do kit utilizado, onde foram colocados os dados de CT (*threshold cycle* ou *crossing point*) gerados pela análise e, a partir disso foi calculada a concentração da biblioteca. Este valor de CT é uma medida obtida quando a fluorescência da amostra é detectada acima do background de fluorescência da análise (NOLAN; HANDS; BUSTIN, 2006), graficamente representada por um ponto que cruza uma linha de base.

Para verificar o tamanho da biblioteca obtida, foi realizada uma corrida de eletroforese em gel de agarose 1%. O tamanho encontrado para o maior número de fragmentos da biblioteca foi adotado como base para o cálculo de concentração da biblioteca. A quantificação foi realizada duas vezes, com a finalidade de confirmar com exatidão a concentração da biblioteca, fator primordial para a qualidade do sequenciamento.

2.10 Desnaturação e diluição da biblioteca de cDNA

As últimas etapas a serem realizadas antes da amostra ser aplicada no sequenciador são a desnaturação e a diluição da biblioteca. Essas etapas foram realizadas utilizando o kit MiSeq System v3 (Illumina). Para a desnaturação utilizaram-se 4 nM da biblioteca. Em um microtubo foram adicionados 5 μ L da biblioteca + 5 μ L de NaOH 0,2 M. A mistura foi incubada por 5 minutos à temperatura ambiente. Após o período de incubação, a amostra foi diluída a uma concentração de 20 pM e, partir desta diluição, a amostra foi novamente diluída a uma concentração de 12 pM. Junto à amostra da biblioteca a 12 pM foram adicionados 5 μ L de PhiX previamente diluído a 12 pM. O PhiX é amostra de DNA padrão utilizado como controle da etapa de sequenciamento. Foram realizadas duas corridas de sequenciamento, cada uma com 600 ciclos e utilizando a técnica *pair-end*, onde são geradas sequências de ambas as extremidades dos fragmentos. Salientando que todas etapas que envolveram o processo de sequenciamento da biblioteca de cDNA (itens 2.2 até 2.10) foram realizados no laboratório do professor Sergio Akira Uyemura da Faculdade de Ciências Farmacêuticas da USP-Ribeirão Preto.

2.11 Análise das sequências

Após a realização do sequenciamento foram geradas leituras no formato fastq. As sequências foram trimadas utilizando o programa Trimmomatic-v0.36; para tanto foram cortadas as extremidades das sequências com *phred* inferior a 25. O restante dos parâmetros utilizados seguiu do *default* do programa. A qualidade das sequências após o processo de trimagem foi verificada utilizando o FastQC. Foram descartados também os *reads* com extensão menor que 50 pb. Para a montagem do transcriptoma foi utilizada a estratégia *de novo* com o auxílio do programa Trinity 2.0.6 (BOLGER; LOHSE; USADEL, 2014), utilizando os parâmetros recomendados pelo default do programa e um tamanho de *k-mer* de 25 pb. Os parâmetros estatísticos obtidos para montagem realizada pelo Trinity, tais como valores de N50, número total de genes, número total de transcritos foram obtidos a partir do script TrinityStats.pl do pacote de ferramentas do Trinity. As fases de leitura aberta, do inglês *open reading frame* (ORF), ou regiões codificadoras de proteínas presentes nos transcritos foram preditas com TransDecoder (<https://github.com/TransDecoder/TransDecoder/wiki>) a partir de um modelo *Ab-initio* permanecendo, apenas, com as ORFs com no mínimo 100 aminoácidos. CD-HIT-EST versão 4.6.1 (HUANG et al., 2010) foi utilizado para remover a redundância das sequências. Para estimar a abundância de transcritos, as sequências *pair-end* foram realinhadas contra transcritos montados para quantificação utilizando o software RSEM (*RNA-Seq by Expectation Maximization*) e o script do Trinity (align_and_estimate_abundance.pl) presente nas ferramentas do software Trinity. Valores de TPM (transcritos por milhão) com um corte de ≥ 100 foram considerados como altamente expressos, enquanto aqueles valores entre $3 \leq \text{TPM} < 100$ foram considerados moderadamente expressos (DUAN et al., 2019; SHADEO et al., 2007).

2.12 Anotação funcional do transcriptoma de *A. gracilis*

A anotação funcional do transcriptoma foi realizada com o auxílio do programa Blast2GO versão 3.2 (CONESA et al., 2005). A primeira etapa da anotação foi realizada utilizando a ferramenta de BLASTP contra o banco de dados de proteínas não redundante (NR) do NCBI, adotando-se como parâmetro um *E-value* mínimo de 1×10^{-6} . Concomitantemente à análise de BLAST, a anotação a partir da identificação de domínios e famílias foi efetuada através da ferramenta InterProScan (JONES et al., 2014) que rastreia

diferentes bancos de dados: Pfam (EL-GEBALI et al., 2019), SMART (LETUNIC; KHEDKAR; BORK, 2021) e SUPERFAMILY (PANDURANGAN et al., 2019). A classificação dos genes identificados em diferentes ontologias foi feita utilizando o Gene Ontology (GO) (HILL et al., 2008).

3 RESULTADOS E DISCUSSÃO

3.1 Extração e quantificação do RNA total de *A. gracilis*

Como pode ser observado na **Figura 7**, as bandas correspondentes as moléculas de RNA total extraídas de *A. gracilis* apresentam-se nítidas e intensas, o que pode estar relacionado ao rendimento da extração. Não há a presença de arraste e sim bandas únicas, o que caracteriza como perfil de bandas de RNA íntegro, sem moléculas quebradas ou degradadas. A maior parte da amostra de RNA total provavelmente é de RNA ribossômico, pois este se encontra em maior volume nas células.

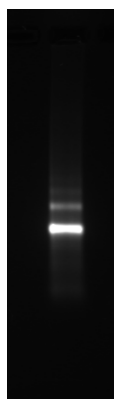


Figura 7: Gel de agarose 1% em corrida de eletroforese realizada em condições desnaturantes para verificação da extração de RNA total de *A. gracilis*. Representação da canaleta onde foi aplicada uma alíquota contendo a amostra de extração RNA total de *A. gracilis*. Presença de bandas nítidas e sem arraste; pode se inferir que a banda de maior intensidade está relacionada ao conteúdo de RNA ribossômico pois este geralmente encontra-se em maior quantidade nas células.

Fonte: Elaborado pela Autora (2022).

O comprimento de onda utilizado no NanoDrop para quantificar a extração das moléculas de RNA foi de 260 nm e, a concentração obtida foi de 2630,6 ng μL^{-1} . Analisando os valores das razões 260/280 e 260/230, foram obtidos índices de 2,05 e 2,30

respectivamente, sendo que ambos estão próximos ao valor de 2,0 que representa a pureza para moléculas de RNA, com valores ideais variando de 1,9 a 2,1. Alterações nestes índices podem estar relacionados com o pH dos componentes utilizados durante as medições e extração do RNA. Para alguns espectrofotômetros é comum o aparecimento de valores até 2,3 em condições de amostras em tampão 10 mM Tris e pH 7,5 (WILFINGER; MACKEY; CHOMCZYNSKI, 1997).

3.2 Purificação e quantificação do mRNA

Para as amostras de mRNA, além da quantificação feita através do NanoDrop também foi realizada a quantificação por fluorimetria. A quantificação medida por NanoDrop foi de 69,9 ng μL^{-1} com a razão de 260/280 de 2,02 e a razão de 260/230 de 1,74. Os índices ficaram próximos ao recomendado e inferiu-se que o mRNA encontra-se livre de contaminantes. Além disso, na curva de absorbância gerada pelo próprio NanoDrop há a presença de um único pico em 260 nm indicando que a amostra contém apenas moléculas de RNA.

A quantificação por fluorimetria revelou um resultado diferente ao encontrado no NanoDrop. A concentração foi calculada a partir da interpolação dos dados de intensidade de fluorescência e concentração, onde a curva gerada apresentou um R^2 de 0,9934. O valor encontrado para a concentração foi de 105 ng μL^{-1} . Esta quantificação é mais precisa, pois utiliza um fluoróforo que se liga às moléculas de RNA e, sendo assim, as próximas etapas do trabalho foram tomadas com base nesta quantificação.

3.3 Quantificação do cDNA

Após a síntese da dupla fita de cDNA, as amostras foram quantificadas por fluorimetria. Foi encontrada uma concentração de 2,58 ng μL^{-1} , nota-se que esta concentração é cerca de 2,5% da concentração obtida para as moléculas de mRNA. Durante todos os processos realizados ocorrem perdas do material e, portanto, é importante partir de uma concentração inicial significativa.

3.4 Quantificação da biblioteca de cDNA

Como pode ser visualizado na **Figura 8** a biblioteca gerada contém fragmentos entre 300 a 1400 pb com um pico em torno de 700 pb. Estes tamanhos ficaram de acordo com o esperado pelo kit utilizado. Tamanhos de fragmentos fora dessa faixa, principalmente abaixo de 300 pb podem levar a uma baixa qualidade no processo de sequenciamento. Logo, obteve-se uma biblioteca com fragmentos de tamanho adequados para a realização do sequenciamento.

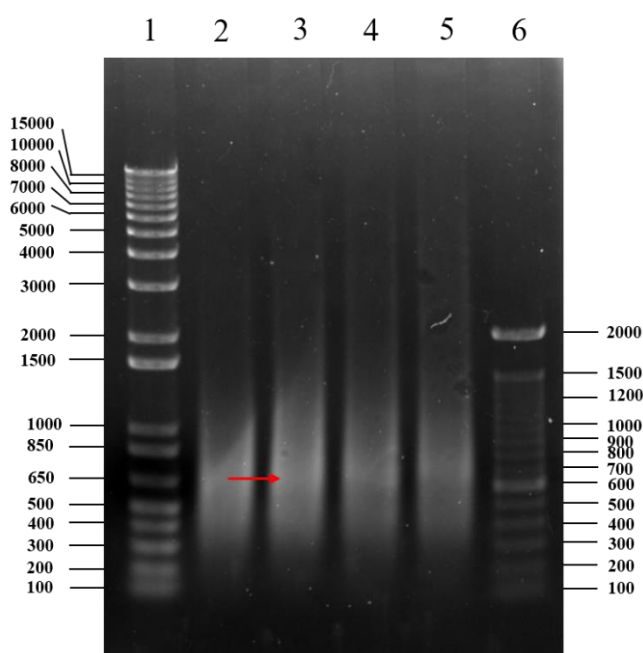


Figura 8: Gel de agarose 1% corado com brometo de etídio para verificação do tamanho da biblioteca de cDNA de *A. gracilis*. Canaletas enumeradas de 1 a 6. 1: Marcador de tamanho molecular 1 kB Plus DNA Ladder (Invitrogen™); 2 a 5: alíquotas dos produtos de PCR gerados para a biblioteca de cDNA em diferentes concentrações; 6: Marcador de tamanho molecular 100 pb DNA Ladder (Invitrogen™). A seta vermelha indica o pico onde se concentram os maiores números de fragmentos obtidos, em torno de 700 pb.

Fonte: Elaborado pela Autora (2022).

A quantificação da biblioteca de cDNA foi feita por PCR em tempo real, uma técnica muito sensível e precisa. O valor encontrado para concentração da biblioteca foi de $3,7 \text{ nM } \mu\text{L}^{-1}$ na primeira avaliação e $3,1 \text{ nM } \mu\text{L}^{-1}$ na segunda avaliação. Essa concentração foi calculada com base no tamanho de 700 pb, tamanho este encontrado em maior volume na biblioteca, dessa forma, essa concentração pode variar, pois na biblioteca existem vários tamanhos de fragmentos na faixa entre 300 e 1400 pb. A quantificação foi repetida para

confirmar a concentração em que a biblioteca se encontrava. A concentração de cDNA aplicada no sequenciador é um fator primordial para qualidade dos dados obtidos pois trabalhar com uma concentração inferior à recomendada pode levar à baixa cobertura no sequenciamento e, uma concentração superior interfere nas etapas de formação dos *clusters* e detecção das sequências, podendo até dificultar a geração das sequências.

Abaixo, na **Figura 9** estão representadas as curvas de amplificação em função dos ciclos da reação de PCR para as duas quantificações realizadas (**Figura 9A**; **Figura 9B**). Como pode ser observado os valores de CT coletados para o cálculo foram os valores que cruzam a linha de background das reações, representada na figura como uma linha na cor vermelha. Ambos os valores das duas quantificações ficaram próximos entre si, demonstrando a boa execução da técnica, característica que também pode ser observada pelo padrão em que as curvas se apresentam. As linhas que representam as curvas de amplificação das triplicatas encontram-se coincidentes umas com as outras, o que está relacionado com uma mistura efetiva dos componentes da reação e aplicação adequada das reações na placa de leitura da fluorescência emitida durante o processamento de amplificação do PCR em tempo real.

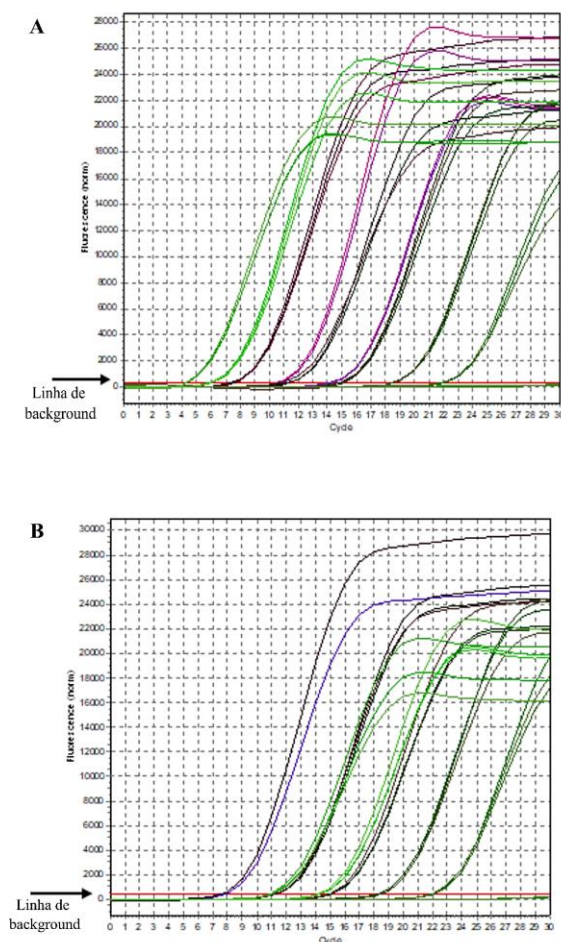


Figura 9: Representação gráfica dos ciclos de amplificação x intensidade de fluorescência emitida durante o processamento das reações de PCR em tempo real. Com a finalidade de obter dados precisos a análise foi realizada duas vezes, sendo que o gráfico 10A representa a primeira análise e o gráfico 10B a segunda análise. Foram destacadas em linha vermelha horizontal a linha de *background* das reações, onde foram coletados os valores de CT utilizados na quantificação da biblioteca de sequenciamento de cDNA de *A. gracilis*.

Fonte: Elaborado pela Autora (2022).

3.5 Análise das sequências e montagem do transcriptoma

Para as duas corridas de sequenciamento e considerando apenas os dados de boa qualidade *phred* $Q \geq 30$ foram obtidos 21,8 GB de dados totais. A espécie *A. gracilis* não possui genoma sequenciado e, portanto, utilizamos como comparativo os dados do genoma de *A. corticis*. O tamanho do genoma obtido para *A. corticis* foi de 1,2 GB (WANG et al., 2021), sendo assim, a cobertura estimada para o transcriptoma de *A. gracilis* foi de aproximadamente 18x o genoma de *A. corticis*. Porém essa estimativa comparando dados

de transcriptoma diretamente com dados de genoma não é tão informativa e/ou trivial devido as diferentes características presentes em transcriptomas e genomas. Quando se trata de transcriptoma o verdadeiro número e nível de transcrição das isoformas não é usualmente conhecido e a transcrição varia ao longo do genoma (WANG; GERSTEIN; SNYDER, 2009). Alguns autores sugerem que algumas considerações experimentais são mais importantes que parâmetros estatísticos de cobertura, sendo eles: número de *reads*, comprimento dos *reads* em pb e número de replicatas experimentais (SIMS et al., 2014). Sendo assim, para a avaliação de transcriptomas torna-se mais relevante o parâmetro de profundidade do sequenciamento que é descrito como o número total de *reads* gerados em um sequenciamento (WANG; GERSTEIN; SNYDER, 2009). A profundidade obtida neste trabalho foi em média de 51 milhões de *reads* por corrida, totalizando 102 milhões de *reads*, dado próximo ao obtido no transcriptoma de *E. andrei* com cerca de 47 milhões de *reads* em uma única corrida de sequenciamento (MIKAMI et al., 2015)

Analisando os gráficos (**Figura 10**) que avaliam a qualidade do sequenciamento por base, após a trimagem das sequências, nota-se que para ambas as corridas os valores médios de qualidade, representados com uma linha azul, encontram-se na faixa verde de *phred* Q alto (*phred* Q ≥ 30), valor este que indica um grau de precisão de 99,9%. Observa-se também que para as primeiras bases dos fragmentos a qualidade é maior, sendo que ao longo da extensão dos fragmentos sequenciados o índice de qualidade diminui. Dessa forma, ao final dos fragmentos de 300 pb a variação dos valores *phred* Q é maior, sendo esta representada pelo *boxplot* em amarelo.

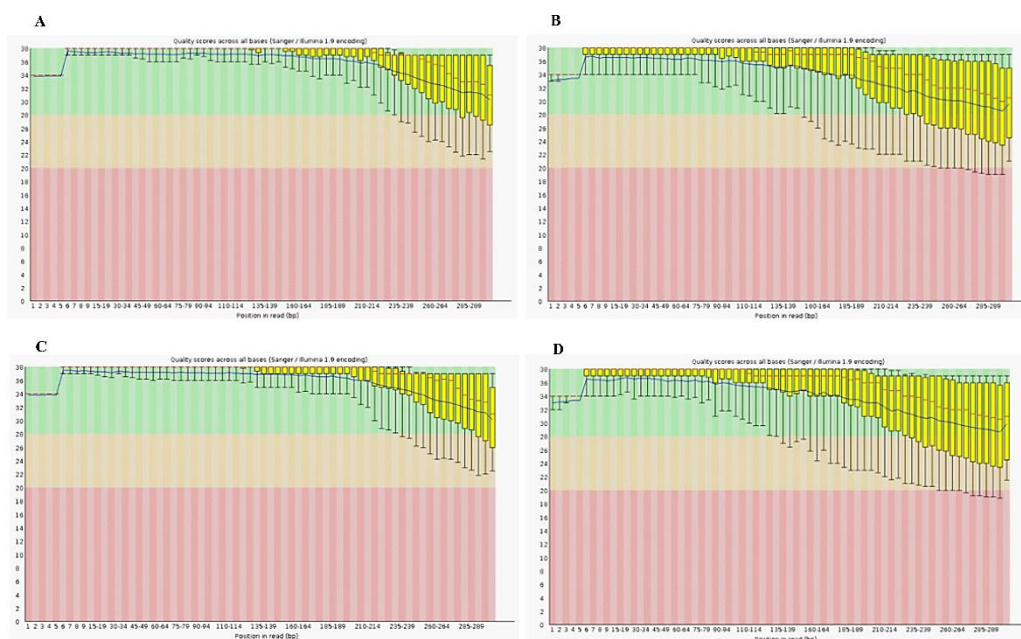


Figura 10: Gráficos gerados pela ferramenta do FastQC, que representam a qualidade do sequenciamento por base. O eixo X representa a posição dos pares de bases de 1 a 300, e o eixo Y representa os valores do *score* de qualidade *phred* Q. A técnica de sequenciamento utilizada foi *pair-end*, sendo assim os gráficos (11A e 11C) representam os *reads* no sentido *forward* e os gráficos (11B e 11D) representam os *reads* no sentido reverse. Os gráficos 11A e 11B são referentes à análise de qualidade da primeira corrida de sequenciamento e os gráficos 11C e 11D são referentes à análise de qualidade da segunda corrida de sequenciamento. A faixa verde indica *score* de alta qualidade ($Q \geq 28$), a amarela indica *score* de qualidade intermediária ($20 \leq Q \leq 28$) e a vermelha são os valores de *score* com qualidade baixa ($0 \leq Q \leq 20$). Os dados estão representados na forma gráfica tipo *boxplot*, onde a caixa amarela representa o intervalo entre 25 e 75% da distribuição dos dados e a linha vermelha que atravessa a caixa representa a mediana dos dados. As barras acima e abaixo da caixa indicam o grau de variação dos dados e a linha azul representa o valor médio de qualidade das bases na posição.

Fonte: Elaborado pela Autora (2022).

A partir de um total de 300.160.914 pb foi realizada a montagem da biblioteca de cDNA de *A. gracilis*. Foram gerados um total de 477.362 transcritos, dos quais 222.293 foram classificados como genes através da análise do programa Trinity. Para o conjunto de transcritos gerados, o valor de N50 foi de 1.050 pb, com mediana no valor de 409 pb, média de *contig* de 628,79 pb e conteúdo de CG de 40,99% (**Tabela 1**). Quando se avalia a isoforma de maior extensão por gene esses parâmetros se alteram, passando a ter um N50 668 pb, com mediana de 409 pb e média de *contig* de 576,13 pb. Devido ao volume de dados gerados acredita-se que o transcriptoma de *A. gracilis* seja extenso, o que demanda a realização de um sequenciamento de alto rendimento com a finalidade de cobrir todo o transcriptoma da espécie. A profundidade do sequenciamento também é uma característica

importante pois está diretamente correlacionada com a qualidade da montagem do transcriptoma, quanto maior a extensão dos transcritos obtidos, maior a possibilidade de identificação de sequências completas dos genes expressos. Dados do transcriptoma de *E. andrei* revelaram um total de 150.229 com um N50 de 1.855 pb e porcentagem de CG de 42,8% (MIKAMI et al., 2015). Esses dados também foram obtidos utilizando o sequenciamento da plataforma Illumina, com o sequenciador HiSeq 2000 o que difere do equipamento utilizado neste trabalho, o MiSeq. O HiSeq tem uma capacidade de rendimento e geração de dados maior que o MiSeq e por este motivo nas análises deste trabalho de doutorado foram feitas duas corridas de sequenciamento.

Tabela 1: Resultados estatísticos para a montagem do transcriptoma de *A. gracilis* realizada através do Software Trinity.

Parâmetros estáticos globais	
Número total de “genes” gerados pelo Trinity	222 293
Número total de transcritos gerados pelo Trinity	477 362
Porcentagem de CG%	40,99
Parâmetros baseados na análise de todos os transcritos	
<i>Contig</i> N10	3786 pb*
<i>Contig</i> N20	2 744 pb*
<i>Contig</i> N30	1 895 pb*
<i>Contig</i> N40	1 399 pb*
<i>Contig</i> N50	1 050 pb*
Mediana <i>contig</i>	417 pb*
Média <i>contig</i>	628,79 pb*
Total de bases montadas	300 160 914
Parâmetros baseados na isoforma de maior extensão por “gene”	
<i>Contig</i> N10	2 381 pb*
<i>Contig</i> N20	1 581 pb*
<i>Contig</i> N30	1 126 pb*
<i>Contig</i> N40	846 pb*
<i>Contig</i> N50	668 pb*
Mediana <i>contig</i>	409 pb*
Média <i>contig</i>	576,13 pb*
Total de bases montadas	128 069 757

(pb*): pares de bases

Fonte: elaborada pela Autora (2022).

3.6 Anotação funcional

A anotação funcional do transcriptoma de *A. gracilis* revelou que dos 477.362 transcritos foram identificadas 260.016 ORFs das quais 187.866 sequências são não redundantes, sendo que 77.571 destas tiveram correspondência com sequências depositadas no banco de dados NR do NCBI através da análise de BLAST. As sequências anotadas a partir da ferramenta Gene Ontology (GO) totalizaram 59.881 e o número de proteínas anotadas e validadas pelo Blast2GO foi de 92.939.

A **Figura 11** representa as espécies que tiveram maior correspondência das proteínas identificadas. A maior taxa de correspondência em números absolutos foi com as espécies *Capitella teleta* e *Helobdella robusta*. Ambas as espécies são utilizadas como organismo modelo em estudos de processos biológicos. *C. teleta* é uma espécie de poliqueta marinho que encontra-se em faixas costeiras, especificamente nas zonas de areia (BLAKE; GRASSLE; ECKELBARGER, 2009). *H. robusta* conhecida como sanguessuga tem um histórico em tratamentos na área médica e a descoberta a nível molecular de suas proteínas revelam um potencial biotecnológico de exploração (VEENSTRA, 2011). Em estudos de filogenia com base na superfamília de genes Homeobox, importantes na compreensão de fenômenos evolutivos, *C. teleta*, *H. robusta* e *Eisenia fetida* ocuparam o mesmo clado, ou seja, são filogeneticamente próximas e dessa forma os autores sugerem que elas compartilham um ancestral em comum (ZWARYCZ et al., 2016).

Considerando o exposto, pode-se inferir que existe uma proximidade entre estas duas espécies de minhocas, *A. gracilis* e *E. fetida*, pois no presente trabalho também foi observada uma grande correspondência das proteínas de *A. gracilis* com as espécies *C. teleta* e *H. robusta*. É importante ressaltar, ainda, que nossos resultados mostraram que *A. gracilis* e *E. fetida* compartilham 0,16% de similaridade de sequências em seus transcriptomas e este número seria certamente maior se mais dados genéticos de ambas as espécies estivessem disponíveis nos bancos de dados, a exemplo do que acontece com *C. teleta* e *H. robusta*.

