



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE CIÊNCIAS
DEPARTAMENTO DE BIOQUÍMICA E BIOLOGIA MOLECULAR
CURSO DE GRADUAÇÃO EM BACHARELADO EM BIOTECNOLOGIA

JAMILE MARIA DA SILVA DOS SANTOS

Caracterização *in silico* de Enzimas-Chave da Biossíntese de Triacilgliceróis em Dendzeiro,
Elaeis guineensis, por Anotação Funcional e Modelagem Estrutural

Fortaleza
2025

JAMILE MARIA DA SILVA DOS SANTOS

Caracterização *in silico* de enzimas-chave da biossíntese de triacilgliceróis em Dendzeiro,
Elaeis guineensis, por anotação funcional e modelagem estrutural

Trabalho de conclusão de curso
apresentado ao Curso de
Graduação em Biotecnologia do
Centro de Ciências da
Universidade Federal do Ceará,
como requisito parcial à
obtenção do grau de Bacharel em
Biotecnologia.

Orientadora: Prof. Dr. Nicholas
Costa Barroso

FORTALEZA
2025

JAMILE MARIA DA SILVA DOS SANTOS

Caracterização *in silico* de enzimas-chave da biossíntese de triacilgliceróis em Dendzeiro,
Elaeis guineensis, por anotação funcional e modelagem estrutural

Trabalho de conclusão de curso
apresentado ao Curso de
Graduação em Biotecnologia do
Centro de Ciências da
Universidade Federal do Ceará,
como requisito parcial à
obtenção do grau de Bacharel em
Biotecnologia.

Orientadora: Prof. Dr. Nicholas
Costa Barroso

Aprovado: 19/12/2025

Prof. Dr. Nicholas Costa Barroso Lima
Orientador
Departamento de Bioquímica e Biologia Molecular - UFC

Prof. ^a. Dra. Daniele de Oliveira Bezerra de Sousa
Departamento de Bioquímica e Biologia Molecular - UFC

Prof. Dr. Murilo Siqueira Alves
Departamento de Bioquímica e Biologia Molecular - UFC

AGRADECIMENTOS

“Consagre ao Senhor tudo o que você faz, e os seus planos serão bem-sucedidos.”

Agradeço, primeiramente, a Deus, que nunca me abandonou nem me desamparou mesmo quando eu não estive fortificada na fé, e que sempre foi a luz que me guiou e minha fortaleza.

Desde pequena, aprendi o valor do trabalho e do suor derramado nele, a ser grata e a honrar os que me honram. Aos meus pais, Antônio Pedro e Vanderlene, que são meu exemplo de vida, de amor e de luta, que trabalharam (e trabalham) incansavelmente para que eu tivesse acesso a melhores oportunidades e me orientaram para que pudesse reconhecer isso, não há palavras para agradecer o apoio e o amor que eu recebo diariamente dos que me deram o dom da vida, é tudo por vocês.

Às minhas irmãs Tayná e Bianca, e à minha sobrinha, Maria Heloísa, que são as minhas companheiras de vida: foi também com elas que aprendi a ter coragem e força. Todo o sucesso que chegar até mim também é delas.

Aos amigos que fiz na graduação e foram meus companheiros de trajetória, esses quatro anos de acolhimento, risadas, momentos bons e ruins estarão para sempre marcados em mim: Caio, Clarissa, Lewi e Gabriela. Aos que fizeram parte da minha história na UFC, mesmo em outros cursos, me recebendo sempre de braços abertos no bloco da Geografia ou pelos corredores do CT: Thiago, Victória, Felipe Rufino, Maria Alice e Miguel. Aos que conheci em outra graduação, mas me apoiaram a seguir meus sonhos e mudar de curso e, assim, se tornaram amigos de vida: Bruno e Ilana. E, por último e não menos importante, aos amigos que me acompanham desde o ensino fundamental, que me viram entrar na faculdade, amadurecer, passar por infernos e paraísos e sempre se mostraram um porto seguro: João Pedro, Iasmym e Gabriel.