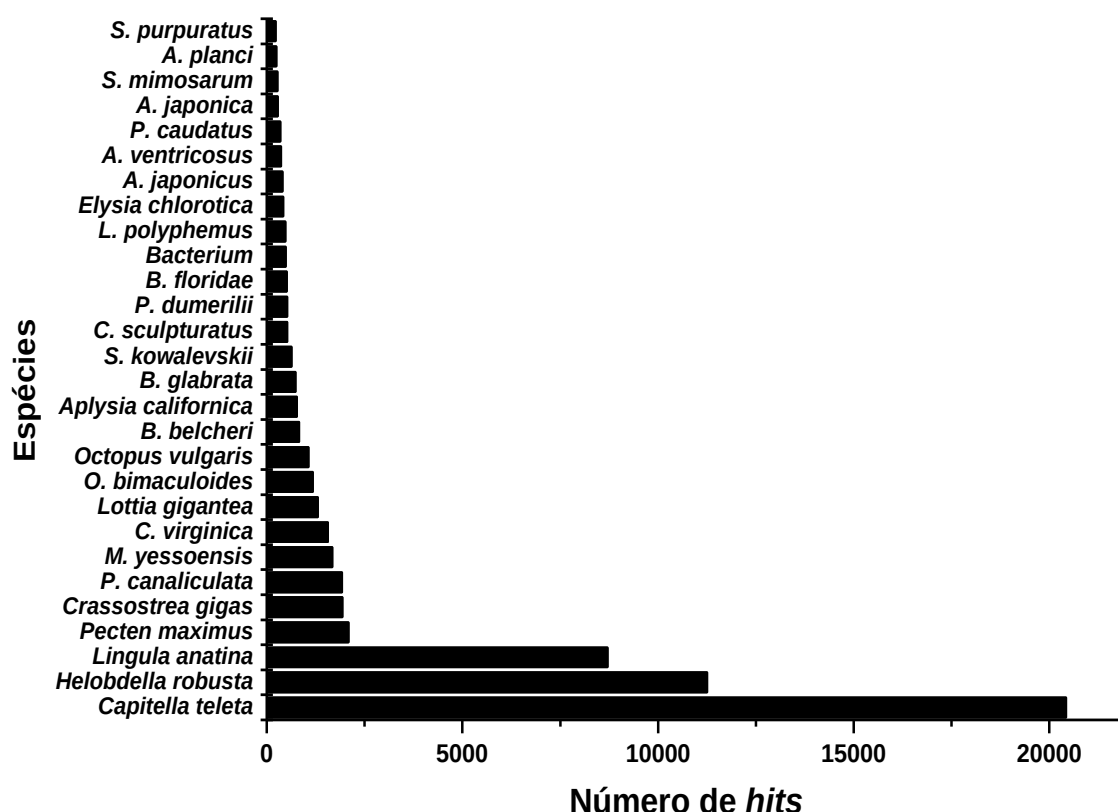


Figura 11: Representação gráfica da relação entre o número de *hits* dentro do transcriptoma de *A. gracilis* que tiveram correspondência com sequências depositadas no banco de dados NR do NCBI através da análise de BLAST. Sendo o eixo horizontal representado pelo número de *hits* e o eixo vertical pelas espécies que possuem sequências similares às sequências anotadas neste trabalho.

Fonte: Elaborado pela Autora (2022).

O Gene Ontology (GO) é uma ferramenta de anotação especializada em identificar produtos gênicos categorizando-os em diferentes classes de genes (ontologias) correlacionadas entre si: Processos biológicos (PB) corresponde ao genes relacionados às rotas metabólicas e processos onde a função molecular dos genes são utilizadas; Funções moleculares (FM) que compreende à atividade do produto gênico em nível molecular; Componentes celulares (CC) é a ontologia que corresponde às localidades onde os produtos gênicos estão em relação às estruturas biológicas. A base de dados contida no GO contém mais de 7 milhões de anotações de genes provenientes de mais de 3200 espécies, onde 10% deste banco tem o suporte de dados experimentais (THE GENE ONTOLOGY CONSORTIUM, 2019).

A **Figura 12** representa a anotação feita pelo GO, na qual estão representados o número de sequências encontradas para 20 classes de genes dentro das ontologias

“Processos Biológicos” (**Figura 12A**), “Funções Moleculares” (**Figura 12B**) e “Componentes Celulares” (**Figura 12C**). Dentro da ontologia PB temos o “processo do metabolismo celular” com maior número de sequências anotadas, seguido do “processo do metabolismo do nitrogênio” e “processos metabólicos envolvendo substâncias orgânicas”, que possui o mesmo número de genes anotados para a classe de “processo do metabolismo primário”. “Comunicação celular”, “processo de biossíntese”, “processo de regulação celular”, “estímulo à resposta celular”, “transdução de sinal” apresentam números de genes anotados muito similares, entre 9.000 e 10.000 sequências. O menor número de sequências anotadas dentro destas 20 primeiras categorias foi para “adesão celular”. Para a ontologia FM temos “ligação a íons”, “atividade transferase” e “atividade hidrolase” como as três principais, sendo que outras classes de enzima também aparecem nos dados, como oxidoredutases, ligases, isomerases. Já para CC, as três classes com maior número de anotação foram “estruturas de anatomia intracelular”, “organelas” e “citoplasma”, sendo seguidas das classes “membrana”, “sistema endomembrana”, “citossol”, “nucleoplasma” e “região extracelular”. A classe com menor número de anotação foi “estrutura de encapsulamento externo”.

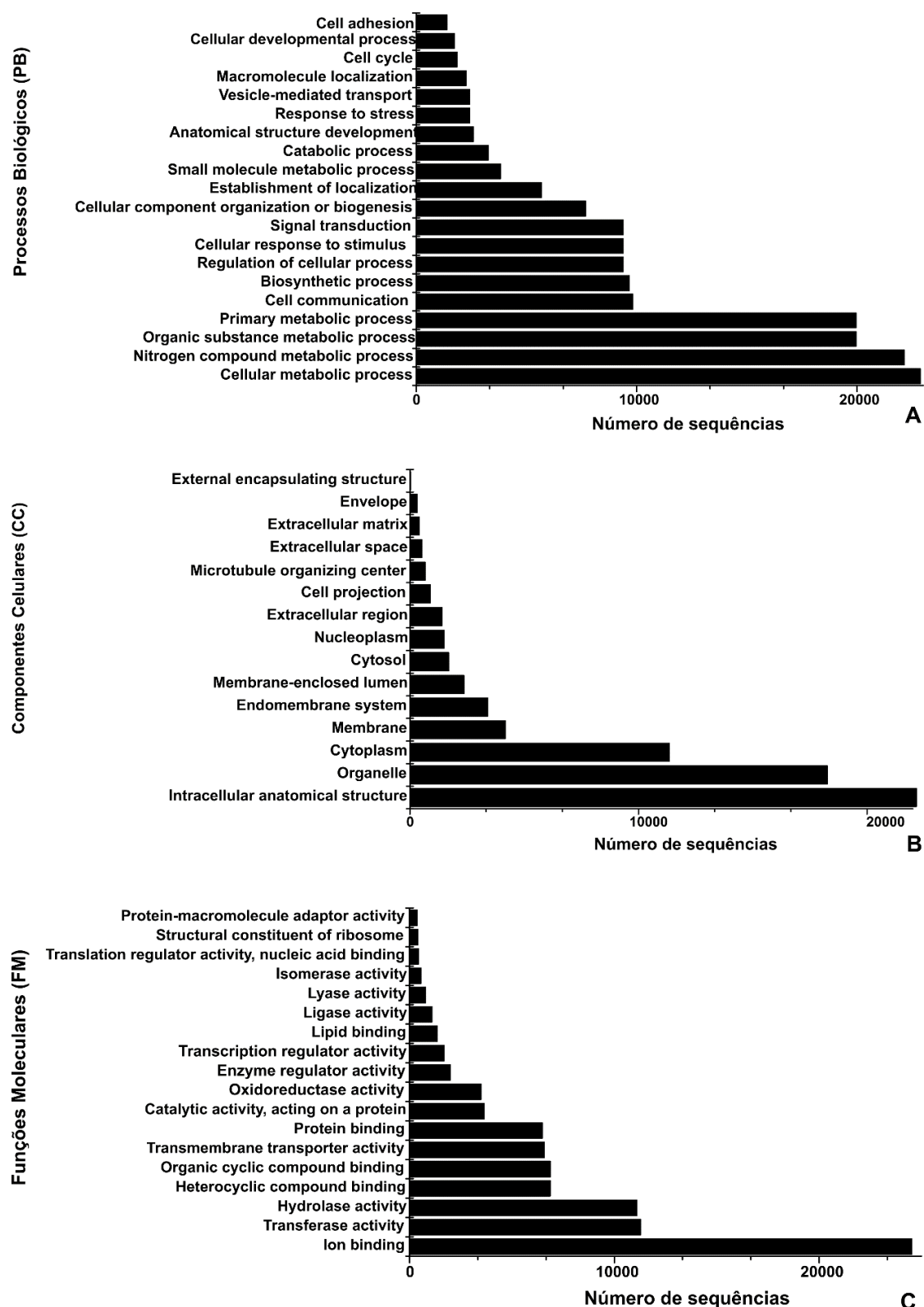


Figura 12: Anotação obtida pelo Gene Ontology (GO). O eixo X representa o número de seqüências identificadas e classificadas nas diferentes ontologias descritas no eixo Y. (A): Processos biológicos (PB); (B): Funções moleculares (FM); (C): Componentes celulares (CC).

Fonte: Elaborado pela Autora (2022).

Dados de transcriptoma da espécie de minhoca *E. fetida* revelaram diferentes classes de genes expressos. Os resultados obtidos a partir dos transcritos do cordão nervoso de *E. fetida* mostram que para a ontologia BP os maiores números de genes anotados foram nas classes “processos celulares”, “processos metabólicos” e “processo únicos do organismo”. Para a ontologia FM, o maior número de genes identificados foram dentro das classes “ligação” e “atividade catalítica”. E para ontologia CC foram encontrados grande número de genes para formação de células, organelas e complexo macromolecular (PAUL et al., 2018; PONESAKKI et al., 2017). Pode-se inferir que o enfoque dado ao cordão nervoso de *E. fetida* seja a principal razão da diferença encontrada em relação ao transcriptoma de *A. gracilis* aqui apresentado, uma vez que nossa análise fez uso de exemplares inteiros da espécie.

A partir dos dados analisados, foram listados os 20 genes mais expressos no transcriptoma de *A. gracilis*. Como pode ser visualizado na **Tabela 2**, foram identificadas dentro das sequências mais expressas genes de uma protease chamada Lumbrokinase (LK). As LK são enzimas encontradas somente em anelídeos, especificamente em espécies de minhocas terrestres e algumas espécies de poliquetas marinhos (HANKS; HUNTER, 1995).

O uso das LKs é bastante difundido na China, através dos preceitos da medicina chinesa. Essas enzimas possuem atividade trombolítica, ou seja, são capazes de desobstruir trombos sanguíneos causados por doenças de origem cardiovasculares (WANG et al., 2019). A indústria farmacêutica comercializa essas proteases em forma de pó ou encapsuladas e alguns estudos apontam a utilização da inserção de genes de LKs em plantas geneticamente modificadas (WANG et al., 2013).

O aumento da expressão dessas proteases em minhocas pode estar relacionado também com a exposição desses indivíduos a condições de estresse ambiental. Estudos onde se avaliou o efeito do fungicida benzovindiflupir (BZF) na espécie de *E. fetida* mostraram que as minhocas submetidas a presença de maiores doses de BZF tiveram uma maior expressão gênica de LK, o que indica o potencial dessas moléculas como um biomarcador para investigação de poluição ambiental (YAO et al., 2021).

Além das LKs, foram encontrados genes de ferretina, proteína importante no metabolismo do ferro além de alguns pigmentos respiratórios como hemeritrina e hemoglobinas extracelulares. A hemoglobina extracelular gigante HbAg é alvo principal desse projeto de doutorado e, já era previsto o seu aparecimento entre os genes mais expressos, por se tratar de uma proteína essencial para manutenção e sobrevivência da

espécie *A. gracilis*. Mas além da identificação das Hbs, destaca-se aqui a expressão da hemeritrina, ou seja, há a expressão concomitante de dois pigmentos respiratórios em *A. gracilis*. Diferentemente das Hbs, as hemeretrinas não possuem o grupo Heme, mas sim um sitio ativo de ferro capaz de se ligar a duas moléculas de oxigênio. Os dados encontrados neste trabalho corroboram com os dados encontrados por Costa-Paiva e Coates (2020), que sugerem que todas as espécies de anelídeos apresentam genes de hemeritrinas em seus genomas. A partir da análise de sequências depositadas em bancos de dados foram identificados genes de hemeritrinas em 44 espécies de anelídeos incluindo mais de 40 famílias diferentes (COSTA-PAIVA et al., 2017).

Tabela 2: Lista dos 20 genes mais expressos no transcriptoma de *A. gracilis*. Na tabela estão contidos os identificadores do gene e suas isoformas; extensão do gene em número de pares de bases (pb); os parâmetros gerados pelo programa TPM e FPKM, a descrição do *hit* no banco de dados do NR-NCBI que obteve correspondência com as proteínas preditas no transcriptoma de *A. gracilis*; o código identificador da sequência no NCBI; o parâmetro estatístico *e-value* e a porcentagem de similaridade.

Gene ID ¹	Transcrito ID ¹	Extensão (pb)	TPM ²	FPKM ³	Descrição <i>hit</i> NR ⁴	<i>Hit</i> ACC ⁵	E-value	Similaridade (%)
comp272973_c0	comp272973_c0_seq1	3667	30453,87	54379,3	Senescence-associated protein	OWF36834, OWF54769	7,66E-62	84
comp64724_c0_seq1	comp64724_c0_seq1	3608	11771,45	21019,44	Hypothetical protein EgrG_000335800 [<i>Echinococcus granulosus</i>]	CDS20307.1	5,00E-64	56,02
comp256634_c0	comp256634_c0_seq1 comp256634_c0_seq2 comp256634_c0_seq3 comp256634_c0_seq4 comp256634_c0_seq5 comp256634_c0_seq6 comp256634_c0_seq7 comp256634_c0_seq8	471	8848,16	15799,52	Ubiquitin-60S ribosomal protein L40 [<i>Equus asinus</i>]	XP_014687043.1	4,00E-13	67,92
comp258731_c3	comp258731_c3	1826	8521,72	15216,62	Lumbrokinase-4 precursor [<i>Eisenia fetida</i>]	AAR13225.1	4,00E-76	60,83
comp218539_c0	comp218539_c0_seq1 comp218539_c0_seq2	2682,43	4846,09	8653,32	Hypothetical protein RclHR1_00930026 [<i>Rhizophagus clarus</i>]	GBC10048.1	3,00E-51	74,16
comp244196_c0	comp244196_c0_seq1	243	3830,69	6840,18	Unnamed protein product [<i>Rotaria</i> sp. Silwood2]	CAF3983879.1	4,00E-50	100
comp258756_c0	comp258756_c0_seq1 comp258756_c0_seq2 comp258756_c0_seq3	1025,57	3232,28	577,65	Ferritin [<i>Eisenia andrei</i>]	ACL14179.2	3,00E-95	80
comp272976_c1	comp272976_c1_seq1	3080	2499,43	4463,06	Cytochrome oxidase subunit 1	BBF98036	3,03E-47	91

comp181532_c0	comp181532_c0_seq1	1567	2437,79	4352,99	Cytochrome c oxidase subunit III [<i>Amyntas gracilis</i>]	YP_009144484.1	5,00E-14	89,96
comp261583_c0	comp261583_c0_seq1 comp261583_c0_seq2	1613,99	1954,88	3490,69	Actin	AOR07100, AOR07104	0	99
comp263270_c0	comp263270_c0_seq1 comp263270_c0_seq10 comp263270_c0_seq2 comp263270_c0_seq3 comp263270_c0_seq4 comp263270_c0_seq5 comp263270_c0_seq6 comp263270_c0_seq7 comp263270_c0_seq8 comp263270_c0_seq9	1728,48	1774,91	1774,91	Actin	P92182, CAA65361, CAA65363, CAA65364	0	99
comp253679_c0	comp253679_c0_seq1 comp253679_c0_seq2	1410,77	1750,23	3125,26	Extracellular globin [<i>Amyntas</i> sp. FB-2019]	QAT81040.1	1,00E-60	68,79
comp226471_c0	comp226471_c0_seq1	946	1225,09	2187,55	Hypothetical protein HELRODRAFT_174071 [<i>Helobdella robusta</i>]	XP_009018864.1	6,00E-09	29,24
comp186986_c0	comp186986_c0_seq1 comp186986_c0_seq2	852,5	1182,81	2112,06	Proprotein convertase subtilisin/kexin type 5-like	XP_009017608, ESO04339	2,12E-34	63
comp253920_c0	comp253920_c0_seq1 comp253920_c0_seq2 comp253920_c0_seq3 comp253920_c0_seq4 comp253920_c0_seq5 comp253920_c0_seq6 comp253920_c0_seq7 comp253920_c0_seq8	600,38	1103,93	1971,21	PREDICTED: lipoxxygenase homology domain-containing protein 1-like [<i>Hydra vulgaris</i>]	XP_012565699.1	4,00E-27	33,76
comp264175_c2	comp264175_c2_seq1 comp264175_c2_seq2 comp264175_c2_seq3 comp264175_c2_seq4	1064,34	1052,24	1878,91	Lumbrokinase-P2(2) [<i>Lumbricus</i> <i>rubellus</i>]	AAQ15052.1	5,00E-87	60,98

	comp201897_c0_seq4							
	comp201897_c0_seq5							
	comp201897_c0_seq6							
comp241194_c0	comp241194_c0_seq1	1329	987,6	1763,49	SCBP3 protein [<i>Lumbricus terrestris</i>]	CAD29317.1	6,00E-75	69,77
comp255010_c0	comp255010_c0_seq1	1202,23	947,87	1692,54	Hemerythrin [<i>Aphrodita japonica</i>]	AQV13580.1	2,00E-59	78,99
	comp255010_c0_seq2							
	comp255010_c0_seq3							
	comp255010_c0_seq4							
	comp255010_c0_seq5							
	comp255010_c0_seq6							
comp265083_c0	comp265083_c0_seq1	4890	945,43	1688,18	Hypothetical protein CAPTEDRAFT_151312 [<i>Capitella teleta</i>]	ELU15842.1	0	74,88

¹ID: identificador

²TPM: transcritos por milhão de kilobase

³FPKM: fragmentos por milhão de kilobase

⁴NR: banco de dados não redundante (*non-redundant*)

⁵Hit ACC: identificador da sequência no NCBI

Fonte: Elaborado pela Autora (2022).

Para a identificação de famílias e domínios proteicos, o OmicsBox faz uma varredura das sequências através do software InterproScan que possui diferentes bancos de dados: PFAM; SMART e SUPERFAMILY. Pode-se observar que para os três bancos de dados obtivemos um número grande de proteínas identificadas com o domínio quinase (**Figura 13**). Este domínio é considerado um dos domínios proteicos com maior índice de registros de genes nos organismos eucariotos. Existem diferentes tipos de quinases, sendo caracterizadas como uma classe de enzimas que usa o fosfato de ATP ou GTP para gerar monoésteres de fosfato, usando grupo OH^- dos aminoácidos serina, treonina ou tirosina como aceptores fosfato (HANKS; HUNTER, 1995). Dentro deste domínio encontramos as proteínas PTK (*protein tyrosine kinase*) que exercem papel importante no mecanismo celular junto com outra classe de proteínas, a PTP (*protein tyrosine phosphatase*). Em conjunto elas trabalham em reações de fosforilação e desfosforilação dentro dos organismos nos quais se encontram presentes (YU; ZHANG, 2018; ZHANG, 2001).

Também foi identificado o domínio *zinc-finger*. As proteínas *zinc-fingers* possuem íons de zinco coordenados com cisteínas e histidinas, dando assim a conformação estrutural da molécula. A estrutura dessas proteínas permite interação das mesmas com moléculas de DNA, RNA ou outras proteínas e, elas podem agir como moduladores de doenças neurológicas como Alzheimer e Mal de Parkinson. A troca dos íons Zn^{2+} por outros metais pode levar a inibição dos domínios *zinc-finger* e, dessa forma, essa propriedade tem sido estudada para o desenvolvimento de fármacos (ABBEHAUSEN, 2019).

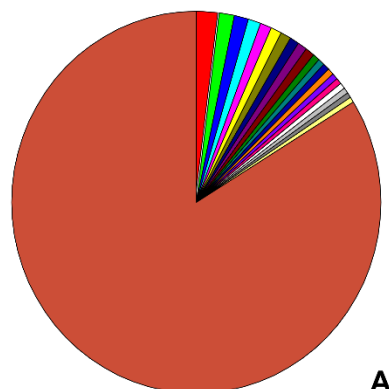
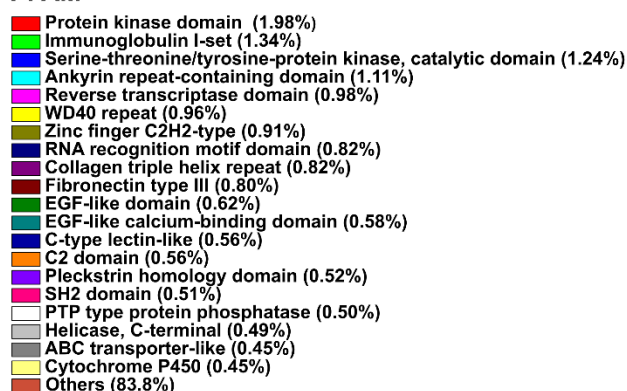
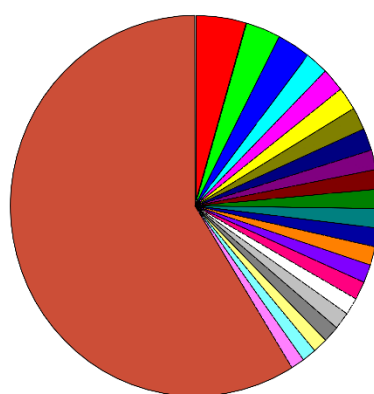
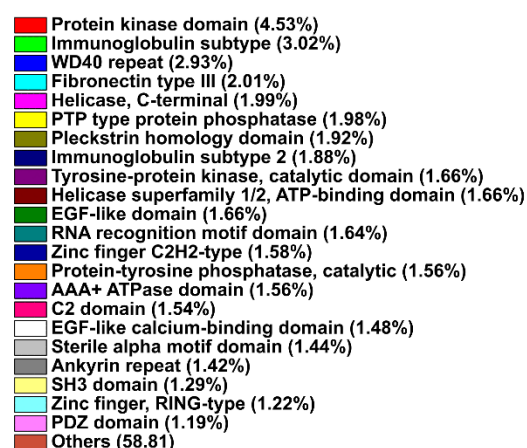
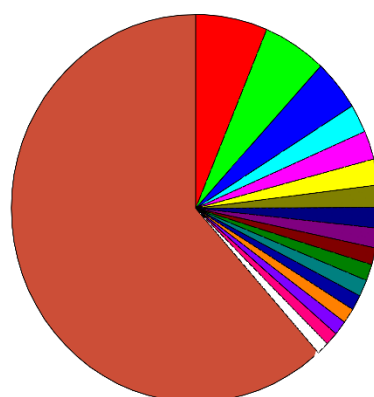
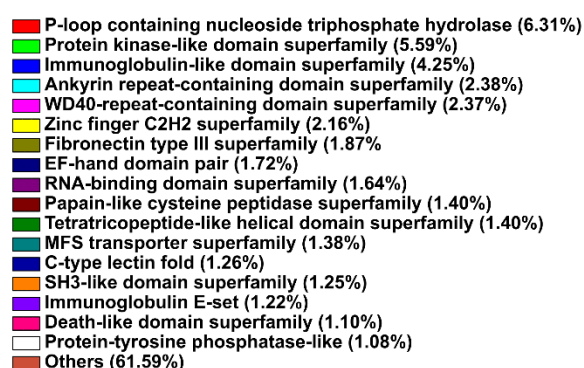
PFAM**A****SMART****B****SUPERFAMILY****C**

Figura 13: Distribuição das principais proteínas anotadas através da ferramenta InterproScan. Proteínas anotadas contra as sequências depositadas nos bancos de dados PFAM (A) SMART (B) e SUPERFAMILY (C).

Fonte: Elaborado pela Autora (2022).

Por fim, a análise de anotação funcional do transcriptoma de *A. gracilis* permitiu a identificação de diferentes classes de enzimas, destacando-se um maior número de sequências anotadas para as classes de enzimas hidrolases, transferases e oxidoredutases respectivamente (**Figura 14**).

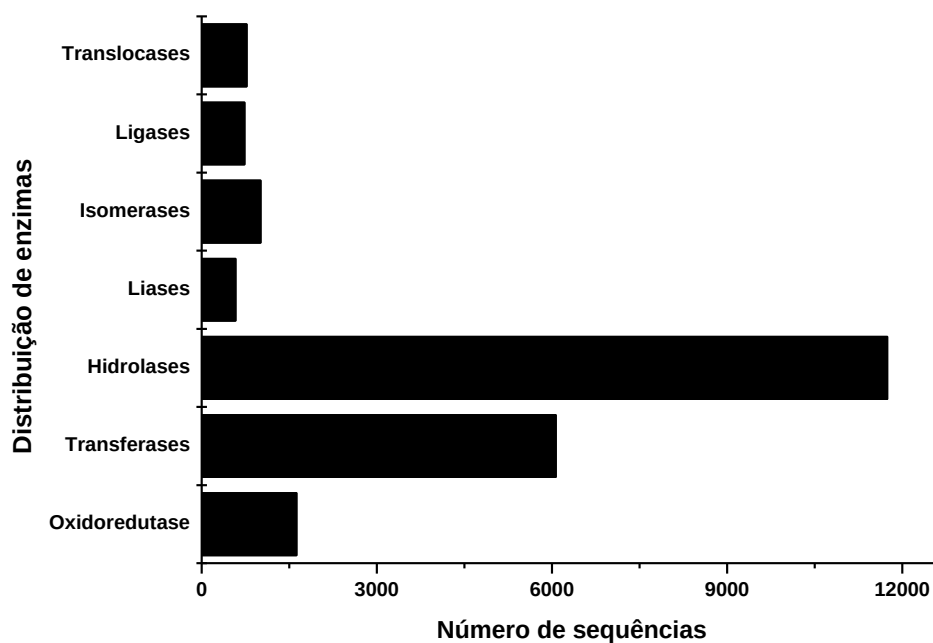


Figura 14: Distribuição das enzimas encontrada no transcriptoma de *A. gracilis*. O eixo X representa o número de sequências anotadas e o eixo Y representa as principais classes de enzimas identificadas.

Fonte: Elaborado pela Autora (2022).

Para a espécie de minhoca *E. fetida*, a distribuição de enzimas encontradas foi similar ao apresentado nestes dados, sendo que cerca de 47,96% das sequências de proteínas anotadas tiveram relação com proteínas com atividade hidrolases, 23,84% foram para transferases e 19,44% foram identificadas como proteínas oxidoredutases (PONESAKKI et al., 2017).

O maior número de sequências anotadas foi na classe das hidrolases, enzimas responsáveis por exercer funções importantes no grupo das minhocas. As minhocas junto com microorganismos do solo promovem o processo de vermicompostagem e as hidrolases presentes no conteúdo intestinal das minhocas estão envolvidas neste processo (ORDOÑEZ-ARÉVALO et al., 2018). A formação dessa compostagem se divide em duas etapas: uma etapa conhecida como etapa ativa e outra etapa de maturação. A fase ativa é caracterizada quando as minhocas se alimentam dos resíduos do solo e no seu processo

digestivo depositam enzimas e microorganismos àquele material. A fase seguinte de maturação ocorre quando os demais microorganismos presentes no solo interagem com o produto da digestão das minhocas e o convertem em um vermicomposto (ORDOÑEZ-ARÉVALO et al., 2018; SHARMA; GARG, 2019).

4 CONCLUSÕES

Os dados brutos obtidos pelo sequenciamento estão de acordo com os parâmetros de qualidade estabelecidos pelo sistema Illumina, onde obtivemos cerca de 68% das sequências com um *phred* $Q \geq 30$. A técnica *de novo* para a montagem do transcriptoma gerou um total de 477.632 transcritos, onde 222.293 são considerados genes. Este número é divergente pois dentro dos transcritos obtidos encontram-se as diferentes isoformas para cada gene identificado. Através da análise de identificação de ORF pelo Transdecoder foram obtidas 260.016 ORFs, das quais 187.866 são sequências não redundantes. Foram anotadas pelo menos 92.939 sequências, o que representa 49% do transcriptoma obtido. A maior similaridade entre os *hits* de *A. gracilis* contra as sequências depositadas no banco de dados do NR do NCBI foi com as espécies *C. teleta* e *H. robusta*, as quais possuem um número expressivo de dados disponibilizados nos bancos de dados. Pela classificação do GO o maior número de sequências preditas para a ontologia PB foi de proteínas envolvidas em processos metabólicos e metabolismo do nitrogênio. Para a ontologia CC foi observado um expressivo número de sequências de estruturas intercelulares e organelas. E para a ontologia FM os dados se concentram em sequências de proteínas que realizam ligação com íons e enzimas com diferentes atividades. Algumas proteases foram anotadas no transcriptoma de *A. gracilis* com destaque para as proteases LK, que apresentam potencial de uso biotecnológico. Para os genes mais expressos foi observada a expressão concomitante de dois pigmentos respiratórios: a hemeritrina e Hb gigante. Esses dados corroboram dados descritos na literatura sobre a hipótese de que todas as espécies de anelídeos expressarem genes de hemeritrina. Também foram preditas algumas proteínas estruturais como as de domínio *zinc-fingers* e uma série de enzimas de diferentes categorias. Os dados reportados aqui podem auxiliar em pesquisas com diferentes objetivos em *A. gracilis*, pois trazem informações básicas sobre algumas de suas características genéticas e podem, também, auxiliar em estudos de outras espécies de minhocas ou anelídeos relacionados.

CAPITULO 3 – IDENTIFICAÇÃO DAS SEQUÊNCIAS DE HbAg E CONSTRUÇÃO DO SEU MODELO MOLECULAR

RESUMO

HbAg é uma hemoglobina gigante extraída da espécie de minhoca puladeira *A. gracilis*. Com propriedades que remetem ao seu potencial uso biotecnológico, HbAg tem sido alvo de estudos pelo nosso grupo de pesquisa visando sua caracterização e exploração através do desenvolvimento de um protótipo de biossensor de contaminação ambiental. Obter informação a respeito da estrutura destas proteínas contribui para trabalhos desenvolvidos pelo grupo e pela comunidade Científica. Dessa forma, o objetivo deste capítulo foi identificar as sequências de todas as subunidades que compõe HbAg assim como construir um modelo teórico da estrutura de seu complexo proteico. Para tanto, a partir de sequências geradas do transcriptoma de *A. gracilis*, foram identificadas todas as cadeias de HbAg através de análise de BLAST contra o bando de dados UniprotKB. O modelo comparativo de HbAg foi construído utilizando o Swiss model. Todas as cadeias foram identificadas com alto grau de conservação entre as proteínas homólogas HbLt e HbGp. Os domínios estruturais presentes em HbGp também se mantiveram conservados, assim como os sítios de ligação aos íons Ca e Zn, que são regiões importantes para a estabilidade da molécula. Aqui apresentamos informações básicas que sustentam dados encontrados anteriormente pelo grupo e também abrem novos caminhos de estudos para trabalhos futuros.

1 INTRODUÇÃO

A descrição estrutural do complexo proteico HbAg tem como ponto de partida o conhecimento das estruturas de proteínas homólogas como a HbLt, HbGp e HbAm. Essas Hbs apresentam a seguinte conformação: são compostas por uma dupla camada hexagonal de protômeros constituídos por dodecâmeros onde cada um destes dodecâmeros são formados por um trímero de tetrâmeros contendo 4 tipos de globinas (*a*, *b*, *c* e *d* - formação do cap) e, outras cadeias chamadas de *linkers* (L1, L2 e L3), com a estequiometria $([abc]_3[d]_3)_3(L1L2L3)$ (BACHEGA et al., 2015; ROYER et al., 2006; ZAL; ROUSSELOT, 2014). No total são 200 cadeias polipeptídicas, sendo 144 globinas, ou seja, 144 subunidades com o grupo prostético Heme capaz de se ligarem às moléculas de O₂. Sendo essa uma das características principais que fazem com que essas Hbs sejam matrizes para o desenvolvimento de moléculas carreadores de O₂ (ROUSSELOT et al., 2006).

As Hbs gigantes também conhecidas como Hbs extracelulares, diferentemente da Hb humana, encontram-se dispersas no sangue dos anelídeos e não abrigadas em células sanguíneas. As características de carreamento de um número elevado de moléculas de O₂, alta estabilidade térmica e resistência ao estresse oxidativo fazem com que essas Hbs sejam alvo de estudos de grupos de pesquisa globalmente, com diferentes objetivos e ênfase no desenvolvimento de produtos terapêuticos (ZAL; ROUSSELOT, 2014). Atualmente, alguns produtos produzidos a partir da HbAm estão disponíveis no mercado e eles possuem diferentes funções com destaque para as seguintes: carreamento de O₂, melhoria na cicatrização de feridas, solução para armazenamento de órgãos pré-transplante, tratamento dentário, melhoria de condições de crescimento de microrganismos (LE MEUR et al., 2020; VARNEY et al., 2021; ZAL; ROUSSELOT, 2014).

Porém, outras áreas de interesse biotecnológico podem ser exploradas, a exemplo da criação de um protótipo de biossensor de contaminação ambiental, um dos objetivos visados pelo nosso grupo de pesquisa. Os estudos com HbAg se iniciaram em 2015 e vem sendo desenvolvidos em duas frentes: uma delas voltada para caracterização da biomolécula a partir de diferentes técnicas biofísicas e outra voltada pelo o desenvolvimento do biossensor. Dessa forma, experimentos visando a detecção dos metais Cu, Cd e Zn para elaboração de sensores com base eletroquímica e óptica têm sido desenvolvidos. A HbAg mostrou-se ser sensível quantitativamente à detecção desses metais, classificando-a como uma molécula promissora nesta área (SOUZA, 2018).

Sendo assim, o objetivo deste capítulo foi identificar as sequências codificadoras das globinas e *linkers* que compõe HbAg, assim como construir um modelo da sua estrutura através de modelagem comparativa. Essas informações básicas podem trazer suporte para os estudos desenvolvidos pelo grupo, além de disponibilizar esse conhecimento para outros pesquisadores que trabalham com moléculas semelhantes.

2 MATERIAIS E MÉTODOS

2.1 Identificação das sequências de HbAg

As etapas iniciais para obtenção das sequências de HbAg foram descritas no capítulo 2. Assim, a metodologia descrita desde a extração do RNA de *A. gracilis* até montagem do transcriptoma já foi reportada neste documento. Para identificação das cadeias de HbAg, as sequências dos transcritos foram alinhadas contra as sequências das cadeias de hemoglobinas gigantes de outras espécies de anelídeos depositadas no banco de dados UniprotKB através do programa BLASTX (CAMACHO et al., 2009), utilizando como fator de seleção um *e-value* de 1.0E-20. Posteriormente foram selecionados os melhores *hits* com um *script perl*. Parâmetros físico-químicos da proteína foram estimados através da ferramenta ProtParam disponível no portal ExPASy (GASTEIGER et al., 2005). A região do peptídeo sinal presente nas globinas foi predita através do servidor signalP 4.1 (PETERSEN et al., 2011). Os alinhamentos entre as sequências foram realizados utilizando o programa Jalview versão 2.11.1.7 a partir do algoritmo CLUSTAL e os parâmetros da análise seguiram os estabelecidos pelo *default* do algoritmo.

Em virtude de um interesse específico do grupo de pesquisa no potencial biotecnológico do monômero-*d* de HbAg, realizou-se uma análise bayesiana para detectar em qual família de genes de globina encontra-se a sequência codificadora deste monômero. Para tanto, as sequências de isoformas do monômero-*d* foram incorporadas a um banco de dados contendo 65 sequências de globinas extracelulares de metazoários (BELATO et al., 2019). O novo dataset contendo as duas sequências para o monômero-*d* foram alinhadas com o software MAFFT, trimadas usando o software TrimAL e o melhor modelo de evolução de proteínas foi estimado pelo software ProTest 3.4 (DARRIBA et al., 2011). Posteriormente foi feita a análise de inferência bayesiana no software MrBayes 3.2.1 (RONQUIST; HUELSENBECK, 2003).

2.2 Construção do complexo proteico HbAg

O modelo do complexo proteico HbAg foi construído utilizando o servidor SWISS-MODEL (WATERHOUSE et al., 2018), tendo como molde a porção assimétrica do trímero de protômeros da estrutura cristalográfica de HbGp (código PDB: 4u8u). A estrutura inteira de HbAg foi gerada no PyMOL através da sobreposição de 4 cópias dos trímeros de protômeros de molde de HbGP com outros pares simétricos da estrutura cristalográfica de HbGp. Todos os esquemas e figuras da estrutura oligomérica foram desenhados utilizando o PyMOL versão 2.5.

3 RESULTADOS E DISCUSSÃO

Nos dados das **Tabela 3** salientamos os parâmetros da análise de BLAST contra as sequências de outras Hbs gigantes, especificamente HbLt e HbGp que são ortólogas a HbAg. Encontram-se também nos bancos de dados algumas sequências de Hbs do gênero *Amyntas*. Assim, pode-se observar que para a globina *a* de HbAg a porcentagem de identidade da sua sequência com sequências das Hbs das espécies *G. paulistus* e *L. terrestris* foi de 42-44%, com uma cobertura em torno de 80%. Já para a globina *b*, esses índices aumentaram para 66-72% de identidade para HbLt e HbGp, com uma cobertura de mais de 80%; contra as sequências do mesmo gênero esses valores de identidade são de 76-89% e cobertura de 100%. Para a globina *c* foi encontrada uma cobertura de 100% com identidade de 75-88% para sequências de globinas do gênero *Amyntas* e, uma cobertura de 87% com identidade de 60-62% para HbGp e HbLt. Para globina *d*, quando comparada com as sequências do mesmo gênero foi encontrada uma taxa de cobertura de 91% e identidade de 48-53%, enquanto que para HbLt os valores percentuais de identidade foram de 49-50% e cobertura de 96% e, para HbGp, com 88% de cobertura, a taxa de identidade foi de 45%. Para as sequências dos *linkers* os valores de identidade variaram de 41 a 61% com cobertura de 84 a 99%.

A partir da análise das sequências obtidas foi possível estimar o ponto isoelétrico (pI) e a massa molecular das subunidades. Para as globinas as estimativas do pI ficaram entre 6,12 e 7,0 e, as massas moleculares foram em torno de 18/19 kDa. Já para os *linkers* o pI estimado ficou entre 4,86 e 6,24 e, as massas moleculares variam de 28 a 30 kDa. A partir de dados de potencial Zeta, o pI encontrado para proteína HbAg nativa é de 6 e para

a subunidade d o pI foi estimado em 7,1 a partir da análise de gel bidimensional (OLIVEIRA et al., 2020). Já para as massas moleculares, os índices obtidos por MALDI-TOF revelaram massas em torno de 16 e 17 kDa para as globinas de HbAg e 26 kDa para os *linkers* (OLIVEIRA et al., 2020). Dessa forma os dados aqui apresentados corroboram com os dados experimentais obtidos para HbAg. Para as proteínas homologas à HbAg, HbGp e HbLt as massas moleculares das globinas variam de 15 a 17 kDa e são encontrados valores de 25 a 32 kDa para os *linkers* (BACHEGA et al., 2015; CARVALHO et al., 2011a; ROYER et al., 2006).

Tabela 3: Tabela descritiva correlacionando as sequências identificadas para as subunidades de HbAg com os parâmetros obtidos através de análise de BLAST. ID Blast: proteína correspondente no banco de dados NR-NCBI; *E-value*: parâmetro estatístico obtido a partir do alinhamento entre as sequências *query* e *subject*, sendo *query* a sequência alvo do alinhamento e *subject* a sequência depositada nos bancos de dados; Cobertura: porcentagem de cobertura do alinhamento; Identidade: porcentagem de identidade entre as sequências *query* e *subject*; Acesso: número de identificação da sequência *subject* no banco de dados. Também estão descritos na tabela os parâmetros físico-químicos teóricos pI e MM (kDa), que foram obtidos através da ferramenta online ProtParam que fica sediada no servidor da ExPASy. O tamanho em número de resíduos de aminoácidos corresponde ao tamanho das sequências identificadas das subunidades de HbAg.

Subunidades HbAg	ID Blast	E-value	Cobertura	Identidade	Acesso	pI teórico	MM	Tamanho
Parâmetros Blast						Parâmetros físico-químicos		
HbAg a	Globina a HbGp	2e-40	80	44,76	4U8U_A	7,01	19,98	178
	Globina a HbLt	9e-40	84	42,00	1X9F_A			
HbAg b	Globina extracelular A. <i>gracilis</i>	3e-87	100	76,36	QAT81045.1	6,44	18,71	166
	Globina b HbLt	1e-62	86	66,20	1X9F_B			
	Globina b HbGp	8e-71	85	72,34	4U8U_B			
HbAg c	Globina extracelular A. <i>gracilis</i>	2e-106	100	88,95	QAT81049.1	6,12	18,84	172
	Globina extracelular A. <i>gracilis</i>	9e-89	100	75,44	QAT81043.1			
	Globina c HbGp	1e-61	87	62,67	4U8U_C			
	Globina c HbLt	1e-56	87	60,00	1X9F_C			

HbAg d	Globina extracelular <i>A. gracilis</i>	8e-52	91	53,10	QAT81040.1	6,49	18,52	159
	Globina extracelular <i>A. gracilis</i>	8e-44	91	48,97	QAT81041.1			
	Globina d HbLt	3e-45	96	49,34	4V93_B0			
	Globina d HbGp	9e-38	88	45,71	4WCH_D			
Linker L1	Linker chain <i>A. gracilis</i>	5e-97	99	64,26	QAT80614.1	5,03	28,34	251
	Linker L1 HbGp	3e-93	88	59,64	4U8U_M			
	Linker L1 HbLt	1e-85	97	52,63	4V93_C0			
Linker L2	Linker L2 HbGp	3e-95	87	59,11	4U8U_N	6,24	28,66	257
	Linker L2 HbLt	2e-94	85	61,82	5M3L_N			
	Linker L2 HbLt	3e-94	85	61,82	2GTL_N			
	Linker chain <i>A. gracilis</i>	8e-50	84	41,67	QAT80711.1			
	Linker chain <i>A. gracilis</i>	3e-94	90	58,78	QAT80613.1			
Linker L3	Linker chain <i>A. gracilis</i>	3e-90	83	59,73	QAT80712.1	5,29	27,05	270
	Linker L3 HbLt	1e-77	84	55,07	ABB71123.1			
	Linker L3 HbGp	1e-76	84	54,63	4U8U_O			

Fonte: Elaborado pela Autora (2022).

No alinhamento entre as sequências das globinas de HbAg e as sequências de globinas de Hbs homólogas pode se observar o nível de conservação entre os aminoácidos (**Figura 15**), os quais estão destacados a partir das cores correspondentes ao sistema de cores do alinhador Clustal. Nas globinas existe uma assinatura de resíduos de aminoácidos que costumam se manter conservados, são 12 resíduos: Cys-19, Trp-31, Phe-63, Val-66, Phe-76, His-79, Arg-82, His-106, Gln-110, His-111, Trp-144, Cys-147 (BELATO et al., 2019; CHABASSE et al., 2006a; NEGRISOLO et al., 2001; OWNBY et al., 1993). A posição desses resíduos pode variar de acordo com o tamanho da globina, assim podem ser notadas algumas diferenças para a sequência assinatura entre as globinas *a*, *b*, *c* e *d*. Esses aminoácidos conservados encontram-se marcados em caixa na cor vermelha (**Figura 15**) com exceção de uma fenilalanina na posição 66 do alinhamento entre as globinas *a*, pois esta não aparece nas sequências de HbAg e HbGp, estando somente presente em HbLt.

Os resíduos de cisteína presentes nas globinas tem papel fundamental na formação estrutural dessas moléculas pois a formação dos dímeros de monômeros ocorre a partir de ligações dissulfetos realizadas entre as cisteínas (CHABASSE et al., 2006a; NEGRISOLO et al., 2001). Nota-se a presença de cisteínas nas globinas *a*, *b* e *c*, as quais são capazes de realizar ligações intermoleculares através de ligações dissulfeto, enquanto no monômero-*d* não há a presença dessas cisteínas, pois este se mantém unido com as demais cadeias a partir de interações eletrostáticas (**Figura 15**) (BACHEGA et al., 2015; CARVALHO et al., 2011a; ROYER et al., 2006).

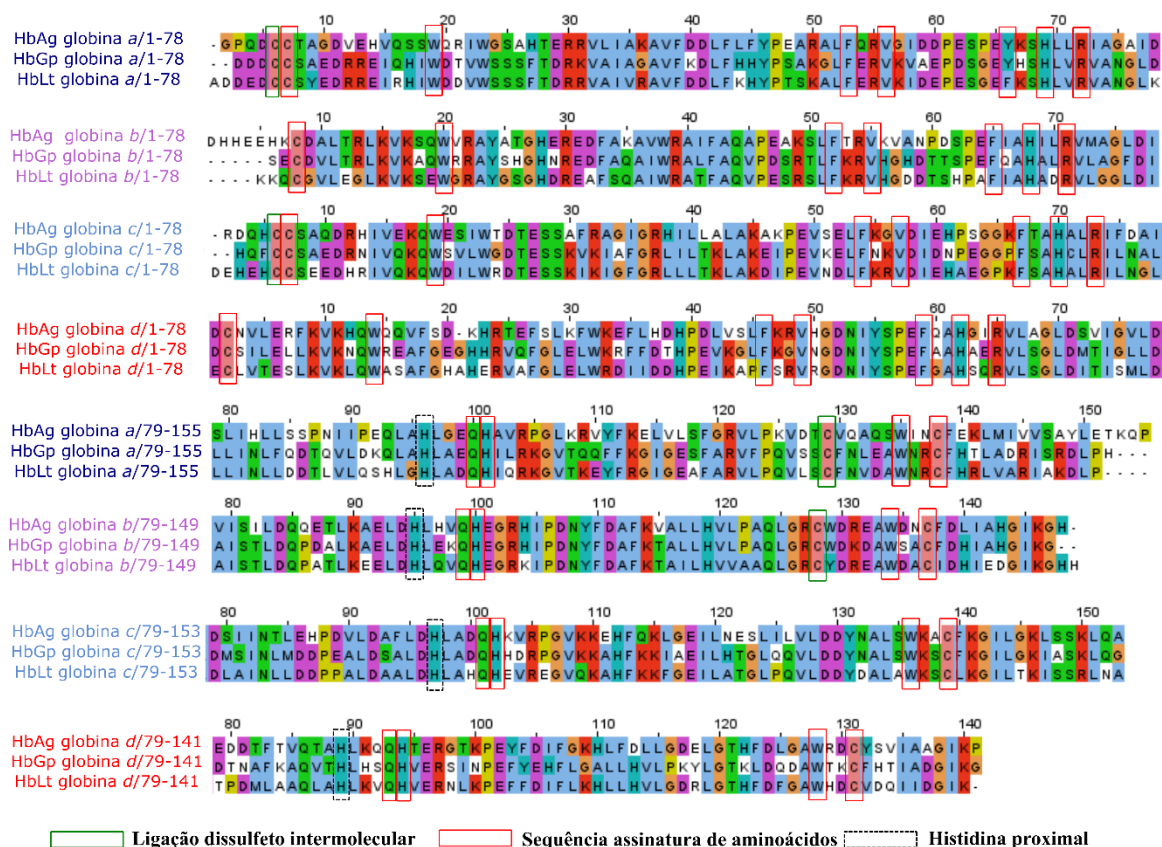


Figura 15: Alinhamento das sequências das globinas *a*, *b*, *c* e *d* de HbAg contra as Hbs homólogas HbGp e HbLt. Os resíduos de aminoácidos coloridos apresentam-se conservados entre as sequências e o sistema de cores do Clustal segue o seguinte padrão: azul – resíduos hidrofóbicos; vermelho – resíduos carregados positivamente; magenta - resíduos carregados negativamente; verde – resíduos polares; rosa – cisteínas; laranja – glicinas; amarelo – prolínas; azul ciano – aromáticos; branco – resíduos não conservados ou *gaps*. Os aminoácidos destacados em caixa vermelha se referem a sequência assinatura para globinas de proteínas HBL-Hb. Os resíduos em caixa verde são as cisteínas que fazem ligações intermoleculares e, a histidina proximal encontra-se em caixas com borda preta tracejada.

Fonte: Elaborado pela Autora (2022).

Foram descritos dois tipos de classificação para as globinas, existindo a classificação adotada neste trabalho onde as globinas foram divididas em 4 quatro subtipos (globinas *a*, *b*, *c* e *d*) e, também, a classificação das globinas em duas grandes famílias de genes, as globinas *a* e as globinas *b*. Também foi sugerida uma subdivisão dentro dessas 2 famílias de genes e dessa forma as subunidades *c* e *d* estariam dentro de uma subdivisão das famílias *a* e *b*, sendo a subunidade *c* codificada por um gene *b1* e a subunidade *d* por um gene *a2* (BAILLY et al., 2002). Porém, essa subdivisão não foi garantida por trabalhos

recentes envolvendo análises filogenéticas, onde se manteve a proposta da divisão das globinas em dois grandes grupos, globinas *a* e globinas *b* (BELATO et al., 2019).

Devido ao enfoque deste trabalho ser o estudo mais aprofundado do monômero-*d*, isoformas desta subunidade foram submetidas às análises que identificam em qual família de globinas essas isoformas se encontram. De acordo com essas análises, as isoformas do monômero-*d* de HbAg estão dentro da família de genes *a*. A

Figura 16 relaciona algumas globinas onde as sequências agrupadas no quadrante verde estão dentro da família de genes *b* e aquelas agrupadas no quadrante lilás pertencem à família de genes *a*. As sequências em vermelho são referentes aos transcritos que codificam o monômero-*d* de HbAg, resultado este corroborado pelo que foi descrito acima, presente em publicações diferentes (BAILLY et al., 2002; BELATO et al., 2019; CHABASSE et al., 2006a; GOTOH; SUZUKI; TAKAGI, 1991).

Para as sequências dos *linkers* existe a presença de um domínio rico em resíduos de cisteína: Cys-X₍₅₋₇₎-Cys-X₍₅₋₆₎-Cys-X₍₆₎-Cys-Asp-X₍₃₎-Asp-Cys-X₍₄₎-Asp-Glu-X₍₂₋₄₎-Cys (CHABASSE et al., 2006b; NEGRISOLO et al., 2001) que é homóloga à região LDL-A (*low density lipoprotein receptor class A*), que está presente em vários metazoários como humanos e anfíbios (MEHTA et al., 1991; SÜDHOF et al., 1985). Na **Figura 17**, a região LDL-A está destacada em um quadro na cor azul e as cisteínas estão marcadas com um asterisco. Dados de cristalografia identificaram a presença de sítios de ligação a íons de Ca²⁺ dentro do domínio LDL-A dos *linkers* de HbGp (BACHEGA et al., 2015). A partir do alinhamento entre essas sequências pode-se observar que os *linkers* de HbAg também possuem esses sítios de ligação (**Figura 17**).

Evolutivamente, pesquisas apontam que o surgimento das sequências de *linkers* podem ter vindo de um processo de duplicação dos genes de globina seguido da perda do primeiro éxon, presente nessas globinas, mais a inserção da região LDL-A (SUZUKI; RIGGS, 1993). A classificação dos *linkers* em L1, L2 e L3 em anelídeos foi suportada por dados filogenéticos contendo sequências provenientes de transcriptomas de 319 metazoários incluindo 16 filos diferentes (BELATO et al., 2019). Dados experimentais utilizando uma técnica de espectrometria de massas MALDI TOF-MS detectou a presença de três cadeias de *linkers* L1, L2 e L3 com MM de 25,8, 26 e 26,8 respectivamente (OLIVEIRA et al., 2020), que corroboram com os valores de MM estimados a partir das sequências dos *linkers* aqui identificadas. Assim, os dados reportados neste trabalho também contribuem para classificação do *linkers* em L1, L2 e L3.