Ao meu bem, meu amor, minha paixão. Você me fortalece, me cura, me motiva a ser melhor, me exalta e me ama, e como diria Tim Maia, “é mais do que sei é mais que pensei é mais que esperava, baby”. Te amo para sempre, Nicholas.

Ao meu orientador, Nicholas Barroso, pela confiança, paciência e gentileza durante todo o processo. Foi um privilégio ser sua aluna.

Aos professores Daniele Sousa, Murilo Alves e Raquel Benevides, por gentilmente aceitarem fazer parte da banca avaliadora.

A todos os professores que cruzaram minha vida acadêmica e somaram na construção do meu conhecimento, senso crítico e ético, que me marcaram positivamente, e me guiaram nessa jornada.

A todos que compartilharam e compartilham comigo a delícia que é viver, que tenhamos a coragem necessária para enfrentar nossos medos e lutar pelos nossos sonhos, sempre desfrutando o percurso e sendo gratos pelo passado. Desejo uma boa e longa vida a todos.

“Viver é arriscar tudo; caso contrário, você é apenas um amontoado inerte de moléculas montadas aleatoriamente, à deriva para onde o universo te levar...”.

– Rick and Morty.

RESUMO

O dendezeiro (*Elaeis guineensis* Jacq.) é uma das principais oleaginosas cultivadas no mundo, e o óleo extraído de seus frutos apresenta grande relevância econômica e industrial. A compreensão dos mecanismos moleculares que controlam a biossíntese de ácidos graxos e a montagem de triacilgliceróis (TAG) nessa espécie é fundamental para subsidiar estratégias de melhoramento voltadas ao aumento de produtividade e à qualidade do óleo. Neste trabalho, o proteoma do dendezeiro foi anotado funcionalmente para caracterizar *in silico* enzimas centrais da lipogênese em *E. guineensis*, integrando anotação funcional, análise de vias metabólicas e predição estrutural. O proteoma da espécie foi obtido a partir do NCBI e anotado com o KOfamScan, utilizando perfis de modelos ocultos de Markov (HMM) específicos para KEGG Orthology (KO). Foram considerados de alta confiança apenas os *hits* com *Score* HMM superior ao limiar específico de cada KO (threshold recomendado pelo KOfamScan) e *E-value* $< 1 \times 10^{-10}$. A partir da lista de KOs válidos, as vias metabólicas foram reconstruídas no KEGG Mapper, evidenciando forte cobertura da biossíntese de ácidos graxos e do metabolismo de glicerolipídios. Com base na anotação, foram selecionadas cinco enzimas-chave: acetil-CoA carboxilase (ACC), 3-oxoacil-ACP sintase I (KAS I), estearoil-ACP dessaturase (SAD/FAD2-like), glicerol-3-fosfato aciltransferase (GPAT) e diacilglicerol aciltransferase (DGAT). Essas proteínas foram submetidas à caracterização físico-química e estrutural com ProtParam, SignalP, TMHMM, DeepLoc2 e CDD-Search, o que permitiu confirmar a presença de domínios catalíticos conservados compatíveis com a função de cada KO, enquanto a regulação pós-traducional foi avaliada com NetPhos 3.1. A análise de motivos conservados pelo MEME Suite identificou assinaturas compartilhadas entre enzimas plastidiais e do retículo endoplasmático, sugerindo elementos regulatórios e estruturais comuns ao longo da rota lipídica. Por fim, a modelagem tridimensional das cinco proteínas com AlphaFold/ColabFold, seguida de visualização em PyMOL, permitiu mapear domínios, motivos e sítios de fosforilação em nível estrutural. Em conjunto, os resultados consolidam essas cinco enzimas como marcadores *in silico* da lipogênese em *E. guineensis*.

Palavras-chave: *Elaeis guineensis*; Lipogênese; Triacilgliceróis (TAG); Anotação funcional de proteomas; KEGG Orthology; Modelagem estrutural de proteínas.