Fonte: Elaborado pela Autora (2022).

Observou-se na sequência do L3 a presença de duas regiões com repetições dos resíduos de ácido aspártico e histidina, localizadas especificamente entre as posições 167 e 184 e, na porção C-terminal (**Figura 17**). Após análise de BLAST não foi possível encontrar identidade com nenhuma sequência de *linker* de outras espécies de anelídeos disponibilizadas no banco de dados NR. Na tentativa de confirmar a ocorrência dessas regiões de repetição, foram desenhados *primers* específicos para sua amplificação e os

fragmentos gerados na PCR foram sequenciados. Porém, os resultados do sequenciamento não foram conclusivos e, portanto, optou-se por não os considerar nas análises e na construção da estrutura do complexo proteico.



Figura 17: Alinhamento das sequências dos *linkers* de HbAg e os *linkers* das Hbs homólogas HbGp e HbLt. Os resíduos de aminoácidos coloridos apresentam-se conservados entre as sequências e o sistema de cores do Clustal foi descrita no alinhamento da figura 16. A região destacada em quadro com borda azul corresponde ao motivo LDL-A (*low density lipoprotein receptor class-A*) presente nos *linkers*. Os aminoácidos destacados em caixa vermelha correspondem aos sítios de ligação aos íons Ca²⁺. Os aminoácidos destacados em caixa preta estão envolvidos na formação do motivo estrutural super hélice ou *coiled coil*.

Fonte: Elaborado pela Autora (2022).

O modelo de HbAg foi construído utilizando a estrutura de HbGp comparativamente, verificando-se que a estequiometria e a hierarquia da montagem do complexo entre as subunidades se mantiveram. Dessa forma a unidade biológica de HbAg contém 180 cadeias das quais 144 são globinas e 36 são *linkers* assim como observado para as proteínas HbGp e HbLt (BACHEGA et al., 2015; ROYER et al., 2006). As globinas são arranjadas em dodecâmeros formando a estrutura denominada de cap ($[abc]_3[d]_3$), onde as globinas $[abc]$ se unem através de pontes dissulfeto. O protômero de HbAg é formado com a união do cap e um heterotrímero de *linkers* (L1L2L3). A monocamada de HbAg é formada por 6 protômeros e a sobreposição de duas camadas formam a bicamada hexagonal contendo o total de 12 protômeros (WEBER; VINOGRADOV, 2001). A **Figura 18** é a representação esquemática do arranjo de HbAg partindo das unidades básicas até o formato de bicamada hexagonal.

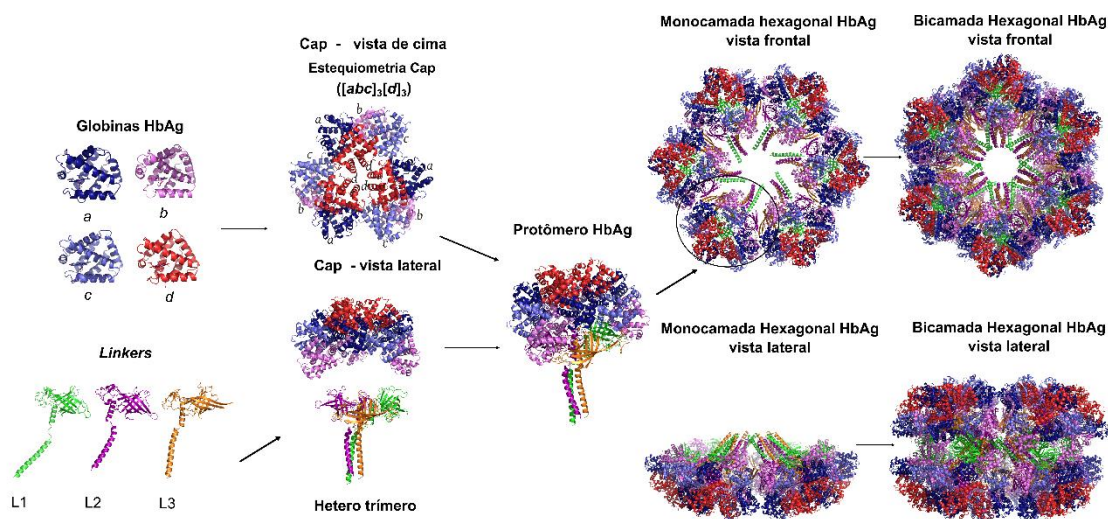


Figura 18: Esquema geral da organização do complexo proteico HbAg. Iniciando ao lado esquerdo da imagem tem-se a representação de todas as cadeias de HbAg; na parte superior foram desenhadas as globinas *a* (azul marinho), *b* (lilás), *c* (azul claro) e *d* (vermelho); e na parte inferior encontram-se os *linkers* L1 (verde), L2 (roxo) e L3 (laranja). Na região central da figura encontra-se representado o protômero, ele é formado por um dodecâmero denominado de *cap* com estequiometria $[abc]_3[d]_3$ e um trímero de *linkers*. Seguindo na sequência da figura, do lado superior direito foram desenhadas a monocamada (arranjo de seis protômeros) e a bicamada hexagonal (arranjo de 12 protômeros) com vista frontal e lateral respectivamente. Os esquemas do complexo proteico foram elaborados com o programa PyMOL versão 2.5

Fonte: Elaborado pela Autora (2022).

É sabido que as globinas tem relação direta com a função de carreamento de O_2 através da corrente sanguínea, pois são elas que carregam o grupo Heme em sua estrutura (**Figura 19**). Em invertebrados e vertebrados, as Hbs possuem na estrutura terciária de suas globinas um enovelamento que ocorre nas mioglobinas. Esse enovelamento consiste na presença de seis a oito segmentos de hélices conectados por *loops* curtos (WEBER; VINOGRADOV, 2001).

Globinas HbAg

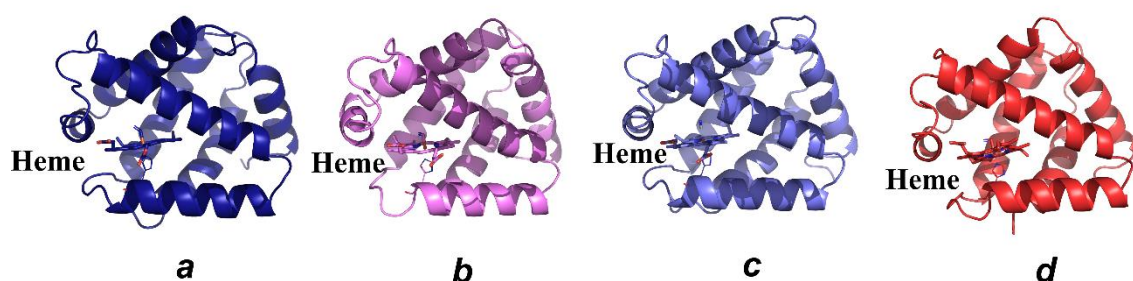


Figura 19: Representação das quatro globinas de HbAg [abcd]. Cada uma das globinas está representada com seu respectivo grupamento Heme, destacando a His proximal que se conecta ao anel porfirínico. A globina *a* foi representada pela cor azul marinho, a globina *b* pela cor lilás, a globina *c* pela cor azul claro e a globina *d* pela cor vermelha. A representação das globinas foi realizada com o programa PyMOL versão 2.5.

Fonte: Elaborado pela Autora (2022).

Como pode-se notar, os *linkers* tem papel fundamental na formação da estrutura quaternária das HBL-Hb; cada cadeia de *linker* é composta estruturalmente por duas α -hélices na porção N-terminal, um domínio LDL-A e um domínio β -barrel com oito fitas antiparalelas (BACHEGA et al., 2015; ROYER et al., 2006). Os *linkers* agem na formação dos protômeros, especificamente na união do cap com o trímero de linkers que se dá na região dos domínios de LDL-A e β -barrel. Além disso, os *linkers* estão envolvidos no primeiro contato entre os protômeros da primeira camada hexagonal e também na formação do sanduíche entre as duas camadas hexagonais de protômeros, pois a interação entre essas duas camadas ocorre apenas através da interação dos *linkers*. Ou seja, os *linkers* mantêm os dois discos juntos na formação da bicamada hexagonal, dando suporte ao modelo

bracelete encontrado nessas proteínas (BACHEGA et al., 2015; CHEN et al., 2015; ROYER et al., 2006).

O heterotrímico de *linkers* é formado devido a interação das porções de α -hélices na região N-terminal que se arranjam de maneira a formar uma região de *coiled-coil* ou super-hélices. A região do *coiled-coil* é formada pelos resíduos hidrofóbicos presentes nas α -hélices dos *linkers* L1, L2 e L3; essa região é caracterizada por septetos típicos de *coiled-coil* (*abcdefg*) e geralmente as posições *a* e *d* são ocupadas pelos resíduos hidrofóbicos. A **Figura 20** destaca em sua porção central a região *coiled-coil* formada entre os 3 *linkers* de HbAg (**Figura 20B**). Nela estão destacados alguns resíduos de aminoácidos que interagem na formação do *coiled coil* nas posições *a1*(**Figura 20C**) e *d1*(**Figura 20D**). Leucinas e isoleucinas são os resíduos mais abundantes, corroborando com os dados encontrados para a estrutura de HbGp (BACHEGA et al., 2015). Os resíduos que ocupam as posições *a* e *d* mantiveram-se conservados em relação a HbGp com exceção da Val-17, Leu-24, Leu-34, Val-57 no *linker* L1; Ile-22, Ile-29, Leu-55, Leu-65 em L2, e Val-13, Thr-20, Val-23, Phe-30 em L3 (**Figura 17**). Porém alguns desses resíduos que não estão conservados com HbGp apresentaram-se conservados em relação a HbLt: Val-17 em L1; Leu-55 em L2 e Thr-20 em L3.

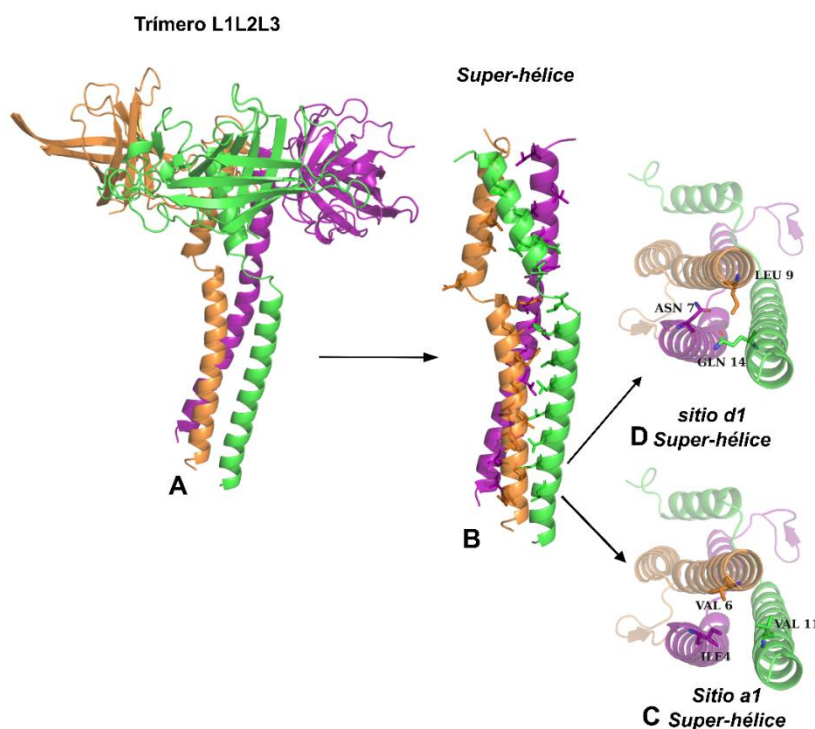


Figura 20: Detalhes da região *coiled coil* ou super hélice formada pela interação dos *linkers* L1, L2 e L3. (A) Hetetotrímero de *linkers*. (B) Região do *coiled coil* com os resíduos hidrofóbicos ocupados nas regiões *a* e *d* destacados. (C) Detalhes do sitio *a1* da super hélice e (D) detalhes do sitio *d1* da super hélice. Os esquemas apresentados nesta figura foram elaborados utilizando o programa PyMOL versão 2.5.

Fonte: Elaborado pela Autora (2022).

A formação do complexo HbAg através da sobreposição das duas camadas hexagonais ocorre por meio de contatos realizados na região do *coiled coil* entre os *linkers* L2 e L3 (**Figura 21C**) e os domínios β -barrel presentes nos *linkers* L1 de camadas diferentes (**Figura 21B**). Comparando a região de contatos estabelecidos entre os *linkers* L2 e L3 de HbAg com a mesma região em HbGp, notou-se a presença de um resíduo Asn-16 ao invés de uma Arg presente no L2 de HbGP. Em L3 a His-18 e Glu-11 são conservados e há presença de uma Ser-7 em HbAg no lugar de um Asp em HbGp (BACHEGA et al., 2015) (**Figura 21C**). A associação entre os protômeros vizinhos e subjacentes é feita entre L1 de diferentes camadas (**Figura 21E**), sendo coordenada pelos resíduos Tyr-9, Asn-13 e Gln-14 (**Figura 21F**); estes resíduos se conservam com o resíduos de HbGp (BACHEGA et al., 2015).

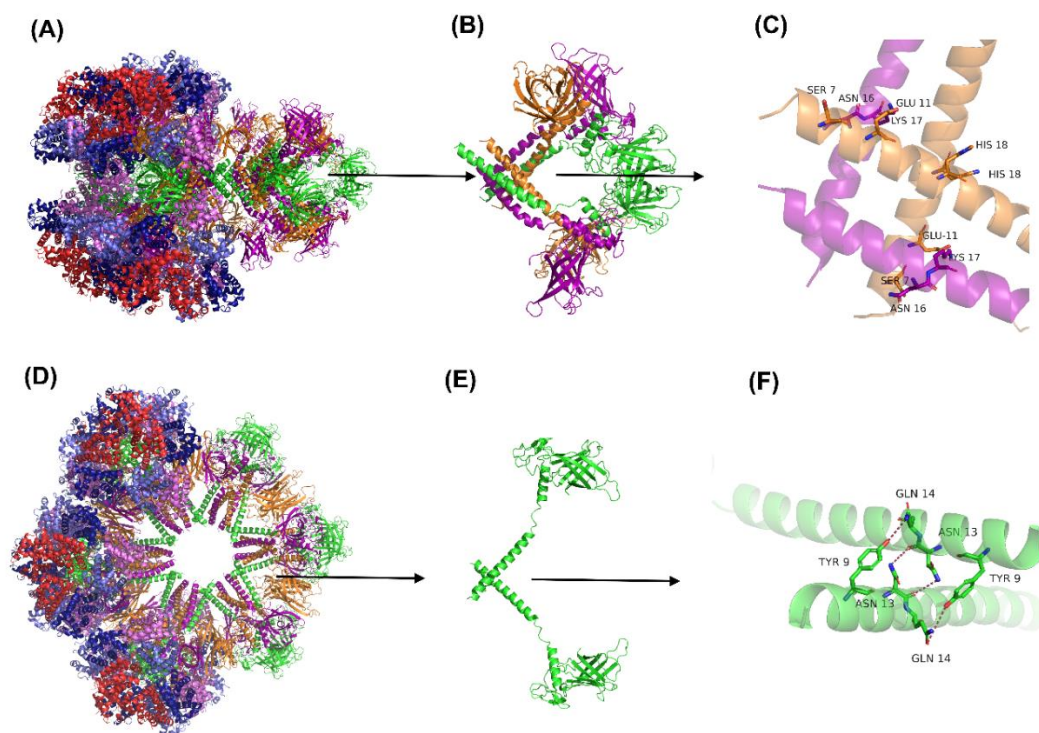


Figura 21: Esquema detalhando os contatos realizados entre os *linkers* na formação da bicamada hexagonal de HbAg. Parte da estrutura dos protômeros foi suprimida da figura para melhor visualização dos contatos entre os *linkers*. (A) Bicamada Hexagonal vista lateralmente com parte da região de protômeros suprimida. (B) Interação entre trimeros de *linkers* de camadas diferentes. (C) Detalhes dos contatos estabelecidos entre os *linkers* L2 provenientes de camadas diferentes, destacando-se os resíduos de aminoácidos que participam desta interação. (D) Bicamada Hexagonal vista frontal com parte da região de protômeros suprimida. (E) Contato entre dois *linkers* L1 de diferentes camadas. (F) Detalhes dos contatos estabelecidos entre os diferentes L1 provenientes de camadas distintas, destacando-se os resíduos de aminoácidos que participam desta interação. Os esquemas apresentados nesta figura foram elaborados utilizando o programa PyMOL versão 2.5.

Fonte: Elaborado pela Autora (2022).

A presença dos sítios de ligação a Ca^{2+} já foi citada anteriormente; esses sítios são conservados nas sequências de HbAg e encontram-se no domínio LDL-A (**Figura 22**). Os íons Ca^{2+} agem na regulação das propriedades alostéricas das Hbs gigantes e na formação de sua estrutura quaternária (ROYER et al., 2006). Os sítios de ligação aos íons Ca^{2+} são encontrados numa região de interface entre os *linkers* e a globina *b*. A união do cap com o heterotrímero para formação do protômero está relacionada a esses sítios, os quais conferem maior estabilidade a molécula (BACHEGA et al., 2015; ROYER et al., 2006). Os resíduos envolvidos nesta interação podem ser visualizados na **Figura 22**. Nota-se a Arg-34 da globina *b* que interage com Asp-84, Asp-88, Glu-95, Asp-94 em L1; Asp-74, asp-78, gGu-85, Asp-84 em L2; Asp-76, Asp-80, Glu-87, Asp-86 em L3.

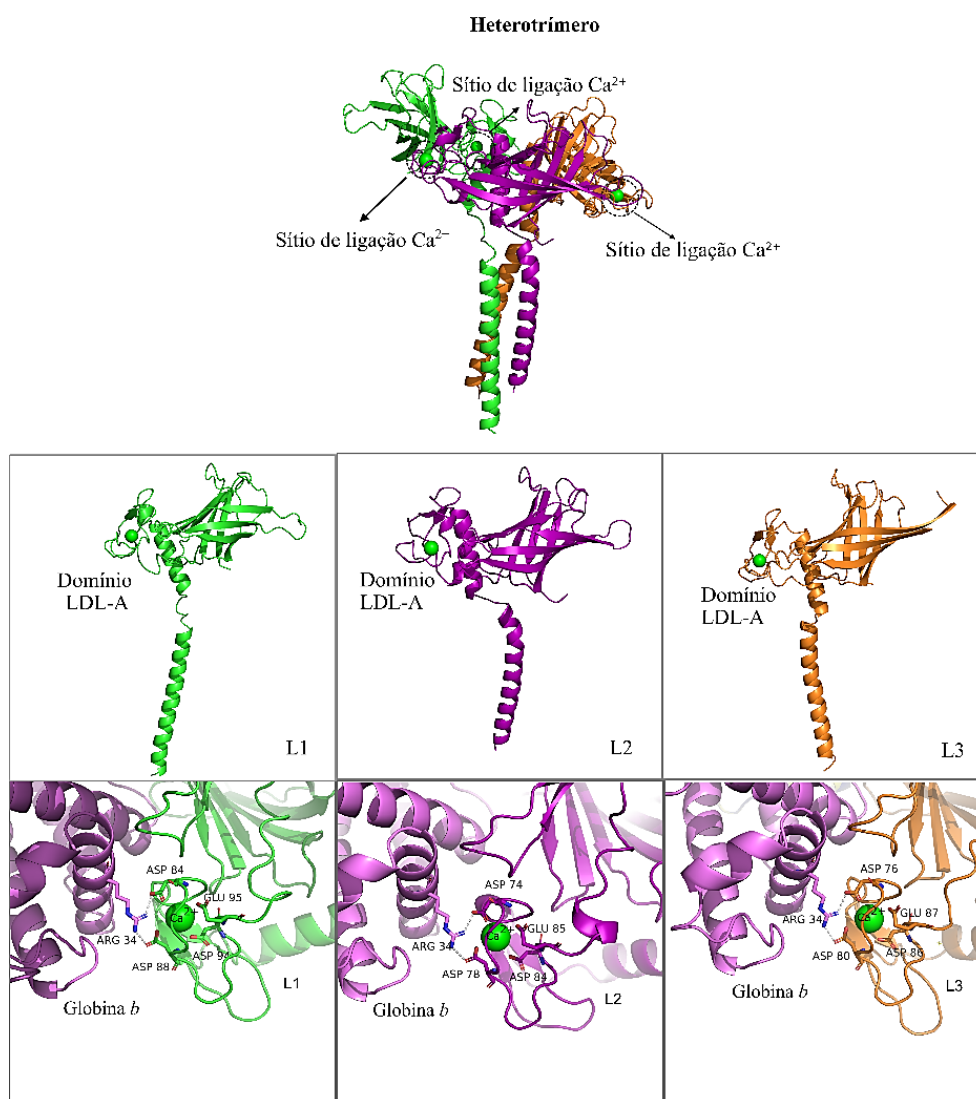


Figura 22: Sítios de ligação aos íons Ca^{2+} em L1, L2 e L3. Os íons de Ca^{2+} foram representados com esferas verdes e estão destacados em círculos com borda tracejada preta. No topo da figura está representado o heterotrímtero de *linkers* onde podem ser visualizados os sítios de Ca^{2+} em cada um dos *linkers*. Em seguida os *linkers* L1, L2 e L3 estão representados separadamente indicando os sítios de ligação a Ca^{2+} na região do domínio LDL-A. A parte inferior da figura mostra a região de interface entre a globina b e os linkers L1, L2 e L3. Os resíduos que coordenam essa ligação foram indicados na figura. Os esquemas apresentados nesta figura foram elaborados utilizando o programa PyMOL versão 2.5.

Fonte: Elaborado pela Autora (2022).

Acredita-se que além dos sítios de ligação a Ca^{2+} , HbAg também preserva dois sítios de ligação a Zn^{2+} , um deles localizado em L3 e outro na interface entre L3 e L2 (**Figura 23**). De forma análoga aos sítios de Ca^{2+} , os sítios Zn^{2+} também conferem maior estabilidade a molécula e podem estar relacionados a atividade peroxidase presente nessas

Hbs (BACHEGA et al., 2015; ROYER et al., 2006). Moléculas de HbGp pré-tratadas com EDTA tiveram a remoção dos íons de zinco promovendo mudanças na estabilidade da proteína assim como dissociação oligomérica, aumentando o número de espécies de dodecâmeros em solução (CARUSO et al., 2016).

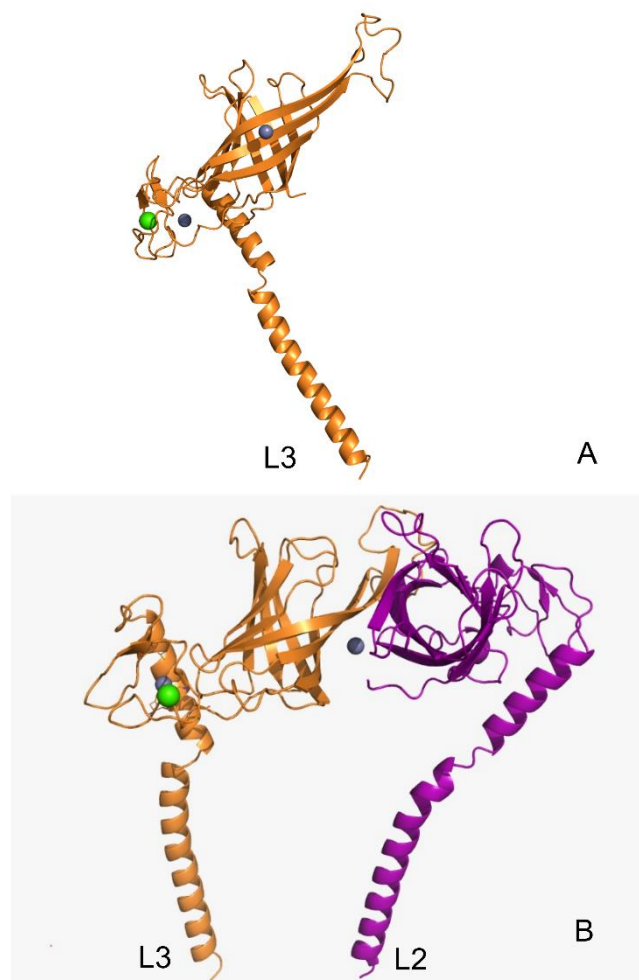


Figura 23: Representação da localização dos sítios de ligação aos íons de Zn^{2+} em HbAg. (A) *Linker* L3 com sítios de ligação aos íons Ca^{2+} e Zn^{2+} . A esfera verde representa o íon Ca^{2+} e as esferas cinzas representam os íons Zn^{2+} . (B) Destaque para a região de interface entre os *linkers* L3 e L2 que o segundo sítio de ligação a Zn^{2+} ocupa. Os esquemas apresentados nesta figura foram elaborados utilizando o programa PyMOL versão 2.5.

Fonte: Elaborado pela Autora (2022).

Alguns estudos apontam as hemoglobinas oriundas de espécies de anelídeos como bioindicadores de contaminação ambiental, em específico a metais pesados. Dados de transcriptoma para o poliqueta marinho *Ophelina* sp.1 submetido a ambientes com alta concentração de Cu revelaram aumento na expressão dos transcritos das globinas do tipo *b* (NEAVE et al., 2012). Padrão semelhante foi encontrado para o poliqueta *Alitta virens*, que foi submetido a diferentes concentrações de Cu e Zn e também tiveram um aumento na expressão de cadeias de globinas e *linkers* (GREEN ETXABE et al., 2021). Os autores que pesquisaram o transcriptoma de *A. virens* sugerem que as Hbs produzidas a partir de genes induzidos, em resposta a algum tipo de estresse, podem ser resistentes aos danos causados por espécies reativas ao oxigênio (ROS), alteração na afinidade por O₂ ou conversão da Hb em metahemoglobina (GREEN ETXABE et al., 2021). A exposição de minhocas *L. terrestris* a metais pesados como Cd, Mg e Cu seguiu do aumento concomitante da concentração de hemoglobina e é sugerido que exista uma afinidade entre os metais pesados e os resíduos de cisteínas livres presentes nas HBL-Hb (CALISI et al., 2011, 2013).

Os dados reportados aqui trazem informações básicas a respeito da constituição de aminoácidos presentes nas globinas e *linkers* e da conformação estrutural de HbAg. Essas informações podem ser usadas como suporte para trabalhos futuros que visem um aprofundamento da relação das Hb com a exposição a diferentes condições de estresse ambiental, tais como exposição aos efeitos de metais pesados.

4 CONCLUSÕES

Todas as sequências que compõem HbAg, incluindo globinas e *linkers*, foram identificadas. As sequências apresentaram elevado grau de conservação entre suas proteínas homologas HbGp e HbLt. As globinas apresentam uma sequência de resíduos de aminoácidos que são conservados dentro do grupo das HBL-Hb e o enovelamento das globinas seguiu conforme o enovelamento presente nas mioglobinas, característica essa encontrada em Hbs de vertebrados e invertebrados. A formação do complexo proteico seguiu a mesma ordem e hierarquia proposta nas estruturas de HbGp e HbLt e, portanto, HbAg também apresentou um número de 144 globinas e 36 *linkers*, totalizando um número de 180 cadeias com uma estequiometria de 12 ($[abc]_3[d]_3$) (L1L2L3). Enquanto as globinas carregam o grupo Heme e estão mais relacionadas com as funções que a HbAg desempenha, os *linkers* agem na formação e estabilização do complexo proteico. A partir

da construção do modelo foi possível checar a conservação dos domínios proteicos presentes na estrutura de HbAg. Os dados reportados aqui são corroborados por dados experimentais obtidos pelo nosso grupo de pesquisa e trazem informações importantes que poderão fomentar trabalhos futuros utilizando essa biomolécula.

CAPITULO 4 CARACTERIZAÇÃO BIOFÍSICA INICIAL DO MONÔMERO-D NATIVO DE HbAg.

Resumo

Estudar isoladamente uma cadeia de um complexo proteico pode trazer informações a respeito das dinâmicas envolvidas no processo de formação da proteína. Técnicas biofísicas como a espectroscopia são utilizadas para caracterizar proteínas e possuem como vantagem a fácil execução e rápida obtenção de resultados. Dessa forma, neste capítulo foi realizada a caracterização inicial do monômero *d* de HbAg através das técnicas de espectroscopia de absorção UV-Vis e fluorescência, além de medidas de espalhamento de luz (LSI). Para tanto, as medidas foram tomadas em dois conjuntos de experimentos, um envolvendo a caracterização do monômero – *d* em diferentes valores de pH e outro avaliando a interação do monômero *d* com o surfactante dodecil sulfato de sódio (SDS). Para os dados que avaliaram a variação do pH foi possível notar um processo de auto-oxidação do grupo Heme em valores de pH acima de 9,0. Os dados são complementares aos encontrados para o parâmetro intensidade de fluorescência, já que em valores de pH alcalinos os valores de intensidade de fluorescência foram maiores evidenciando uma maior desenovelamento/desnaturação da molécula. Para os valores de pH 5,0 e 9,0, abaixo e acima do ponto isoelétrico (pI) do monômero *d* (pI = 7.1), foram observados maiores valores de LSI o que pode estar relacionado com a formação de dímeros da cadeia *d*. No estudo de interação do monômero *d* com SDS, observou-se a oxidação do grupo Heme a partir das concentrações de 4,0 mmol L⁻¹. A partir da concentração de 1,2 mmol L⁻¹ de detergente ocorreu um aumento gradativo da intensidade de fluorescência, demonstrando uma maior desenovelamento/desnaturação da molécula o que pode estar relacionado com um início do seu processo de desenovelamento. A variação encontrada nos dados de LSI podem estar relacionadas com a formação de agregados entre o monômero-*d* e o surfactante assim como a presença de auto agregados de monômero em solução.

1 INTRODUÇÃO

Como já foi descrito neste trabalho sabe-se que as subunidades que compõem Hbs gigantes, como HbAg estão ligadas de maneiras distintas; enquanto os monômeros *a*, *b* e *c* são mantidos por ligações dissulfeto, o monômero-*d* encontra-se ligado às outras subunidades por interações eletrostáticas (CARVALHO et al., 2011b). Essas características físico-químicas permitem que ocorra a dissociação do monômero-*d*, após a exposição dessas proteínas a ambiente alcalino pH>9,0 (RUGGIERO BACHEGA et al., 2015).

A dissociação oligomérica das subunidades de Hbs gigantes é interessante pois é o primeiro passo para compreendermos os fenômenos envolvidos no processo de agregação das Hbs. Ao mesmo tempo, esse fenômeno possibilita os estudos de caracterização das cadeias isoladas, além de avaliar a possibilidade de utilização dos monômeros de forma isolada em aplicações biotecnológicas.

Após identificar as estruturas primárias das proteínas de interesse, podemos prosseguir as análises de caracterização a partir dos estudos de suas estruturas secundárias, terciárias e quaternárias. Para realização dessas caracterizações, a biofísica apresenta um grupo de técnicas, que conjuntamente, podem informações relevantes quanto a Massa Molecular (M), processos de dissociação estados de agregação, separação e identificação de subunidades, pI, grau de desnaturação, presença de diferentes estruturas secundárias entre outras características (CARVALHO et al., 2013; OLIVEIRA et al., 2020; SANTIAGO et al., 2010).

O surfactante SDS possui ação desnaturante e o estudo envolvendo este tipo de agente caotrópico permite investigar processos de desenovelamento de proteínas. Ao interagirem com estruturas proteicas são formados complexos entre o surfactante e a proteína e as mudanças estruturais desencadeadas por essas interações podem ser monitoradas com o auxílio de técnicas biofísicas (FOTOUHI et al., 2015).

Sendo assim, o objetivo deste capítulo foi apresentar algumas análises iniciais de caracterização biofísica através de técnicas de espectroscopia do monômero-*d* de HbAg em função dos valores de pH e interação desse monômero com o surfactante aniônico SDS.

2 MATERIAIS E MÉTODOS

2.1 Análises espectroscópicas

2.1.1 Espectroscopia de absorção UV-VIS

Mudanças conformacionais observadas mediante diferentes condições podem ser estudadas a partir de métodos espectroscópicos que são altamente sensíveis e rápidos. Devido à alta sensibilidade são requeridas amostras de trabalho com baixas concentrações de proteínas em solução. Os diferentes métodos de espectroscopia incluem absorção ótica UV-VIS, fluorescência intrínseca e extrínseca (utilização de sondas ligadas covalentemente ou não), dicroísmo circular (CD) e transformação de Fourier por espectroscopia de infravermelho. São métodos que quando comparados com a cristalografia ou NMR trazem um conteúdo informático menor, apresentando algumas limitações a respeito da estrutura tridimensional das proteínas (NEET; LEE, 2002).

Porém são métodos tradicionalmente utilizados para estudar as propriedades das moléculas em solução em experimentos não destrutivos e de baixa concentração. Podem ser monitoradas mudanças conformacionais através de dados de cinética e termodinâmica (variação de tempo e temperatura). As vantagens dos métodos espectroscópicos seriam a possibilidade da variação das condições experimentais (NEET; LEE, 2002).

A absorção ótica está relacionada com a capacidade que um material tem de absorver radiação a partir da exposição contínua dessa radiação. A radiação residual do material ao passar por um prisma originará o espectro de absorção. Ao absorver energia, átomos ou moléculas mudam de estados de energia, passando da condição de energia mais baixa (estado fundamental) para um estado de energia mais elevado (estado excitado). Para a espectroscopia UV-Vis (faixa de absorção de luz entre os comprimentos de onda de 190 a 800 nm) as transições de energia ocorrentes são de caráter eletrônico. Assim, a absorção de energia de uma molécula faz com que um elétron transite de um orbital ocupado para um orbital de maior energia potencial (PAIVA et al., 2010).

No caso das moléculas a absorção UV ocorre em uma faixa de comprimentos de onda, pois elas possuem vários modos excitados de vibração e rotação em temperatura ambiente. Isso acarreta um número de transições elevadas, o que poderia ser representado por um número de linhas diferentes muito próximas umas das outras e, como os aparelhos que medem absorção (espectrofotômetros) não conseguem definir todas essas transições o resultado da medição é a representação de uma banda larga de absorção centrada perto do comprimento de onda da principal transição (PAIVA et al., 2010).

A concentração de uma solução é proporcional a absorbância do material, característica regida pela Lei de Beer-Lambert (1), que descreve que o número de moléculas em uma solução em determinado comprimento de onda é proporcional a extensão de sua absorção, assim como quanto maior for a eficiência de uma molécula em

absorver luz de um certo comprimento de onda, maior será sua absorbância (GOLEB, 1966; STRONG, 1952; TRUMBO et al., 2013).

$$(1) A = \log (I_0/I) = \epsilon cl$$

Onde:

A: absorbância

I_0 : intensidade de luz incidindo na cela da amostra

I: intensidade de luz saindo da cela da amostra

ϵ : absortividade molar

c: concentração molar do soluto

l: caminho ótico

2.1.2 Espectroscopia de Fluorescência

O uso da técnica de fluorescência é amplamente difundida dentro da biotecnologia sendo utilizada na obtenção de diagnósticos médicos, sequenciamento de DNA, análises genéticas e estudos biofísicos de proteínas (LAKOWICZ, 2006). A fluorescência é um tipo de luminescência, fenômeno pelo qual ocorre a emissão de luz de uma substância após excitação dos seus estados eletrônicos. Quando a fluorescência ocorre os elétrons encontram-se emparelhados no estado singleto excitado, ou seja, os spins dos elétrons ocupam os orbitais de menor e maior energia em sentido anti-paralelo. Dessa forma, quando se retorna ao estado fundamental ocorre a emissão de um fóton. Duas propriedades da fluorescência são importantes para entendimento e utilização da técnica: o rendimento quântico e tempo de vida. O rendimento quântico é a razão dos fótons emitidos pelo número de fótons absorvidos. O tempo de vida está relacionado com o tempo que o fluoróforo possui para interagir ou se difundir no ambiente em que se encontra. O tempo de vida de um fluoróforo é obtido pela média entre seu estado excitado e seu estado fundamental, sendo em estado singleto na ordem de 10^{-5} a 10^{-9} s. Diferentes medições de fluorescência podem ser obtidas a partir de equipamentos específicos, podendo ser classificadas como fluorescência estática ou resolvida no tempo. Como a escala de tempo é extremamente curta, as medições de emissão de fluorescência resolvida no tempo requerem equipamentos mais sofisticados (LAKOWICZ, 2006; SKOOG; HOLLER; NIEMAN, 2018).

A fluorescência ocorre basicamente em moléculas com anéis aromáticos sendo mais intensa em grupos aromáticos com transições $\pi \rightarrow \pi^*$ de baixa energia. A fluorescência aumenta com o número de anéis e seu grau de condensação. Espectros de emissão de fluorescência possuem variação ampla e depende da natureza dos fluoróforos e do solvente da solução. Compostos alifáticos, alicíclicos carbonílicos ou estruturas de ligações duplas

altamente conjugadas também podem apresentar fluorescência (SKOOG; HOLLER; NIEMAN, 2018).

O fenômeno da fluorescência pode ser melhor compreendido através do Diagrama de Jablonski (JABŁOŃSKI, 1935) (**Figura 24**). O estado fundamental singlete é representado por S_0 e conforme vão subindo os níveis de energia a representação muda para S_1 e S_2 . Neste diagrama não estão representadas as interações do tipo *quenching*, energia de transferência e interações com solvente. Como pode ser observado, após a excitação das moléculas por uma fonte de luz podem ocorrer diversos processos. Uma parte da energia que incide nas moléculas é absorvida passando dos estados S_0 para S_1 e S_2 . A emissão de fluorescência ocorre com a perda de energia quando os elétrons mudam do estado excitado S_1 para o estado fundamental S_0 . Nas conversões internas as moléculas caem de um nível de maior energia para de menor energia sem efeitos de emissão de radiação. Cruzamentos intersistemas ocorrem quando há inversão do spin do elétron excitado passando do estado singlete para o tripleto. A fosforescência ocorre com a perda de energia entre estados tripletos. Outros processos não radioativos que não estão representados na **Figura 24** são relaxação vibracional e conversão externa.

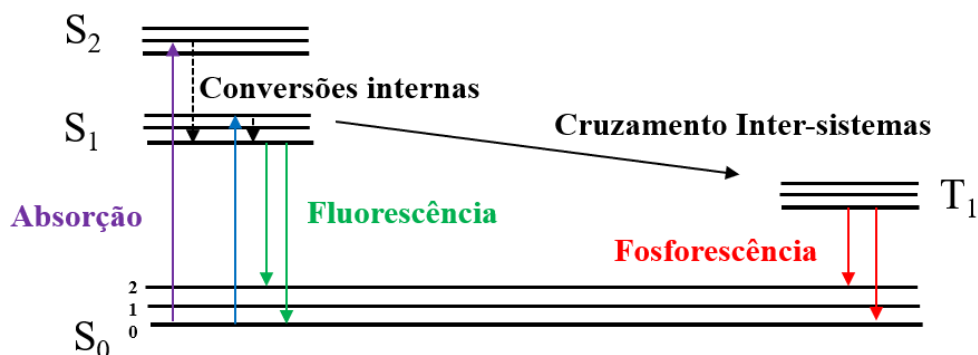


Figura 24: Representação esquemática do Diagrama de Jablonski. Neles estão representados os processos de Absorção – seta direcionada do sentido inferior para o superior na cor lilás ($S_0 \rightarrow S_2$) e azul ($S_0 \rightarrow S_1$); Conversão Interna – Setas tracejadas em preto direcionadas do sentido superior para o inferior ($S_2 \rightarrow S_1$); Fluorescência – Setas verdes direcionadas do sentido superior para o inferior ($S_1 \rightarrow S_0$); Cruzamentos intersistemas – seta preta diagonal partindo do estado singlete para o tripleto (T_1); Fosforescência – Setas vermelhas direcionadas do sentido superior para o inferior ($T_1 \rightarrow S_0$). S_0 representa o estado singlete fundamental; S_1 e S_2 representam os estados singlete excitados; T_1 representa o estado tripleto excitado. Os números de 0 a 2 representam os diferentes níveis de energia dentro de um mesmo estado.

Fonte: (LAKOWICZ, 2006). Adaptado pela Autora (2022).

A energia de emissão é menor que a energia de absorção, sendo assim a fluorescência ocorre em comprimento de ondas mais longos e por conseguinte em níveis energéticos mais baixos. Quem observou este comportamento foi George Gabriel Stokes, denominando este comportamento como deslocamento de Stokes (STOKES, 1852). O deslocamento de Stokes pode estar relacionado com a relaxação vibracional, efeitos do solvente, reações dos estados excitados, formação de complexos e transferência de energia.

A técnica da fluorescência pode ser utilizada com fluoróforos intrínsecos e extrínsecos. Os extrínsecos são utilizados como sondas adicionadas às moléculas e/ou sistemas a serem monitorados. Os intrínsecos são aqueles presentes na constituição das moléculas; como exemplo nas proteínas, temos os triptofanos que são aminoácidos com um grupo indol que absorve luz próximo ao comprimento de onda de 280 nm e emite no comprimento próximo a 340 nm. O grupo indol é sensível a polaridade do solvente, sendo assim o espectro de emissão do grupo indol pode revelar a localização dos resíduos de triptofano nas proteínas. Dessa forma a exposição desses resíduos à fase aquosa solvente pode relacionar-se com processos de desdobramento da proteína, como desnaturação ou dissociação (LAKOWICZ, 2006).

2.2 Extração e purificação da HbAg

A extração da hemolinfa de *A. gracilis* foi realizada com um número em torno de 100 indivíduos e cada um deles foi anestesiado com água gelada por um período de 15 a 20 min. Com o auxílio de um bisturi foi feito um corte na parte superior da minhoca, numa região próxima a cabeça para evitar a contaminação da hemolinfa com enzimas digestivas. Em seguida a hemolinfa foi coletada com uma pipeta pasteur até o esgotamento do animal adicionando citrato de sódio 0,1 mmol L⁻¹ para evitar a coagulação.

A hemolinfa foi centrifugada, a 5.000 rpm, por 15 min e a 4° C, para retirar restos de células e resíduos, seguido por 12 horas de diálise em uma membrana de corte de 30 kDa contra tampão tris-HCl 100 mmol L⁻¹, pH 7,0. Após a diálise, o material foi ultracentrifugado a 60.000 rpm por 5 horas a 4° C para sedimentar a HbAg. A proteína foi obtida na forma de *pellet*, sedimentada no fundo do tubo, sendo ressuspensa em um volume de 3 – 4 mL de tampão tris-HCl 100 mmol L⁻¹, pH 7,0, e armazenada a 4°C na geladeira.

A próxima etapa foi a purificação da HbAg através de cromatografia líquida por exclusão de tamanho em uma coluna de gel filtração Superdex Hiload 16/600 200 pg GE Helthcare. Para tanto foram preparadas amostras com concentração de 25 mg mL⁻¹ injetadas em um *loop* de 2 mL em um cromatógrafo modelo Äkta Pure 8 GE Helthcare. O fluxo de trabalho utilizado foi de 1,5 mL min⁻¹.

Todas as concentrações foram determinadas espectroscopicamente utilizando o valor de absortividade molar $\epsilon_{415\text{ nm}} = 5,1 \pm 0,6 \text{ (mg mL}^{-1}\text{) cm}^{-1}$ determinado pelo método de Lowry. O valor encontrado de absortividade molar para a HbAg em 415 nm está de acordo com o encontrado para a HbGp (CARVALHO et al., 2011b; CARVALHO; SANTIAGO; TABAK, 2012). A HbAg foi armazenada a 4° C para os posteriores ensaios de caracterização.

2.3 Dissociação e purificação do monômero-*d* de HbAg

Para obtenção do monômero-*d* de HbAg foi utilizada a metodologia descrita para outras Hbs como HbGp e HbLt (BACHEGA, 2013). Para tanto, utilizou-se uma alíquota 1000 µL de HbAg na concentração de 15 mg mL⁻¹. O pH da amostra foi elevado a 9,6, deixando a amostra em repouso por cerca de 17 horas. Transcorrido o período de incubação da amostra, o pH da mesma foi ajustado para o valor de 7,0 e a Hb foi submetida à centrifugação a 10.000 rpm por 10 minutos a 4° C. A manutenção da HbAg em meio alcalino promove à sua dissociação, permitindo assim o isolamento do monômero-*d* das demais cadeias de globinas. Como mencionado anteriormente, o monômero-*d* mantém se ligado as demais cadeias globínicas por interações eletrostáticas.

Após o período de dissociação, a amostra de trabalho (pH 7,0) foi purificada através da técnica de cromatografia líquida por exclusão de tamanho em uma coluna de gel filtração Superdex Hiload 16/600 200 pg GE Helthcare. Dessa forma, foram injetados 500 µL de HbAg (concentração de 3 mg mL⁻¹) em um Cromatógrafo modelo Äkta Pure 8 GE Helthcare. O fluxo de trabalho utilizado foi de 0,9 mL min⁻¹.

Ao final da etapa de cromatografia foram coletadas as amostras correspondentes aos picos extraídos do cromatograma e feita uma corrida de eletroforese em gel de poliacrilamida SDS-PAGE 15% para verificar a eficiência da purificação.

2.4 Análises de Espectroscopia UV-VIS da subunidade-*d* de HbAg em função do pH e diferentes concentrações de SDS.

O preparo das amostras submetidas à análise de Espectroscopia de luz UV-VIS foi realizado através da diluição do monômero-*d* purificado. Assim, as amostras contendo a subunidade-*d* foram diluídas à concentração de $0,1 \text{ mg mL}^{-1}$ em tampão acetato-fosfato-borato 30 mM. Foram preparadas cinco diferentes amostras, nos valores de pH 5,0; 6,0; 7,0; 8,0 e 9,0. As amostras permaneceram em equilíbrio (mantidas em repouso a 25°C) por 3 horas e as medidas foram tomadas a cada hora de equilíbrio, ou seja, após 1, 2 e 3 horas de equilíbrio. A leitura foi realizada através do Espectrofotômetro Shimadzu UV 1800 com varredura entre os comprimentos de onda de 250 a 700 nm. Os dados foram tratados através do programa de Software OriginPro 8.5.

Para a titulação entre o monômero e as diferentes concentrações de SDS as amostras foram preparadas em tampão acetato-fosfato-borato 30 mM pH 7,0 com concentração final de $0,1 \text{ mg mL}^{-1}$ de monômero-*d* de HbAg. As concentrações de SDS variaram de 0 a $8,0 \text{ mmol L}^{-1}$, onde foram utilizados 9 pontos (0; 0,4; 0,6; 0,8; 1,5; 2,0; 4,0; 6,0; $8,0 \text{ mmol L}^{-1}$). O tempo de equilíbrio destas reações também foi de 3 horas.

2.5 Análises de Fluorescência e Espalhamento de Luz (LSI) da subunidade-*d* de HbAg

Para realização das análises de fluorescência e espalhamento de Luz LSI (*light scattering intensity*) foram utilizadas amostras de monômero-*d* $0,1 \text{ mg mL}^{-1}$ em tampão acetato-fosfato-borato 30 mM. O preparo das amostras foi realizado como descrito no item anterior desta sessão. Dessa forma, as análises foram processadas com as amostras para titulação em função do pH, partindo do valor de 5,0 até o pH 9,0 e, para a titulação com SDS os valores do surfactante variaram de 0 a $8,0 \text{ mmol L}^{-1}$. O equipamento utilizado para fazer as leituras de fluorescência e LSI foi um espectrofluorímetro Shimadzu RF 6000. Com intuito de monitorar acessibilidade do solvente aquoso ao triptofano da proteína, os parâmetros utilizados para as análises foram comprimento de onda excitação de 295 nm e varredura de emissão entre 305 e 450 nm. Para o LSI a excitação e emissão foram fixados em 350 nm. Os dados foram tratados através do programa de Software OriginPro 8.5. Todas essas análises biofísicas foram medidas em uma cubeta de quartzo, de caminho ótico de 1,0 cm.

3 RESULTADOS E DISCUSSÃO

3.1 Dissociação e purificação do monômero-*d* de HbAg

A dissociação do monômero-*d* de HbAg ocorreu em meio alcalino e a **Figura 25(A)** representa o cromatograma da HbAg dissociada. Os picos foram enumerados de I a IV em algarismo romanos, e pode se notar que não ocorreu uma separação nítida para o pico III e por isso o conteúdo eluído foi descartado. Devido a técnica de cromatografia utilizada (exclusão por tamanho), foram atribuídos aos picos I e II uma mistura da proteína íntegra (maior MM). O pico referente ao monômero *d*- de HbAg é IV que foi eluído num volume em torno de 90 mL.

A **Figura 25(B)** ilustra o perfil de deslocamento em eletroforese SDS-PAGE das frações de HbAg coletadas após a cromatografia. A coluna **1** corresponde à HbAg íntegra diluída a $0,3 \text{ mg mL}^{-1}$, as colunas **2** e **3** são as frações I e II coletadas na cromatografia e possuem o perfil eletroforético semelhante ao perfil da proteína íntegra. A fração do monômero-*d* (Fração IV) encontra-se na coluna **4**, onde pode se observar a presença de uma banda forte entre 10 e 15 kDa, tamanho característico para as globinas de hemoglobinas extracelulares gigantes. Os perfis eletroforéticos exibidos correspondem ao encontrado na literatura para outras hemoglobinas gigantes (CARVALHO et al., 2011a). A presença de duas bandas na altura correspondente ao tamanho molecular da subunidade *d* pode estar relacionada com a ocorrência de uma isoforma do monômero. Dados de MALDI-TOF para amostra de HbAg na presença de mercaptoetanol revelaram a existência de isoformas da subunidade *d* em torno de 16 kDa (OLIVEIRA et al., 2020). Além disso, os dados de RNA-Seq reportados neste trabalho também corroboram com a existência de isoformas para o monômero. Considerando que para as Hbs gigantes os valores de MM das globinas variam de 14 a 17 kDa e para os *linkers* estão em torno de 24 a 25 kDa (HACKERT; RIGGS, 2006; OLIVEIRA et al., 2020; SUZUKI; VINOGRADOV, 2003) podemos observar que nas colunas **1** e **2** correspondentes à proteína íntegra existe a presença de quatro bandas, sendo que a banda em torno de 75 kDa pode representar os tetrâmeros [*abcd*], as bandas localizadas entre 37 e 30 kDa possivelmente são *linkers* e a banda que aparece isolada entre 15 e 10 kDa corresponde a uma das subunidades monoméricas. Na coluna **3** é nítida a presença da banda correspondente ao tetrâmero de globinas, uma das bandas dos que representam os *linkers* e uma banda referente a subunidade. A natureza da interação do monômero-*d* com as demais subunidades se dá por interações eletrostáticas (BACHEGA et al., 2015; ROYER et al., 2006) e dessa forma a

dissociação do monômero em meio alcalino é possível, sendo assim pode se inferir que a banda localizada entre 15 e 10 kDa é correspondente ao monômero-*d*.