ABSTRACT

Oil palm (*Elaeis guineensis* Jacq.) is one of the main oilseed crops cultivated worldwide, and the oil extracted from its fruits has high economic and industrial relevance. Understanding the molecular mechanisms that control fatty acid biosynthesis and triacylglycerol (TAG) assembly in this species is essential to support breeding strategies aimed at increasing productivity and improving oil quality. In this study, the oil palm proteome was functionally annotated to in silico characterize central lipogenic enzymes in *E. guineensis*, integrating functional annotation, metabolic pathway analysis and structural prediction. The species proteome was obtained from NCBI and annotated with KOfamScan using Hidden Markov Model (HMM) profiles specific for KEGG Orthology (KO). Only hits with HMM Score above the KO-specific threshold (as recommended by KOfamScan) and E-value $< 1 \times 10^{-10}$ were considered high-confidence. From the list of valid KOs, metabolic pathways were reconstructed in KEGG Mapper, revealing strong coverage of fatty acid biosynthesis and glycerolipid metabolism. Based on this annotation, five key enzymes were selected: acetyl-CoA carboxylase (ACC), 3-oxoacyl-ACP synthase I (KAS I), stearyl-ACP desaturase (SAD/FAD2-like), glycerol-3-phosphate acyltransferase (GPAT) and diacylglycerol acyltransferase (DGAT). These proteins were subjected to physicochemical and structural characterization with ProtParam, SignalP, TMHMM, DeepLoc2 and CDD-Search, which confirmed the presence of conserved catalytic domains consistent with the function assigned to each KO, while post-translational regulation was evaluated with NetPhos 3.1. The analysis of conserved motifs with MEME Suite identified signatures shared between plastidial and endoplasmic reticulum enzymes, suggesting common structural and regulatory elements along the lipid pathway. Finally, three-dimensional modeling of the five proteins with AlphaFold/ColabFold, followed by visualization in PyMOL, allowed mapping of domains, motifs and phosphorylation sites at the structural level. Taken together, the results support these five enzymes as in silico markers of lipogenesis in *E. guineensis*.

Keywords: *Elaeis guineensis*; Lipogenesis; Triacylglycerols (TAG); Proteome functional annotation; KEGG Orthology; Protein structural modeling.

LISTA DE ILUSTRAÇÕES

Figura 1 - <i>Elaeis guineensis</i> Jacq.....	14
Figura 2 - Frutos do dendê.....	15
Figura 3 - Dendê.....	16
Figura 4 - Óleo extraído do dendê.....	18
Figura 5 - Map00061 com as enzimas ACC e SAD em destaque na sua posinação correspondente na via.....	37
Figura 6 - Map00561 com as enzimas DGAT e GPAT em evidência.....	40
Figura 7 - Map04146 com a enzima DGAT em evidência.....	42
Figura 8 - Modelo 3D da ACC colorido por pLDDT, com o domínio de biotin carboxylase destacado.....	49
Figura 9 - ACC com motivos selecionados em representação <i>sticks</i> e identificação dos resíduos.....	50
Figura 10 - ACC com sítios de fosforilação selecionados em representação <i>spheres</i>	50
Figura 11 - Modelo 3D da KAS I colorido por pLDDT, com o domínio ketoacyl synthase destacado.....	51
Figura 12 - KAS I com motivos selecionados em representação <i>sticks</i>	52
Figura 13 - KAS I com sítios de fosforilação selecionados em representação <i>spheres</i>	52
Figura 14 - Modelo 3D da SAD/FAD2 like colorido por pLDDT, com o domínio FA_desaturase destacado.....	54
Figura 15 - SAD/FAD2 like com motivos selecionados em <i>sticks</i>	54
Figura 16 - SAD/FAD2 like com sítios de fosforilação selecionados em <i>spheres</i>	55
Figura 17 - Modelo 3D da GPAT colorido por pLDDT, com domínio aciltransferase destacado.....	56
Figura 18 - GPAT com motivos selecionados em <i>sticks</i>	57
Figura 19 - GPAT com sítios de fosforilação selecionados em <i>spheres</i>	58
Figura 20 - Modelo 3D da DGAT colorido por pLDDT, com domínio aciltransferase destacado.....	59
Figura 21 - DGAT com motivos selecionados em <i>sticks</i>	60
Figura 22 - DGAT com sítios de fosforilação selecionados em <i>spheres</i>	60