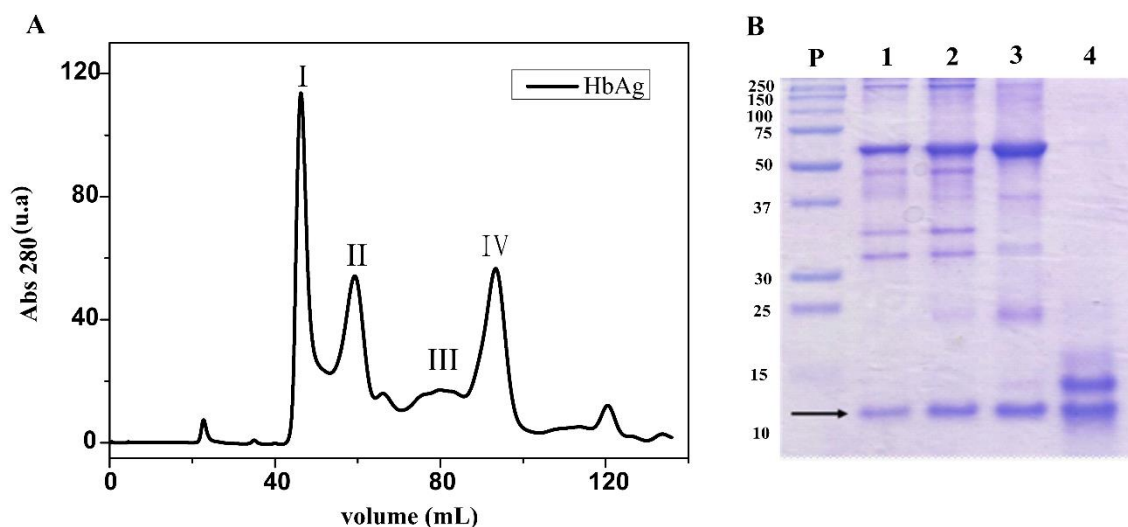


Figura 25: Purificação do monômero-*d* de HbAg: Cromatografia e Gel SDS-PAGE.

(A) Cromatograma gerado após dissociação do monômero-*d* de HbAg, onde o eixo x representa o volume de eluição da amostra durante o processo de purificação e o eixo y indica a absorvância em 280 nm. Os principais picos foram enumerados de I a IV sendo os picos I e II atribuídos à HbAg íntegra e o pico IV ao monômero-*d*. (B) Gel de poliacrilamida SDS-PAGE 15%; P: padrão de peso molecular MM relativa; 1: HbAg íntegra sem ter sido submetida ao processo de dissociação em pH alcalino; 2: amostra coletada do pico I do cromatograma de dissociação; 3: amostra coletada do pico II do cromatograma de dissociação; 4: amostra coletada do pico IV do cromatograma de dissociação. As amostras aplicadas no gel foram preparadas na ausência de β -mercaptoetanol.

Fonte: Elaborado pela Autora (2022).

É importante salientar que as etapas de dissociação e purificação do monômero-*d* foram desafiadoras devido ao baixo rendimento gerado ao fim de todo o processo. Durante o processo de dissociação ocorreram algumas perdas de proteína, após a incubação da amostra em pH 9,6 por 17 horas e retorno para o pH 7 ocorreu a formação de um precipitado, provavelmente proteína íntegra que não se dissociou por completo. Partindo da concentração de trabalho de 10 mg obteve-se 0,36 mg de monômero-*d*, ou seja, o equivalente a apenas 3,6 % da proteína. Como pode ser observado nos dados a seguir para variação de pH e uso do surfactante SDS, a quantidades de pontos (titulação de monômero-*d* versus SDS) utilizada foi devido ao baixo rendimento de obtenção do monômero-*d*. Além disso, a partir da análise da **Figura 25B** observa-se que o monômero *d* encontra-se parcialmente puro, porém análises biofísicas foram realizadas com esta fração. Parâmetros utilizados na etapa de purificação como taxa de fluxo de corrida, tempo de dissociação do

monômero em pH alcalino, concentração da amostra, tratamento das amostras com EDTA foram alterados com o intuito de melhorar a etapa de purificação, porém mesmo com as mudanças citadas não possível obter o monômero *d* de HbAg totalmente puro.

3.2 Análises espectroscópicas do monômero-*d* de HbAg em função do pH.

O estado de oxidação do grupo Heme de HbAg foi monitorado por espectroscopia de absorção UV-Vis. Esse processo de oxidação do grupo Heme ocorre quando os íons Fe^{2+} (estado ferroso) são oxidados e passam para forma Fe^{3+} (estado férrico), sendo que a Hb passa do estado de oxi-Hb para meta-Hb. Ao ocorrer o processo de oxidação a sexta coordenação do grupo Heme passa a ser ocupada por uma molécula de H_2O no lugar de O_2 , sendo classificada como aquometa-Hb, caso a sexta coordenação for ocupada pelos íons cianeto CN^- ela passa a se chamar cianometa-Hb e Também pode existir a formação da espécie hemicromo (RIFKIND et al., 1994). O espectro de absorção de ótica característico para as Hbs extracelulares gigantes em sua forma oxi se caracteriza por apresentar um pico intenso e estreito em torno de 414 nm chamado de banda de Soret, seguido de duas bandas Q em 535 e 575 nm respectivamente **Figura 26** (LE PAPE et al., 2015). Pode ocorrer uma variação pequena desses valores de absorção de luz, conforme a espécie de anelídeo na qual foi extraída a Hb (HARNOIS et al., 2009).

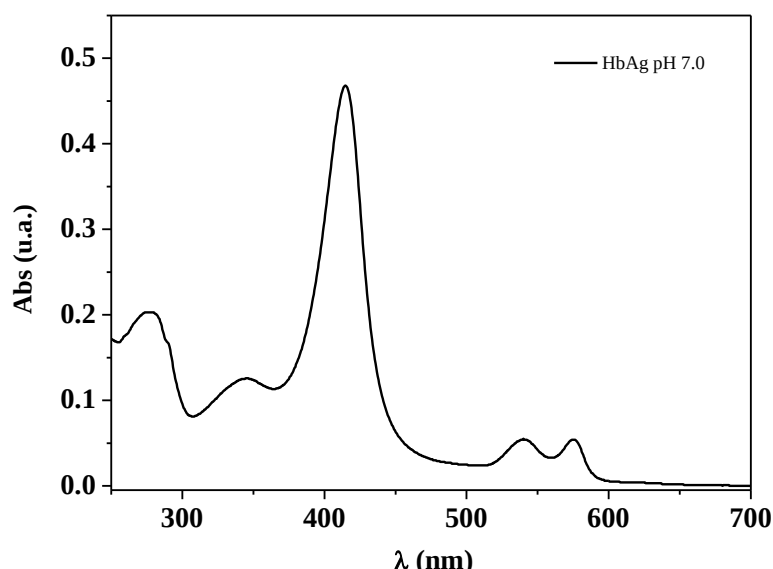


Figura 26: Espectro de absorção de luz UV-VIS para oxi-HbAg em pH 7,0. Os comprimentos de onda foram monitorados na faixa de 250 a 700 nm. O pico estreito em 414 nm representa a Banda de Soret e os picos em 535 e 575 são conhecidos como bandas Q.

Fonte: Elaborado pela Autora (2022).

Como pode ser observado na **Figura 27** houve uma variação no monômero-*d* de HbAg conforme alteração no pH da solução. O tempo de equilíbrio é um fator importante na interpretação dos resultados obtidos, pois como pode-se notar, nas amostras que permaneceram em equilíbrio por apenas uma hora (**Figura 27A**) os resultados apresentados são discordantes aos resultados obtidos após 2 e 3 horas de equilíbrio. Transcorrida uma hora após o preparo da amostra observa-se que a curva correspondente ao pH 5 se sobrepõe a curva referente ao pH 7, já para os gráficos de duas e três horas de equilíbrio isso não é observado. O maior alargamento da banda correspondente à absorção ótica do grupo Heme para o pH igual a 9 fica mais evidente após três horas de equilíbrio da amostra (**Figura 27C**). Assim pode-se inferir que resultados mais conclusivos podem ser obtidos a partir do equilíbrio de 3 horas de reação, e, portanto, esse tempo foi adotado como padrão para os demais experimentos apresentados.

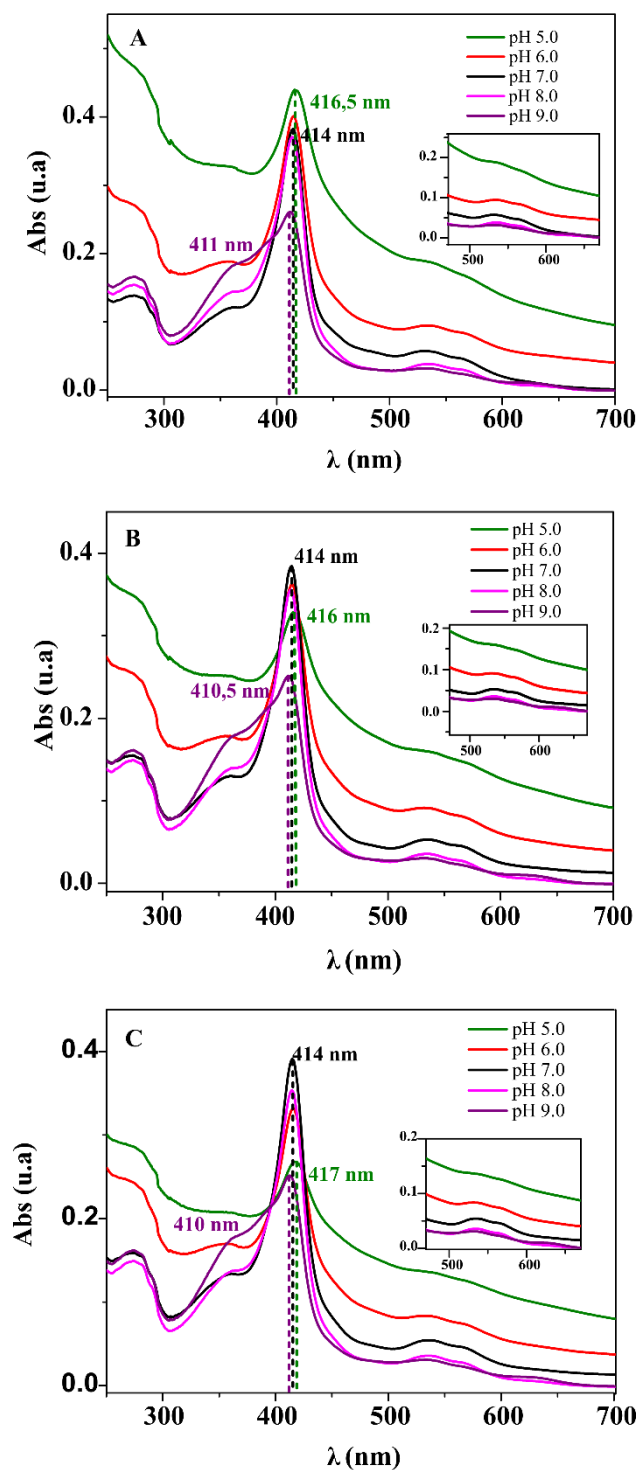


Figura 27: Representação gráfica dos espectros de absorção de luz para o monômero-*d* sob diferentes faixas de pH. (A) uma hora de equilíbrio de reação, (B) duas horas de equilíbrio e (C) três horas de equilíbrio. A absorção foi monitorada na faixa de comprimento de onda entre 250 a 700 nm (eixo x).

Fonte: Elaborado pelo Autora (2022).

O valor de pH igual a 7,0 é estabelecido como referência para armazenamento de proteínas como HbAg (CARVALHO et al., 2011a). Dessa forma, observou-se que a curva referente ao pH 7,0 apresenta um pico longo e estreito centrado em 414 nm seguido de duas bandas Q em 535 e 565 nm; notou-se também que as bandas Q estão alargadas, o qual pode estar associada a auto-oxidação do grupo Heme. Quando as Hbs gigantes são submetidas ao meio alcalino, elas encontram-se com um excesso de cargas superficiais negativas pois as mesmas estão em condições acima do seu pI, sendo o pI do complexo HbAg nativa igual a 6,1 e pI do monômero-*d* igual a 7,1 (OLIVEIRA et al., 2020). Esse excesso de cargas negativas pode levar a uma repulsão entre as subunidades do complexo proteico favorecendo o processo de dissociação e a própria auto-oxidação do grupo Heme (MOREIRA et al., 2010). Dessa forma, foi observado que o processo de dissociação da HbAg para a obtenção do monômero-*d* deixa-o parcialmente oxidado. Alterações no estado oxidativo do monômero *d* foram mais pronunciadas em solução de pH 9,0, onde a banda de Soret é alargada com seu pico máximo deslocado para 410 nm, e as bandas Q permanecem alargadas, ou seja, não apresentam picos bem definidos (MOREIRA et al., 2010; POLI; MOREIRA; IMASATO, 2011; SANTIAGO et al., 2008) (**Figura 27C**). Além disso, foi observada a presença de uma banda em 620 nm, o qual é característica da banda LMCT (Banda de transferência de carga ligante metal) de espécies oxidadas (POLI; MOREIRA; IMASATO, 2011). No pH 5,0 observa-se um ligeiro deslocamento da banda de Soret para 417 nm, porém como os espectros apresentaram uma grande contribuição do espalhamento de Rayleigh, que pode estar associado a uma auto-agregação do monômero-*d* nos valores de pH ácidos, sendo assim não foi possível afirmar qual o estado de oxidação nestas condições.

Diante desses resultados pode se inferir que o monômero-*d* encontra-se mais estável entre os valores de pH de 6,0 a 8,0 e com maior grau de oxidação para o pH mais alcalino igual a 9,0. A grande estabilidade ao estresse oxidativo presente nas Hbs gigantes é devido a estrutura de seu complexo proteico, a maneira como suas cadeias polipeptídicas estão arranjadas na sua forma nativa dificultando o acesso do solvente ao centro da porfirina (POLI; MOREIRA; IMASATO, 2011). Assim, notou-se que para o monômero-*d* de HbGp na faixa de pH entre 5,9 a 7,3 a taxa de autooxidação do grupo Heme foi mais rápida comparada a HbGp nativa (POLI; MOREIRA; IMASATO, 2011). O monômero-*d* puro possui cavidades em torno do Heme que facilitam o acesso das molécula de água ao centro de Fe^{2+} , levando a oxidação e surgimento de espécies de Fe^{3+} (POLI; MOREIRA;

IMASATO, 2011) E pode se inferir que o mesmo fenômeno ocorra para o monômero-*d* de HbAg.

Os gráficos da **Figura 28** mostram os espectros de emissão de fluorescência, enquanto na **Figura 29** temos os gráficos da razão da área/área normalizada (A/A_0) dos dados de fluorescência em função do pH das amostras. Analisando os dados expostos nas **Figura 28** e **Figura 29** observa-se uma tendência entre os dados analisados, conforme se aumenta o valor do pH da subunidade *d* a intensidade de fluorescência também aumenta. Observando as curvas em relação aos diferentes tempos de equilíbrio pode ser notado que, após três horas de equilíbrio, a intensidade de fluorescência entre os valores de pH 5,0 e 6,0 se aproximam (**Figura 28C**), não havendo uma queda pronunciada como se observa para as análises de uma e duas horas após o equilíbrio das amostras (**Figura 28A** e **Figura 28B**).

A intensidade de emissão de fluorescência está intimamente ligada à exposição dos triptofanos ao solvente, assim quando a molécula se encontra em condições de maior exposição desses aminoácidos (molécula aberta), maior será sua intensidade de fluorescência. Para o monômero-*d* observa-se que no valor de pH alcalino igual a 9 (**Figura 28** e **Figura 29**) a intensidade de fluorescência é maior, ou seja, neste pH a molécula encontra-se menos estável que nas outras faixas de pH avaliadas. Estes dados são concordantes com os dados de absorção óptica, onde foi demonstrado que em pH 9,0 a molécula encontra-se em maior estado de oxidação que as demais faixas de pH. A intensidade de fluorescência é menor em valores de pH 5,0 e 6,0 nos estudos realizados. Observou-se também um deslocamento de Stokes de 323 para 331 nm nos picos dos espectros de intensidade de fluorescência dos valores de pH mais ácidos (5,0 e 6,0) ao mais alcalino (pH 9,0). Comportamento semelhante foi encontrado para HbAg nativa, onde foi observado um aumento de 5,5 vezes na fluorescência total do pH 7,0 até o pH 10, além da variação e deslocamento do máximo de emissão dos comprimento de 329 nm (pH 7,0) para 334 nm (pH 10) (OLIVEIRA et al., 2020). Para HbAg nativa essas mudanças observadas nos espectros de emissão de fluorescência foram atribuídas ao processo de dissociação do complexo proteico em pH 10. No caso do monômero-*d* por se tratar de uma subunidade única é possível que no pH mais básico (pH 9,0) a proteína comece um processo de desenovelamento parcial. E os resultados encontrados para o pH mais ácido (5,0) podem estar atribuídos a auto agregação do monômero, o que deixaria os triptofanos menos expostos ao solvente.

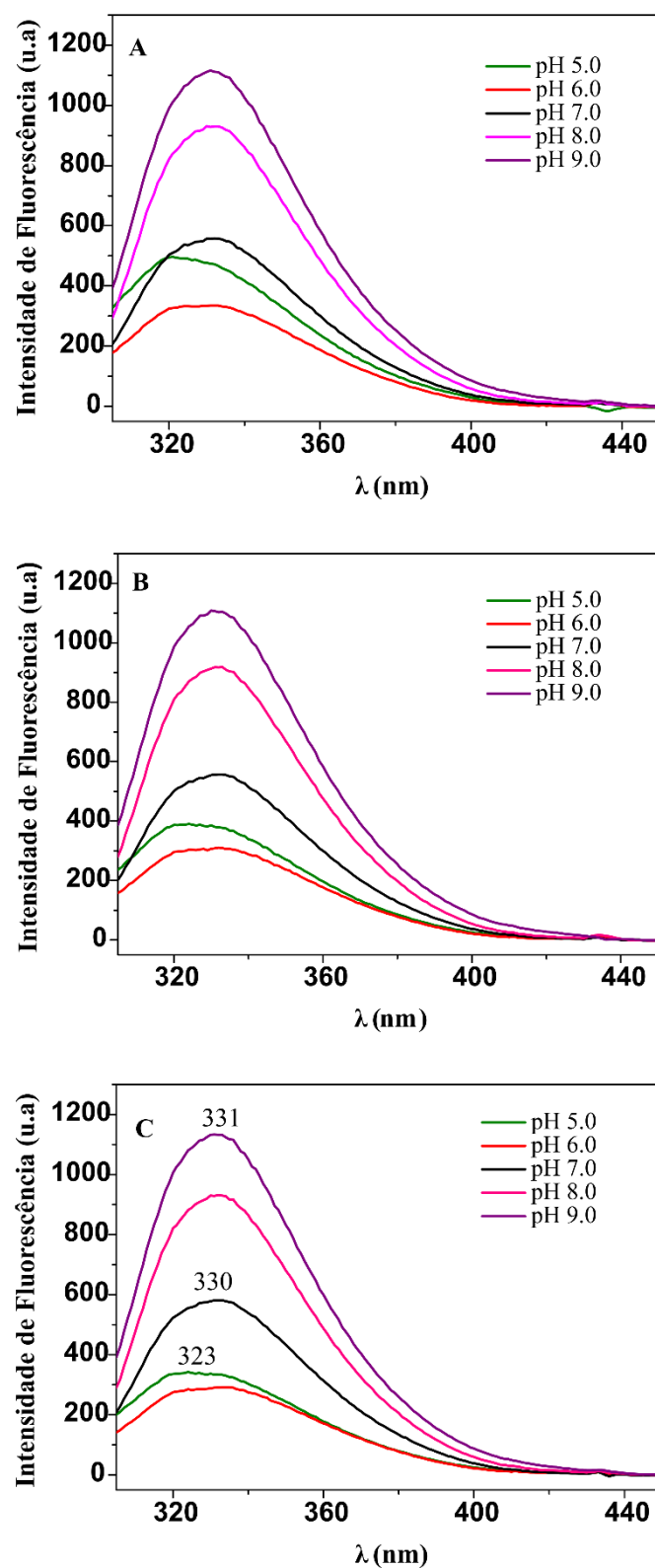


Figura 28: Espectros de emissão de fluorescência do monômero *d* de HbAg em função da variação de pH. (A) uma hora de equilíbrio de reação, (B) duas horas de equilíbrio e (C) três horas de equilíbrio.

Fonte: Elaborado pela Autora (2022).

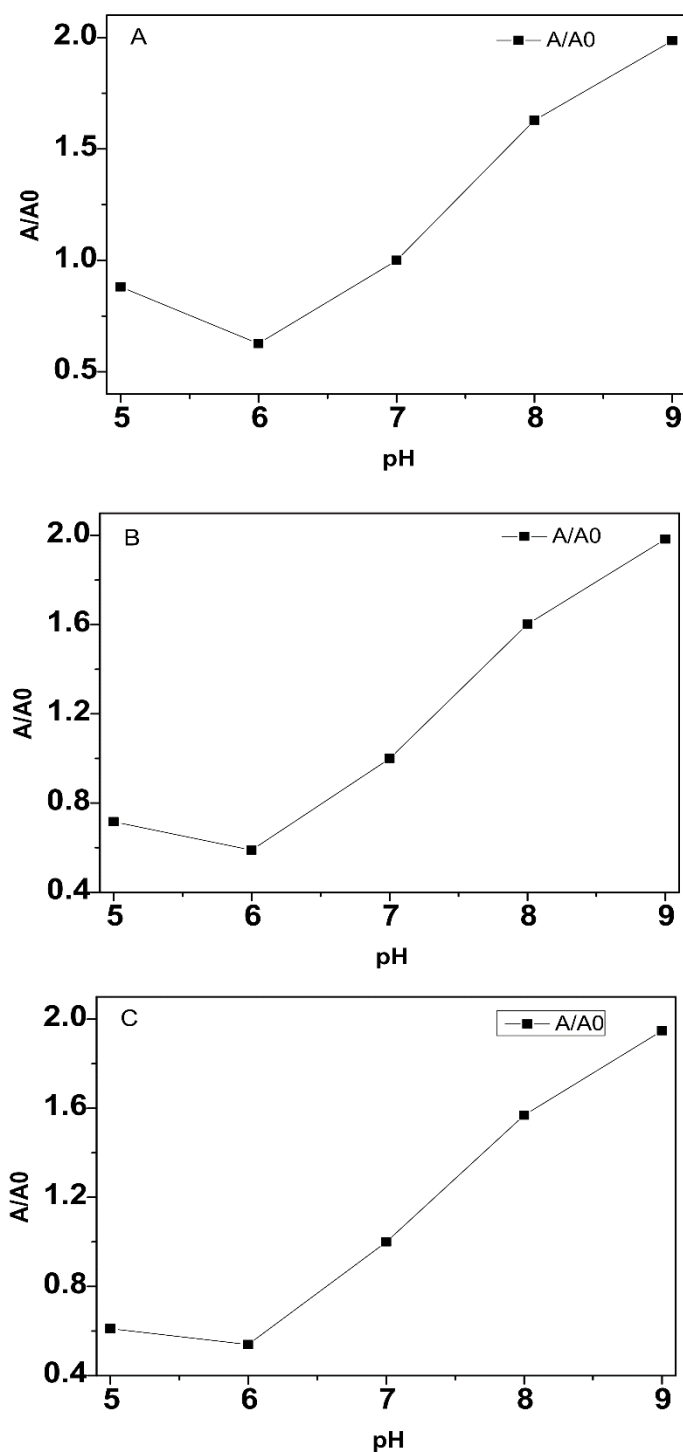


Figura 29: Área total de emissão de fluorescência do monômero-*d* de HbAg normalizada pela emissão do monômero-*d* em pH 7,0. (A) uma hora de equilíbrio de reação, (B) duas horas de equilíbrio e (C) três horas de equilíbrio.

Fonte: Elaborado pela Autora (2022).

A partir da interpretação dos dados gerados pelas análises biofísicas utilizadas, propriedades como o ponto isoelétrico (pI) da proteína pode ser estimado. A técnica de

espalhamento de luz está diretamente ligada ao tamanho da partícula, assim partículas maiores espalham mais luz que partículas menores. Quando uma proteína se encontra com pH próximo ao seu pI, essa proteína possui cargas que se anulam, fazendo com que a proteína se agregue. Provavelmente esse é o motivo dos valores de LSI estarem mais altos para os valores de pH de 5 a 7,0 (**Figura 30**); o monômero-*d* estaria perto do seu pI fazendo com que essa subunidade se aproxime formando agregados maiores que o monômero-*d* íntegro. A partir da análise de gel bidimensional foi possível determinar o pI da subunidade *d* de HbAg e esta encontra-se com pI igual a 7,1 (OLIVEIRA et al., 2020); dessa forma, as estimativas iniciais feitas neste trabalho foram coincidentes com dados experimentais mais robustos para aferição deste parâmetro.

Porém é observado que em pH 9,0 o espalhamento de luz volta a subir (**Figura 30**). Esse aumento pode estar relacionado à formação de dímeros, uma vez que, de acordo com o descrito na literatura o monômero-*d* pode se agregar formando dímeros em valor de pH alcalino (CARVALHO et al., 2011a). Para HbGp a capacidade de agregação do monômero-*d* foi observada pH 7,0 e pH 10. Dados obtidos através de AUC, da HbGp em pH 10, sugerem a existência de um equilíbrio entre as formas de monômero-*d* e de seu dímero-*d*₂, em que a porcentagem de dímeros fica em torno de 10%. A formação dos dímeros, *d*₂, é dependente da concentração de hemoglobina para as amostras contendo o monômero-*d* puro em pH 7,0; após a dissociação, estando o monômero em pH 10, essa dependência da concentração de hemoglobina não foi observada (CARVALHO et al., 2011a). A obtenção da estrutura cristalográfica da subunidade *d*-HbGp veio por complementar estas informações, além de deixar claro que somente é favorecida a formação de dímeros, não sendo possível a formação de trímeros dessa subunidade (BACHEGA et al., 2015). Isto pode explicar o aumento do espalhamento de luz em pH 9,0, já que esta faixa de pH pode estar favorecendo a formação de dímeros de monômero-*d*.

A dinâmica de agregação da HbAg ainda é desconhecida e com o uso destas técnicas biofísicas não é possível afirmar que esses dados de LSI estão relacionados à formação de dímeros de monômero, outras técnicas serão necessárias para avaliarmos esta agregação do monômero-*d* e poderão ser empregadas em trabalhos futuros do grupo de pesquisa. Os dados reportados aqui são dados iniciais para caracterização biofísica do monômero-*d* nativo. Questões envolvendo a dinâmica de agregação da HbAg devem ser respondidas com a utilização de outras abordagens metodológicas.

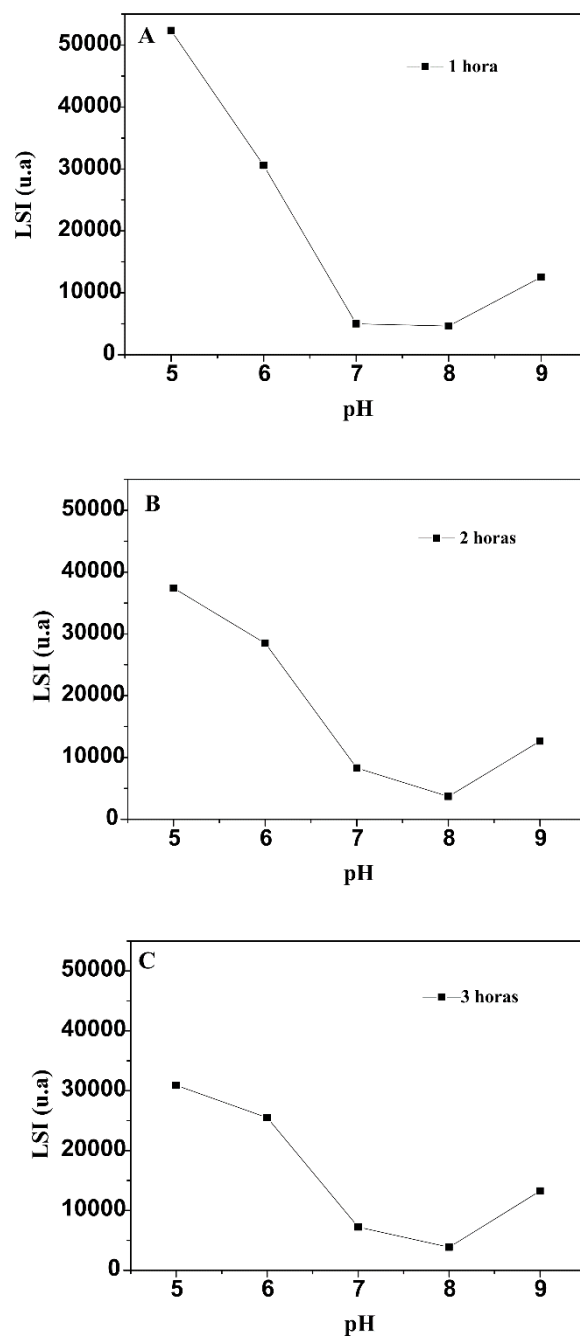


Figura 30: Gráficos relacionando pH e espalhamento de luz *light scattering intensity* monômero-*d*. (A) uma hora de equilíbrio de reação, (B) duas horas de equilíbrio e (C) três horas de equilíbrio.

Fonte: Elaborado pela Autora (2022).

3.3 Interação do monômero-*d* de HbAg com o surfactante aniônico Dodecilsulfato de sódio (SDS) através de análises de Espectroscopia UV-VIS, Fluorescência e Espalhamento de Luz (LSI)

A **Figura 31** apresenta os resultados de absorção ótica referente à titulação do monômero *d* de HbAg com SDS em pH 7,0. Foi observado que o aumento da concentração de detergente induz um aumento na intensidade de absorção ótica até a concentração de 2,0 mmol L⁻¹. A partir da concentração de 1,5 mmol L⁻¹ observa-se um deslocamento do pico da banda de Soret para comprimentos de onda menores (**Figura 31C**), além do seu alargamento. A partir da concentração de 4,0 mmol L⁻¹ de SDS observou-se o desaparecimento das bandas Q. Estas mudanças nos espectros de absorção ótica da interação entre o SDS e o monômero-*d* de HbAg são características da oxidação do grupo Heme (POLI; MOREIRA; IMASATO, 2011). Para a HbAg na sua forma íntegra em pH 5,0 foi observado que nas concentrações entre 0,01 a 0,2 mmol L⁻¹ de SDS ocorreu uma agregação entre a proteína+surfactante e após essa faixa de concentração iniciou-se o processo de oxidação do grupo Heme (RAMOS, 2017). Na faixa de 0,04 a 0,2 mmol L⁻¹ o surfactante promoveu uma redução na intensidade de absorção ótica da HbAg nativa. Fica evidente que a baixa concentração de SDS pode induzir a oxidação da proteína nativa devido à forte interação entre a oxi-HbAg e o surfactante. O pI da oxi-HbAg é igual a 6,11, assim em pH igual a 5,0 a proteína encontra-se com um excesso de cargas positivas superficiais e sua interação com um surfactante de natureza aniônica, como o SDS, é forte. Vale ressaltar que nessas condições ocorre interações de natureza eletrostática e hidrofóbicas entre o complexo proteína+surfactante, ou seja, a proteína com excesso de cargas positivas interage com a cabeça polar negativamente carregada do SDS, enquanto as regiões hidrofóbicas da proteína podem interagir com a cauda do detergente.

Porém, o mesmo não ocorreu nos dados reportados para o monômero, pois as condições experimentais apresentadas foram realizadas em pH 7,0 e o pI do monômero *-d* é igual a 7,1, ou seja, o monômero está muito próximo a região de seu pI, onde as cargas tem a tendência de se anularem, e é por esse motivo que a interação do monômero *d* com SDS em pH 7,0 foi mais fraca e muito provavelmente de natureza hidrofóbica. Assim, foi necessária uma concentração maior de SDS para que a interação entre o monômero *d* e o surfactante ocorresse, quando comparado com os dados da HbAg nativa em pH 5,0. O deslocamento da Banda de Soret iniciou-se em com 1,5 mmol L⁻¹, e o desaparecimento das bandas Q conjuntamente com o aparecimento da banda LCMT em torno de 620 nm se deu a partir de 4,0 mmol L⁻¹.

Em pH 7,0 a interação entre HbAg nativa e o SDS foi mais fraca, a redução na absorção ocorreu em concentrações acima de 0,4 mmol L⁻¹, ou seja 10x maior quando comparado com o pH 5,0 e processo de oxidação se deu a partir da concentração de 0,8

mmol L⁻¹ (RAMOS, 2017). Observou-se também uma possível formação de agregados entre a proteína e o surfactante devido a presença de espalhamento de Rayleigh com aumento da concentração de SDS, notando isso nas concentrações a partir de 2 mmol L⁻¹ (Figura 31).

Para outras Hbs de anelídeos, como HbGp, observou-se que o processo de oxidação se inicia a partir da concentração de 0,6 mmol L⁻¹ de SDS (SANTIAGO et al., 2007). Ou seja, a HbAg necessitou de uma menor concentração de SDS para iniciar o seu processo de oxidação, mostrando ser menos estável que a HbGp (SANTIAGO et al., 2007).

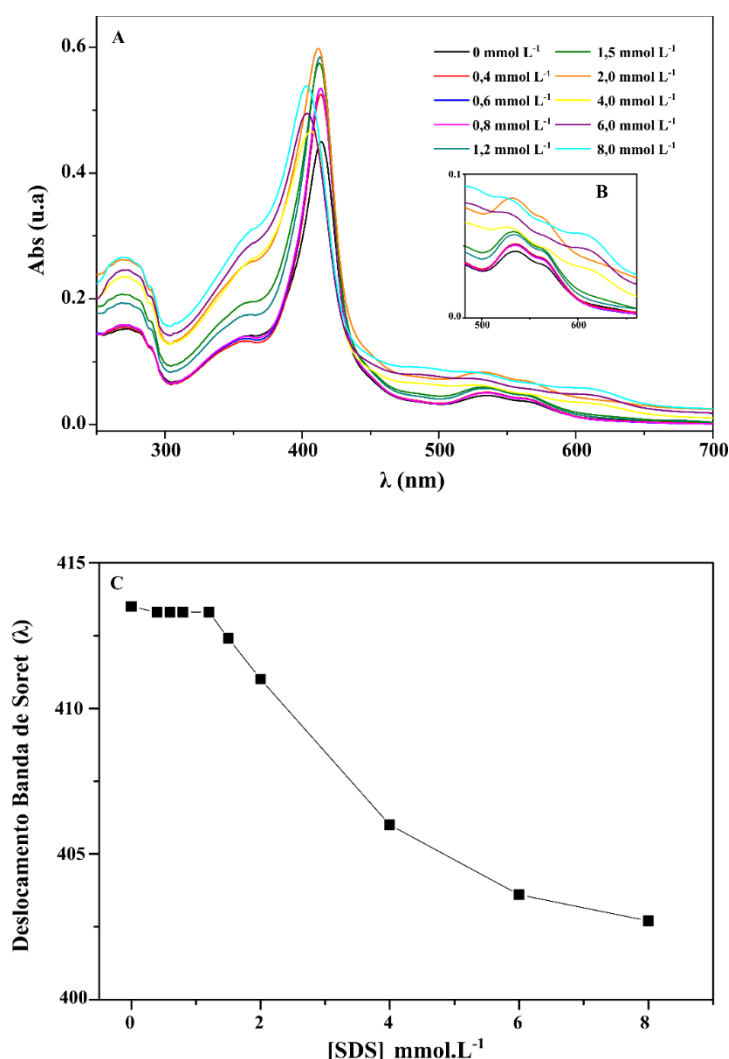


Figura 31: Interação entre o monômero-*d* e o surfactante SDS monitorada por absorção ótica. (A) Espectros de absorção de luz UV-vis do monômero-*d* de HbAg em função de diferentes concentrações do surfactante aniônico SDS. (B) Inserto com a amplificação da região das bandas Q. (C) Gráfico representativo do deslocamento da banda de Soret monitorada na região do λ de 415 nm x concentração de SDS.

Fonte: Elaborado pela Autora (2022).

Através da análise de fluorescência observou-se inicialmente que o incremento de SDS na solução ocasionou uma diminuição na intensidade de fluorescência e a partir da concentração de $1,2 \text{ mmol L}^{-1}$ os valores de intensidade são maiores que os encontrados para a amostra na ausência de SDS. A partir de $1,2 \text{ mmol L}^{-1}$ o aumento das concentrações de SDS induz um aumento gradativo da intensidade de fluorescência (**Figura 32**). Esse mesmo comportamento foi observado para HbAg nativa em pH 5,0, onde a diminuição da intensidade de fluorescência se deu até a concentração de $0,2 \text{ mmol L}^{-1}$ de SDS (RAMOS, 2017). O surfactante SDS é capaz de promover o desenovelamento de proteínas desnaturando suas estruturas secundárias (HØJGAARD et al., 2018; KASPERSEN et al., 2017; OTZEN, 2011). Os surfactantes aniônicos interagem com as proteínas através da interação de seu grupo polar aos resíduos de aminoácidos positivamente carregados, como a lisina, arginina e histidina ou através da interação da sua cauda apolar com os sítios hidrofóbicos das proteínas. Quando os sítios começam a ficar saturados com o aumento da concentração do surfactante, inicia-se a formação de *clusters* proteína+detergente, e assim pode ocorrer o desenovelamento das proteínas seguida da sua desnaturação (OTZEN, 2011).

As maiores intensidades de fluorescência indicam que possivelmente há um desenovelamento da molécula, porém não pode-se afirmar que houve a desnaturação ou em qual concentração essa desnaturação pode ocorrer. Análises que monitoram a conformação das estruturas secundárias de proteínas, como o dicróismo circular (CD), se fazem necessárias para responder essa questão. Para a HbAg na sua forma íntegra, concentrações acima de 8 mmol L^{-1} levam a formação de 100% de espécies parcialmente desnaturadas (RAMOS, 2017).

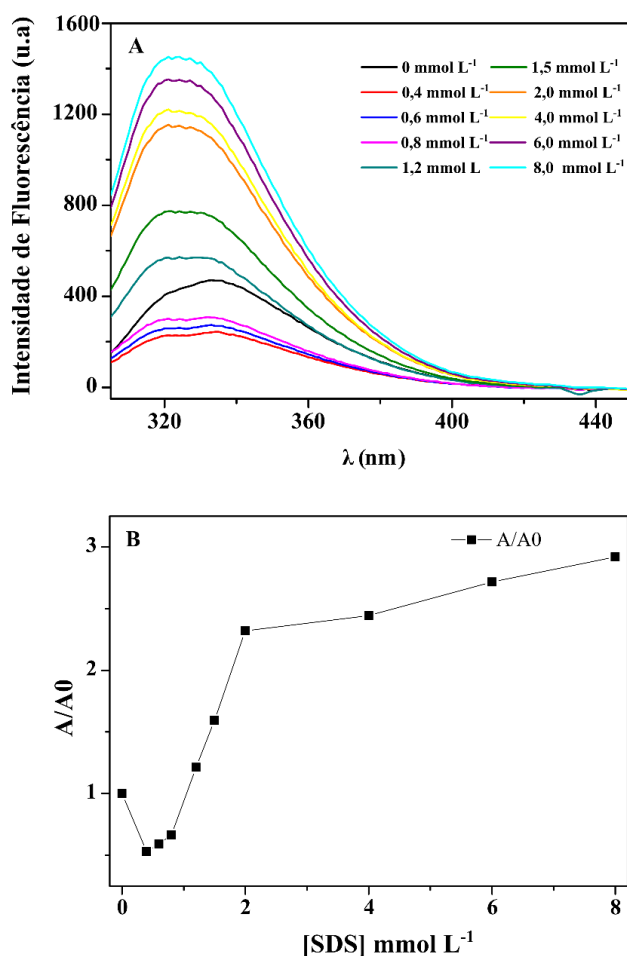


Figura 32: Interação entre o monômero-*d* e o surfactante SDS monitorada por fluorescência. Gráficos relacionando comprimento de onda (λ) em nanômetros (nm) (eixo x) e a intensidade de fluorescência (eixo y) emitida pelo monômero-*d* sob diferentes concentrações de SDS em mmol L⁻¹. (B) Gráfico da área normalizada dos espectros de fluorescência emitidos pelo monômero-*d* de HbAg em função da concentração de SDS.

Fonte: Elaborado pela Autora (2022).

A oscilação entre os pontos monitorados para na análise de LSI mostra que o espalhamento de luz tem um decréscimo até a concentração de 0,8 mmol L⁻¹ e depois cresce até a concentração de 2,0 mmol L⁻¹, seguida de diminuição em 4,0 mmol L⁻¹, tornando-se mais constante a partir deste ponto (**Figura 33**). Pode-se inferir que a medida que são formados os agregados entre a subunidade *d* e o SDS ocorre uma diminuição no espalhamento de luz, chegando a mínima LSI na concentração de 1,2 mmol L⁻¹. Em uma primeira hipótese esta diminuição pode estar associada a formação de agregados proteína-surfactante, que ao se formarem se sedimentam, diminuindo a LSI. Em um segundo momento temos como hipótese a possibilidade de termos presente nesta solução auto agregados de monômero-*d*, que na presença de baixa concentração de SDS, dissocia-se.

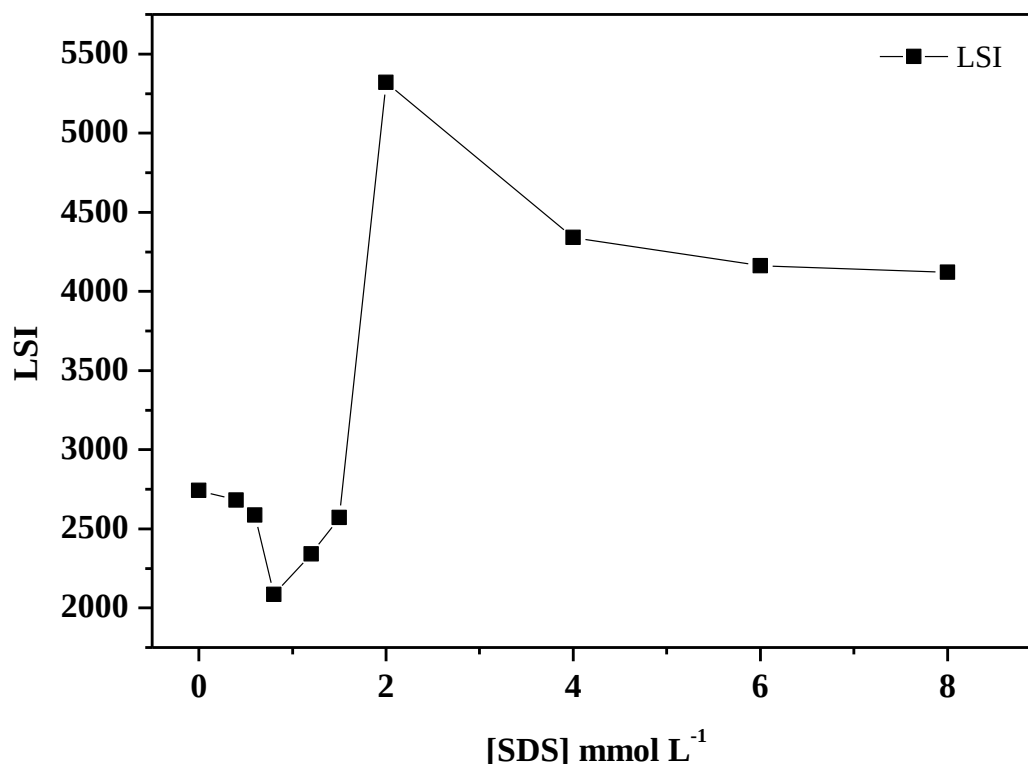


Figura 33: Curva das medidas de espalhamento de luz (LSI) para a interação entre o monômero-*d* de HbAg com o surfactante aniônico SDS. O eixo x representa as diferentes concentrações de SDS utilizadas e no eixo y encontram os valores de intensidade de espalhamento de luz (LSI) em pH 7,0.

Fonte: Elaborado pela Autora (2022)

Em suma, observou-se que o SDS leva a oxidação do grupo heme presente nas Hbs, seguida de um início de desenovelamento da proteína que possivelmente levou a desnaturação dessas moléculas, além da formação de agregados entre o monômero-*d* de HbAg e o surfactante.

4 CONCLUSÕES

Os dados reportados aqui são dados iniciais de caracterização biofísica para o monômero *d* de HbAg. Para os dados de absorção ótica da interação do monômero-*d* em função da variação de pH revelaram que em pH alcalino igual a 9 ocorreu o processo de auto-oxidação do grupo Heme, caracterizado pelo alargamento e deslocamento do pico da banda de Soret para 417 nm, além do desaparecimento das bandas Q e aparecimento da banda LMCT em 620 nm. Analisando as medidas de fluorescência notou-se inicialmente uma diminuição na intensidade de fluorescência para faixa de pH mais ácido (5,0 e 6,0) o que pode estar relacionada com a supressão de fluorescência exercida pelo grupo Heme. Já

para os valores de pH acima de 9,0 ocorreu um aumento na intensidade de fluorescência dados que corroboram com os encontrados para absorção ótica, nessa faixa de pH, onde o grupo Heme sofre auto-oxidação e temos uma maior acessibilidade do solvente os resíduos de triptofano do monômero *d*. Para os dados de LSI observou uma possível agregação do monômero *d* nos valores de pH ácido (5,0) e básico (9,0). Para a interação entre o monômero-*d* e o surfactante SDS os dados revelaram uma interação fraca para o pH 7,0, já que neste pH o monômero *d* encontra-se praticamente na zona do seu pI. Dessa forma comparado com dados da proteína nativa encontrou-se valores mais altos de concentração de SDS para causar mudanças nos espectros de absorção ótica, onde processo de auto-oxidação do grupo Heme foi notado mais claramente a partir das concentrações de 4,0 mmol L⁻¹ de SDS. A partir da concentração de 1,2 mmol L⁻¹ de SDS ocorreu um aumento gradativo da intensidade de fluorescência, demonstrando uma maior acessibilidade do solvente aos resíduos de triptofano o que pode estar relacionado com um início de desenovelamento da molécula. Os dados de LSI levantam a hipótese de uma possível agregação entre o monômero-*d* e o surfactante assim como a presença de auto agregados de monômero em solução.

CONCLUSÃO GERAL E PERSPECTIVAS FUTURAS

Os dados apresentados aqui neste trabalho poderão contribuir na investigação do uso biotecnológico de proteínas identificadas no transcriptoma de *A. gracilis*. Após anotação do transcriptoma dessa espécie de minhoca foram identificadas algumas sequências de proteínas com potencial biotecnológico, como por exemplo as proteínas Lumbrokinases. As enzimas lumbrokinases, encontradas exclusivamente em espécies de anelídeos, pertencem à família de proteínas quinases e as LKs possuem ação comprovada no tratamento de doenças cardiovasculares. Além disso, pesquisas mais recentes apontam uma possível utilização dessas proteínas como marcadores de contaminação ambiental. Estudos mais aprofundados de caracterização e validação dessas proteínas como biomarcadores devem ser realizados. Além disso, o perfil de sequências encontradas neste trabalho traz uma noção do conteúdo genético de *A. gracilis*, podendo assim auxiliar no melhor entendimento da biologia da espécie ou espécies de anelídeos correlacionadas. Estudos envolvendo a evolução dos pigmentos respiratórios apontam a ocorrência de genes de Hr em todos os anelídeos, e esta informação foi corroborada com a identificação de genes de Hr no transcriptoma de *A. gracilis*. Interessantemente encontrou-se a expressão concomitante de genes de Hb e Hr, o motivo pelo qual isso acontece ainda não foi elucidado, mas aponta-se que um dos pigmentos possa desempenhar outras funções que não sejam exclusivamente a do transporte de O₂ para células. Neste trabalho obtivemos avanços na caracterização de HbAg com os estudos iniciais de caracterização biofísica do monômero-*d* de HbAg, além da identificação das sequências polipeptídicas que compõem esse complexo proteico. Analisando o modelo teórico comparativo da unidade biológica de HbAg, notou que uma grande proximidade na conservação da estrutura de HbAg quando comparada com suas proteínas homologas HbGp e HbLt. Dessa forma, acredita-se que os sítios de ligação aos metais se mantiveram conservados. Um dos objetivos do nosso grupo de pesquisa é a investigação do uso de HbAg como um biomarcador de poluição por metais pesados e os dados aqui apresentados servirão de base para trabalhos futuros objetivando elucidar e monitorar como se dá a dinâmica entre a HbAg e os metais pesados. Entre outras perspectivas de trabalhos futuros gostaríamos de destacar: o planejamento de um estudo mais aprofundado do conteúdo genético de *A. gracilis* afim de identificar possíveis biomarcadores e/ou proteínas com potencial biotecnológico e aplicações experimentais das proteínas encontradas. Quanto a caracterização biofísica de HbAg, especificamente do monômero-*d*, destacamos a necessidade da utilização de outras técnicas biofísicas como

ultracentrifugação analítica (AUC) das frações coletadas durante o cromatograma de dissociação, para identificar a contribuição de diferentes conformações da proteína (presença de dímeros, trímeros, tetrâmeros ou dodecâmeros), análises de DLS para melhor compreender a dinâmica de agregação das cadeias de HbAg em função da variação de pH e concentrações do surfactante SDS. Realização de experimentos utilizando surfactantes catiônicos e zwitteriônicos, assim como a realização de análises de dicroísmo circular para o monitoramento da perda de estruturas secundárias da proteína. É importante destacar aqui que obtivemos a construção de um clone do monômero-*d* o que possibilita a produção dessa cadeia a partir de expressão heteróloga. Assim, poderão ser obtidas quantidades suficientes de monômero para a realização dos experimentos aqui sugeridos, já que a baixa concentração da cadeia *d* obtida a partir da dissociação da HbAg foi um dos entraves para realização da sua caracterização biofísica.

REFERÊNCIAS

- ABBEHAUSEN, C. Zinc finger domains as therapeutic targets for metal-based compounds-an update. **Metallomics**, v. 11, n. 1, p. 15–28, 2019.
- ALVAREZ-CARREÑO, C.; BECERRA, A.; LAZCANO, A. Molecular evolution of the oxygen-binding hemerythrin domain. **PLOS ONE**, v. 11, n. 6, p. e0157904, 23 jun. 2016.
- BACHEGA, J. F. R. **Estrutura cristalográfica da hemoglobina gigante de *Glossoscolex paulistus*, um complexo de 3,6 mega Daltons**. 2013. Tese (Doutorado em Ciências) Universidade de São Paulo, Instituto de Física de São Carlos, São Carlos, 2013.
- BACHEGA, J. F. R. *et al.* The structure of the giant haemoglobin from *Glossoscolex paulistus*. **Acta Crystallographica Section D Biological Crystallography**, v. 71, n. 6, p. 1257–1271, 1 jun. 2015.
- BAILLY, X. *et al.* Evolution of the sulfide-binding function within the globin multigenic family of the deep-sea hydrothermal vent tubeworm *Riftia pachyptila*. **Molecular Biology and Evolution**, v. 19, n. 9, p. 1421–1433, 2002.
- BECANA, M. *et al.* Research review Hemoglobins in the legume-Rhizobium symbiosis. **New Phytologist**, v. 228, p. 472–484, 2020.
- BELATO, F. A. *et al.* Newly discovered occurrences and gene tree of the extracellular globins and linker chains from the giant hexagonal bilayer hemoglobin in metazoans. **Genome Biology and Evolution**, v. 11, n. 3, p. 597–612, 2019.
- BLAKE, J. A.; GRASSLE, J. P.; ECKELBARGER, K. J. Capitella sp . I , with a review of the literature for confirmed records. **Zoosymposia**, v. 53, n. September 1969, p. 25–53, 2009.
- BOLGER, A. M.; LOHSE, M.; USADEL, B. Trimmomatic: a flexible trimmer for Illumina sequence data. **Bioinformatics**, v. 30, n. 15, p. 2114–2120, 1 ago. 2014.
- BOLOGNESI, M. *et al.* Nonvertebrate hemoglobins: structural bases for reactivity. **Progress in biophysics and molecular biology**, v. 68, n. 1, p. 29–68, 1997.
- BRACKE, A.; HOOGEWIJS, D.; DEWILDE, S. Exploring three different expression systems for recombinant expression of globins: *Escherichia coli* , *Pichia pastoris* and *Spodoptera frugiperda*. **Analytical Biochemistry**, v. 543, n. November 2017, p. 62–70, fev. 2018.
- CALISI, A. *et al.* Effect of heavy metal exposure on blood haemoglobin concentration and methemoglobin percentage in *Lumbricus terrestris*. **Ecotoxicology**, v. 20, n. 4, p. 847–854, 22 jun. 2011.
- CALISI, A. *et al.* Integrated biomarker analysis in the earthworm *Lumbricus terrestris*: Application to the monitoring of soil heavy metal pollution. **Chemosphere**, v. 90, n. 11, p. 2637–2644, 2013.

CAMACHO, C. *et al.* BLAST+: Architecture and applications. **BMC Bioinformatics**, v. 10, p. 1–9, 2009.

CARUSO, C. S. *et al.* Metals content of *Glossoscolex paulistus* extracellular hemoglobin: Its peroxidase activity and the importance of these ions in the protein stability. **Journal of Inorganic Biochemistry**, v. 161, p. 63–72, ago. 2016.

CARVALHO, F. A. O. *et al.* Further characterization of the subunits of the giant extracellular hemoglobin of *Glossoscolex paulistus* (HbGp) by SDS-PAGE electrophoresis and MALDI-TOF-MS. **Process Biochemistry**, v. 46, n. 11, p. 2144–2151, 2011a.

CARVALHO, F. A. O. *et al.* Molecular masses and sedimentation coefficients of extracellular hemoglobin of *Glossoscolex paulistus*: Alkaline oligomeric dissociation. **International Journal of Biological Macromolecules**, v. 48, n. 1, p. 183–193, jan. 2011b.

CARVALHO, F. A. O.; SANTIAGO, P. S.; TABAK, M. On the stability of the extracellular hemoglobin of *Glossoscolex paulistus*, in two iron oxidation states, in the presence of urea. **Archives of Biochemistry and Biophysics**, v. 519, n. 1, p. 46–58, 1 mar. 2012.

CARVALHO, J. W. P. *et al.* Thermal denaturation and aggregation of hemoglobin of *Glossoscolex paulistus* in acid and neutral media. **International Journal of Biological Macromolecules**, v. 54, p. 109–118, mar. 2013.

CHABASSE, C. *et al.* The multigenic family of the extracellular hemoglobin from the annelid polychaete *Arenicola marina*. **Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology**, v. 144, n. 3, p. 319–325, jul. 2006a.

CHABASSE, C. *et al.* Gene structure and molecular phylogeny of the linker chains from the giant annelid hexagonal bilayer hemoglobins. **Journal of Molecular Evolution**, v. 63, n. 3, p. 365–374, 12 set. 2006b.

CHEN, W. T. *et al.* Structural basis for cooperative oxygen binding and bracelet-assisted assembly of *Lumbricus terrestris* hemoglobin. **Scientific Reports**, v. 5, n. 1, p. 9494, 21 ago. 2015.

COATES, C. J.; DECKER, H. Immunological properties of oxygen-transport proteins: hemoglobin, hemocyanin and hemerythrin. **Cellular and Molecular Life Sciences**, v. 74, n. 2, p. 293–317, 12 ago. 2016.

CONESA, A. *et al.* Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. **Bioinformatics**, v. 21, n. 18, p. 3674–3676, 15 set. 2005.

COSTA-PAIVA, E. M. *et al.* Discovery and evolution of novel hemerythrin genes in annelid worms. **BMC Evolutionary Biology**, v. 17, n. 1, p. 85, 23 dez. 2017.

COSTA-PAIVA, E. M.; COATES, C. J. Recent insights into the diversity and evolution of invertebrate hemerythrins and extracellular globins. In: Hoeger, U.; Harris, J. R. **Vertebrate and Invertebrate Respiratory Proteins, Lipoproteins and Other Body Fluid Proteins**. 1. ed. Springer Cham, 2020, Chap. 10, p. 251–273.

CREMONESI, M. V. *et al.* Earthworm species in Musa spp. plantations in Brazil and worldwide Launched to accelerate biodiversity research. **ZooKeys**, v. 1033, p. 1–33, 2021.

CUNHA, L. *et al.* Morphometry of the epidermis of an invasive megascolecoid earthworm (*Amyntas gracilis*, Kinberg 1867) inhabiting actively volcanic soils in the Azores archipelago. **Ecotoxicology and Environmental Safety**, v. 74, n. 1, p. 25–32, 2011.

DARRIBA, D. *et al.* ProtTest 3: fast selection of best-fit models of protein evolution. **Bioinformatics (Oxford, England)**, v. 27, n. 8, p. 1164–1165, abr. 2011.

DE SILVA, P. M. C. S. *et al.* Chlorpyrifos causes decreased organic matter decomposition by suppressing earthworm and termite communities in tropical soil. **Environmental Pollution**, v. 158, n. 10, p. 3041–3047, 2010.

DEMETRIO, W. C. *et al.* Earthworms in Brazilian no-tillage agriculture: Current status and future challenges. **European Journal of Soil Science**, n. June, p. 1–18, 14 dez. 2019.

DUAN, J. (ELLIE) *et al.* Dosage Compensation and Gene Expression of the X Chromosome in Sheep. **G3 Genes|Genomes|Genetics**, v. 9, n. 1, p. 305–314, 1 jan. 2019.
EL-GEHALI, S. *et al.* The Pfam protein families database in 2019. **Nucleic Acids Research**, v. 47, n. D1, p. D427–D432, 8 jan. 2019.

FERMI, G. *et al.* The crystal structure of human deoxyhaemoglobin at 1.74 Å resolution. **Journal of molecular biology**, v. 175, n. 2, p. 159–174, 15 maio 1984.

FOTOUHI, L. *et al.* Application of merged spectroscopic data combined with chemometric analysis for resolution of hemoglobin intermediates during chemical unfolding. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 136, n. PC, p. 1974–1981, 5 fev. 2015.

GASTEIGER, E. *et al.* Protein identification and analysis tools on the ExPASy server. In: WALKER, J. M. (ed.). **The Proteomics Protocols Handbook**. Totowa, NJ: Humana Press, 2005. v. 23, chap. 53, p. 571–607.

GOLEB, J. A. Near ultraviolet-visible atomic absorption spectra of the noble gases. **Analytical Chemistry**, v. 38, n. 8, p. 1059–1061, 1 jul. 1966.

GOTOH, T.; SUZUKI, T.; TAKAGI, T. Structure of the extracellular hemoglobin of *Tylorhynchus heterochaetus*. In: VINOGRADOV, S. N.; KAPP, O. H. (ed.). **Structure and Function of Invertebrate Oxygen Carriers**. New York, NY: Springer New York, 1991. chap. 28, p. 217–226.

GREEN ETXABE, A. *et al.* Identifying conserved polychaete molecular markers of metal exposure: Comparative analyses using the *Alitta virens* (Annelida, Lophotrochozoa) transcriptome. **Comparative Biochemistry and Physiology Part - C: Toxicology and Pharmacology**, v. 240, p. 108913, 1 fev. 2021.

HACKERT, M. L.; RIGGS, A. F. When size matters. **Structure**, v. 14, n. 7, p. 1094–1096, 1 jul. 2006.

HANKS, S. K.; HUNTER, T. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification 1 . **The FASEB Journal**, v. 9, n. 8, p. 576–596, 1995.

HARNOIS, T. *et al.* High-level production of recombinant *Arenicola marina* globin chains in *Escherichia coli*: A new generation of blood substitute. **Artificial Cells, Blood Substitutes, and Biotechnology**, v. 37, n. 3, p. 106–116, 13 jan. 2009.

HEATHER, J. M.; CHAIN, B. The sequence of sequencers: The history of sequencing DNA. **Genomics**, v. 107, n. 1, p. 1–8, 1 jan. 2016.

HILL, D. P. *et al.* Gene Ontology annotations: what they mean and where they come from. **BMC Bioinformatics**, v. 9, n. Suppl 5, p. S2, 2008.

HØJGAARD, C. *et al.* Can a charged surfactant unfold an uncharged protein? **Biophysical Journal**, v. 115, n. 11, p. 2081–2086, dez. 2018.

HUANG, Y. *et al.* CD-HIT Suite: a web server for clustering and comparing biological sequences. **Bioinformatics Applications Note**, v. 26, n. 5, p. 680–682, 2010.

ILLUMINA INC. **Illumina® MiSeq System Guide**. San Diego, California: Illumina, 2018.

JABŁOŃSKI, A. Über den Mechanismus der Photolumineszenz von Farbstoffphosphoren. **Zeitschrift für Physik** 1935 **94:1**, v. 94, n. 1, p. 38–46, jan. 1935.

JONES, P. *et al.* InterProScan 5: genome-scale protein function classification. **Bioinformatics**, v. 30, n. 9, p. 1236–1240, 1 maio 2014.

KASPERSEN, J. D. *et al.* Refolding of SDS-unfolded proteins by nonionic surfactants. **Biophysical Journal**, v. 112, n. 8, p. 1609–1620, abr. 2017.

KURTZ, D. M. Oxygen-carrying proteins: Three solutions to a common problem. **Essays in Biochemistry**, v. 34, p. 85–100, 1 nov. 1999.

LAKOWICZ, J. R. **Principles of Fluorescence Spectroscopy**. 3ed. ed. Boston, MA: Springer US, 2006.

LE MEUR, Y. *et al.* First-in-human use of a marine oxygen carrier (M101) for organ preservation: A safety and proof-of-principle study. **American Journal of Transplantation**, v. 20, n. 6, p. 1729–1738, 1 jun. 2020.

LE PAPE, F. *et al.* Advancement in recombinant protein production using a marine oxygen carrier to enhance oxygen transfer in a CHO-S cell line. **Artificial Cells, Nanomedicine, and Biotechnology**, v. 43, n. 3, p. 186–195, 4 maio 2015.

LETUNIC, I.; KHEDKAR, S.; BORK, P. SMART: recent updates, new developments and status in 2020. **Nucleic Acids Research**, v. 49, n. D1, p. D458–D460, 8 jan. 2021.

LEVY, S. E.; BOONE, B. E. Next-Generation Sequencing Strategies. **Cold Spring Harbor perspectives in medicine**, v. 9, n. 7, 1 jul. 2019.

LIU, L. *et al.* Comparison of next-generation sequencing systems. **Journal of Biomedicine and Biotechnology**, v. 2, 2012.

LIU, T. *et al.* Transcriptome, bioaccumulation and toxicity analyses of earthworms (*Eisenia fetida*) affected by trifloxystrobin and trifloxystrobin acid. **Environmental Pollution**, v. 265, p. 115100, 1 out. 2020.

MANWELL, C. Comparative Physiology: Blood Pigments. **Annual Review of Physiology**, v. 22, n. 1, p. 191–244, mar. 1960.

MARGULIES, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. **Nature** **2005 437:7057**, v. 437, n. 7057, p. 376–380, 31 jul. 2005.

MARTIN, P. D. *et al.* Mass spectrometric composition and molecular mass of *Lumbricus terrestris* hemoglobin: A refined model of its quaternary structure. **Journal of Molecular Biology**, v. 255, n. 1, p. 154–169, jan. 1996.

MEHTA, K. D. *et al.* The low density lipoprotein receptor in *Xenopus laevis*. I. Five domains that resemble the human receptor. **Journal of Biological Chemistry**, v. 266, n. 16, p. 10406–10414, jun. 1991.

MIKAMI, Y. *et al.* Whole transcriptome analysis using next-generation sequencing of sterile-cultured *Eisenia andrei* for immune system research. **PLoS ONE**, v. 10, n. 2, p. 1–13, 2015.

MOREIRA, L. M. *et al.* Ligand changes in ferric species of the giant extracellular hemoglobin of *Glossoscolex paulistus* as function of pH: correlations between redox, spectroscopic and oligomeric properties and general implications with different hemoproteins. **Journal of Porphyrins and Phthalocyanines**, v. 14, n. 03, p. 199–218, 25 mar. 2010.

NEAVE, M. J. *et al.* The transcriptome and proteome are altered in marine polychaetes (Annelida) exposed to elevated metal levels. **Journal of Proteomics**, v. 75, n. 9, p. 2721–2735, 2012.

NEET, K. E.; LEE, J. C. Biophysical characterization of proteins in the post-genomic era of proteomics. **Molecular & cellular proteomics : MCP**, v. 1, n. 6, p. 415–420, 2002.

NEGRISOLO, E. *et al.* The evolution of extracellular hemoglobins of annelids, vestimentiferans, and pogonophorans. **Journal of Biological Chemistry**, v. 276, n. 28, p. 26391–26397, 13 jul. 2001.

NELSON, D. L.; COX, M. M. Função Proteica. In: **Princípios de bioquímica de Lehninger**. 6. ed. Porto Alegre: Artmed, 2014. cap. 5, p. 157–182.

NOLAN, T.; HANDS, R. E.; BUSTIN, S. A. Quantification of mRNA using real-time RT-PCR. **Nature Protocols**, v. 1, n. 3, p. 1559–1582, 9 ago. 2006.

OLIVEIRA, J. B. S. *et al.* Initial biophysical characterization of *Amyntas gracilis* giant extracellular hemoglobin (HbAg). **European Biophysics Journal**, v. 49, n. 6, p. 473–484,

19 set. 2020.

ORDOÑEZ-ARÉVALO, B. *et al.* Enzymatic dynamics into the *Eisenia fetida* (Savigny, 1826) gut during vermicomposting of coffee husk and market waste in a tropical environment. **Environmental Science and Pollution Research**, v. 25, n. 2, p. 1576–1586, 2 jan. 2018.

OTZEN, D. Protein-surfactant interactions: A tale of many states. **Biochimica et Biophysica Acta - Proteins and Proteomics**, v. 1814, n. 5, p. 562–591, 2011.

OWNBY, D. W. *et al.* The extracellular hemoglobin of the earthworm, *Lumbricus terrestris*. Determination of subunit stoichiometry. **Journal of Biological Chemistry**, v. 268, n. 18, p. 13539–13547, 1993.

PAIVA, D. L. *et al.* Espectroscopia no ultravioleta. In: **Introdução à Espectroscopia**. 4. ed. [s.l.] Cengage Learning, 2010. p. 1773–1775.

PANDURANGAN, A. P. *et al.* The SUPERFAMILY 2.0 database: a significant proteome update and a new webserver. **Nucleic Acids Research**, v. 47, n. D1, p. D490–D494, 8 jan. 2019.

PARELHO, C. *et al.* Biological endpoints in earthworms (*Amyntas gracilis*) as tools for the ecotoxicity assessment of soils from livestock production systems. **Ecological Indicators**, v. 95, n. January, p. 984–990, dez. 2018.

PAUL, S. *et al.* Transcriptome sequencing, de novo assembly and annotation of the freeze tolerant earthworm, *Dendrobaena octaedra*. **Gene Reports**, v. 13, p. 180–191, 1 dez. 2018.

PERUTZ, M. F. Stereochemistry of Cooperative effects in Haemoglobin. **Nature**, v. 228, p. 726–734, 1970.

PERUTZ, M. F. *et al.* The stereochemical mechanism of the cooperative effects in hemoglobin revisited. **Annual Review of Biophysics and Biomolecular Structure**, v. 27, n. 1, p. 1–34, jun. 1998.

PETERSEN, T. N. *et al.* SignalP 4.0: discriminating signal peptides from transmembrane regions. **Nature Methods**, v. 8, n. 10, p. 785–786, 29 out. 2011.

POLI, A. L.; MOREIRA, L. M.; IMASATO, H. Autoxidation of giant extracellular hemoglobin of *Glossoscolex paulistus*: Molecular mechanism and oligomeric implications. **Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy**, v. 82, n. 1, p. 306–315, 2011.

PONESAKKI, V. *et al.* Annotation of nerve cord transcriptome in earthworm *Eisenia fetida*. **Genomics Data**, v. 14, p. 91–105, 1 dez. 2017.

RAMOS, L. **Estudos biofísicos da Hemoproteína extracelular de *Amyntas gracilis* (HbAg) na ausência e na presença de surfactantes**. Dissertação (Mestrado em Biotecnologia). Universidade Estadual Paulista, Instituto de Química, Araraquara, 2017.

RIFKIND, J. M. *et al.* Detection, formation, and relevance of hemochromes and hemochromes. In: **Methods in enzymology**. [s.l.] Methods Enzymol, 1994. v. 231p. 449–480.

RONDA, L.; BRUNO, S.; BETTATI, S. Tertiary and quaternary effects in the allosteric regulation of animal hemoglobins. **Biochimica et Biophysica Acta - Proteins and Proteomics**, v. 1834, n. 9, p. 1860–1872, 2013.

RONQUIST, F.; HUELSENBECK, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. **Bioinformatics**, v. 19, n. 12, p. 1572–1574, 12 ago. 2003.

ROUSSELOT, M. *et al.* *Arenicola marina* extracellular hemoglobin: a new promising blood substitute. **Biotechnology Journal**, v. 1, n. 3, p. 333–345, mar. 2006.

ROYER, W. E. *et al.* Lumbricus erythrocrurin at 3.5 Å resolution: Architecture of a megadalton respiratory complex. **Structure**, v. 14, n. 7, p. 1167–1177, jul. 2006.

ROYER, W. E.; OMARTIAN, M. N.; KNAPP, J. E. Low resolution crystal structure of arenicola erythrocrurin: Influence of coiled coils on the architecture of a megadalton respiratory protein. **Journal of Molecular Biology**, v. 365, n. 1, p. 226–236, 2007.

RUPAREL, H. *et al.* Design and synthesis of a 3'-O-allyl photocleavable fluorescent nucleotide as a reversible terminator for DNA sequencing by synthesis. **Proceedings of the National Academy of Sciences of the United States of America**, v. 102, n. 17, p. 5932–5937, 26 abr. 2005.

SANGER, F.; NICKLEN, S.; COULSON, A. R. DNA sequencing with chain-terminating inhibitors. **The Proceedings of the National Academy of Sciences**, v. 74, n. 12, p. 5463–5467, 1977.

SANTIAGO, P. S. *et al.* Giant extracellular *Glossoscolex paulistus* Hemoglobin (HbGp) upon interaction with cetyltrimethylammonium chloride (CTAC) and sodium dodecyl sulphate (SDS) surfactants: Dissociation of oligomeric structure and autoxidation. **Biochimica et Biophysica Acta - General Subjects**, v. 1770, n. 4, p. 506–517, 2007.

SANTIAGO, P. S. *et al.* Dynamic light scattering and optical absorption spectroscopy study of pH and temperature stabilities of the extracellular hemoglobin of *Glossoscolex paulistus*. **Biophysical Journal**, v. 94, n. 6, p. 2228–2240, mar. 2008.

SANTIAGO, P. S. *et al.* Isoelectric point determination for *Glossoscolex paulistus* extracellular hemoglobin: Oligomeric stability in acidic pH and relevance to protein-surfactant interactions. **Langmuir**, v. 26, n. 12, p. 9794–9801, 2010.

SHADEO, A. *et al.* Comprehensive serial analysis of gene expression of the cervical transcriptome. **BMC Genomics**, v. 8, n. 1, p. 142, 2007.

SHARMA, K.; GARG, V. K. Vermicomposting of Waste: A Zero-waste Approach for waste Management. In: In: TAHERZADEH, M. *et al.* (ed.). **Sustainable Resource Recovery and Zero Waste Approaches**. [s.l.] Elsevier, 2019. chap.10, p. 133–164.

SIMS, D. *et al.* Sequencing depth and coverage: key considerations in genomic analyses. **Nature Reviews Genetics**, v. 15, n. 2, p. 121–132, 17 fev. 2014.

SKOOG, D.; HOLLER, J.; NIEMAN, T. Molecular luminescence spectrometry. **Principles of Instrumental Analysis**, p. 361–388, 2018.

SOUZA. **Caracterização e imobilização da hemoglobina extracelular de *Amyntas gracilis* em substrato sólido visando a elaboração de biossensores de contaminação ambiental**. Dissertação (Mestrado em Biotecnologia). Universidade Estadual Paulista, Instituto de Química, Araraquara, 2018.

STOKES, G. G. On the change of refrangibility of light. **Philosophical Transactions of the Royal Society of London**, v. 142, p. 463–562, 31 dez. 1852.

STRONG, F. C. Theoretical basis of bouguer-beer law of radiation absorption. **Analytical Chemistry**, v. 24, n. 2, p. 338–342, 18 fev. 1952.

SÜDHOF, T. C. *et al.* The LDL receptor gene: A mosaic of exons shared with different proteins. **Science**, v. 228, n. 4701, p. 815–822, 17 maio 1985.

SUZUKI, T.; RIGGS, A. F. Linker chain L1 of earthworm hemoglobin. Structure of gene and protein: homology with low density lipoprotein receptor. **Journal of Biological Chemistry**, v. 268, n. 18, p. 13548–13555, jun. 1993.

SUZUKI, T.; VINOGRADOV, S. N. Globin and linker sequences of the giant extracellular hemoglobin from the leech *Macrobdella decora*. **Journal of Protein Chemistry**, v. 22, n. 3, 2003.

TERWILLIGER, R. C.; TERWILLIGER, N. B.; SCHABTACH, E. Comparison of chlorocruorin and annelid hemoglobin quaternary structures. **Comparative Biochemistry and Physiology Part A: Physiology**, v. 55, n. 1, p. 51–55, jan. 1976.

THE GENE ONTOLOGY CONSORTIUM. The Gene Ontology Resource: 20 years and still GOing strong. **Nucleic Acids Research**, v. 47, n. D1, p. D330–D338, 8 jan. 2019.

TRENT, J. T. *et al.* Crystallographic analysis of synechocystis cyanoglobin reveals the structural changes accompanying ligand binding in a hexacoordinate hemoglobin. **Journal of Molecular Biology**, v. 341, n. 4, p. 1097–1108, ago. 2004.

TRUMBO, T. A. *et al.* Applied spectrophotometry: Analysis of a biochemical mixture. **Biochemistry and Molecular Biology Education**, v. 41, n. 4, p. 242–250, 1 jul. 2013.

VALOUEV, A. *et al.* A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. **Genome research**, v. 18, n. 7, p. 1051–1063, jul. 2008.

VAN DIJK, E. L. *et al.* Ten years of next-generation sequencing technology. **Trends in Genetics**, v. 30, n. 9, p. 418–426, 1 set. 2014.

VAN DIJK, E. L. *et al.* The Third Revolution in Sequencing Technology. **Trends in**

Genetics, v. 34, n. 9, p. 666–681, 2018.

VARNEY, J. *et al.* Mini-review on the properties and possible applications of therapeutic oxygen carrier Hemarina-M101. **Transfusion and Apheresis Science**, v. 60, n. 1, p. 103016, 2021.

VEENSTRA, J. A. Neuropeptide evolution: Neurohormones and neuropeptides predicted from the genomes of *Capitella teleta* and *Helobdella robusta*. **General and Comparative Endocrinology**, v. 171, n. 2, p. 160–175, abr. 2011.

VOET, D.; VOET, J. G. **Bioquímica**. 4. ed. Porto Alegre: Artmed, 2013.

WANG, K. Y. *et al.* Recombinant protein production of earthworm lumbrokinase for potential antithrombotic application. **Evidence-Based Complementary and Alternative Medicine**, v. 2013, p. 1–8, 2013.

WANG, X.-M. *et al.* Earthworm protease in anti-thrombosis and anti-fibrosis. **Biochimica et Biophysica Acta (BBA) - General Subjects**, v. 1863, n. 2, p. 379–383, 1 fev. 2019.

WANG, X. *et al.* *Amyntas corticis* genome reveals molecular mechanisms behind global distribution. **Communications Biology**, v. 4, n. 1, p. 135, 29 dez. 2021.

WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. **Nature Reviews Genetics**, v. 10, n. 1, p. 57–63, jan. 2009.

WATERHOUSE, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. **Nucleic Acids Research**, v. 46, 2018.

WEBER, R. E.; VINOGRADOV, S. N. Nonvertebrate hemoglobins: functions and molecular adaptations. **Physiological Reviews**, v. 81, n. 2, p. 569–628, 1 abr. 2001.

WILFINGER, W. W.; MACKEY, K.; CHOMCZYNSKI, P. Effect of pH and ionic strength on the spectrophotometric assessment of nucleic acid purity. **BioTechniques**, v. 22, n. 3, p. 474–481, mar. 1997.

YAO, X. *et al.* Effects of a novel fungicide benzovindiflupyr in *Eisenia fetida*: Evaluation through different levels of biological organization. **Environmental Pollution**, v. 271, p. 116336, fev. 2021.

YU, Z. H.; ZHANG, Z. Y. Regulatory mechanisms and novel therapeutic targeting strategies for protein tyrosine phosphatases. **Chemical Reviews**, v. 118, n. 3, p. 1069–1091, 2018.

ZAL, F.; ROUSSELOT, M. Extracellular hemoglobins from annelids, and their potential use in biotechnology. In: LA BARRE, S.; KORNPROBST, J.-M. (Eds.). **Outstanding Marine Molecules: Chemistry, Biology, Analysis**. [s.l.] Wiley Blackwell, 2014. chap.16, p. 361–373.

ZHANG, Z. Y. Protein tyrosine phosphatases: prospects for therapeutics. **Current Opinion in Chemical Biology**, v. 5, n. 4, p. 416–423, ago. 2001.

ZIMMERMAN, D. *et al.* Direct comparison of oligochaete erythrocruorins as potential blood substitutes. **Bioengineering & Translational Medicine**, v. 2, n. 2, p. 212–221, jun. 2017.

ZWARYCZ, A. S. *et al.* Timing and scope of genomic expansion within annelida: Evidence from Homeoboxes in the genome of the earthworm *Eisenia fetida*. **Genome Biology and Evolution**, v. 8, n. 1, p. 271–281, jan. 2016.