LISTA DE TABELAS

Tabela 1 - Proteínas selecionadas para caracterização funcional e estrutural.....	35
Tabela 2 - Propriedades físico-químicas previstas pelo ProtParam para as cinco enzimas selecionadas da lipogênese em <i>Elaeis guineensis</i>.....	43
Tabela 3 – Parâmetros físico-químicos e predição de localização (ProtParam,DeepLoc2, TMHMM e SignalP).....	44
Tabela 4 – Domínios conservados identificados nas enzimas-chave da lipogênese em <i>Elaeis guineensis</i> por CDD-Search.....	45
Tabela 5 – Sítios de fosforilação previstos nas cinco enzimas-chave da lipogênese em <i>Elaeis guineensis</i> pelo NetPhos 3.1.....	46
Tabela 6 – Motivos conservados compartilhados entre as cinco enzimas-chave da lipogênese em <i>Elaeis guineensis</i>, identificados pelo MEME Suite.....	47

LISTA DE ABREVIATURAS E SIGLAS

ACC	Acetil-CoA Carboxilase (Acetyl-CoA Carboxylase)
ACP	Proteína Carreadora de Acila (Acyl Carrier Protein)
BLAST	Basic Local Alignment Search Tool
CDS	Sequência Codificadora (Coding Sequence)
CDD	Conserved Domain Database
DAG	Diacilglicerol (Diacylglycerol)
DGAT	Diacilglicerol O-Aciltransferase (Diacylglycerol O-acyltransferase)
DNA	Ácido Desoxirribonucleico
FAS II	Fatty Acid Synthase type II (Síntese de Ácidos Graxos tipo II)
GPAT	Glicerol-3-fosfato Aciltransferase (Glycerol-3-phosphate Acyltransferase)
GRAVY	Grand Average of Hydropathicity (Média Global de Hidropaticidade)
HMM	Hidden Markov Model (Modelo Oculto de Markov)
KAS I	3-Cetoacil-ACP Sintase I (3-oxoacyl-ACP synthase I)
KEGG	Kyoto Encyclopedia of Genes and Genomes
KO	KEGG Orthology (Ortólogo KEGG)
KOfamScan	Ferramenta de anotação de ortologia KEGG baseada em perfis HMM
MEME	Multiple EM for Motif Elicitation
mRNA	RNA mensageiro (Messenger RNA)
NCBI	National Center for Biotechnology Information
NGS	Next-Generation Sequencing (Sequenciamento de Nova Geração)
NetPhos	Ferramenta para predição de sítios de fosforilação em proteínas
PDB	Protein Data Bank
pI	Ponto Isoelétrico
pLDDT	Predicted Local Distance Difference Test (Índice de confiança local das predições do AlphaFold)
RE	Retículo Endoplasmático
RNA	Ácido Ribonucleico
SAD	Estearoil-ACP Dessaturase (Stearoyl-ACP desaturase)
TAG	Triacilglicerol (Triacylglycerol)
TMHMM	Transmembrane Hidden Markov Model (Preditor de hélices transmembranares)

SUMÁRIO

1 INTRODUÇÃO.....	13
2 OBJETIVOS.....	19
2.1 OBJETIVO GERAL.....	19
2.2 OBJETIVOS ESPECÍFICOS.....	19
3 REFERENCIAL TEÓRICO.....	21
3.1 BIOINFORMÁTICA COMO EIXO ESTRUTURANTE DA PESQUISA EM ORGANISMOS VEGETAIS.....	21
3.2 A CASCATA OMICS E SUA RELEVÂNCIA PARA A BIOLOGIA VEGETAL.....	21
3.3 O PAPEL DA ANOTAÇÃO DE ORTOLOGIA (KO) NA CARACTERIZAÇÃO FUNCIONAL.....	22
3.4 ANOTAÇÃO FUNCIONAL, DOMÍNIOS PROTEICOS E MAPEAMENTO METABÓLICO.....	23
3.5 CARACTERIZAÇÃO FÍSICO-QUÍMICA E REGULAÇÃO PÓS-TRADUCIONAL..	23
3.6 PREDIÇÃO ESTRUTURAL DE PROTEÍNAS NA ANÁLISE FUNCIONAL.....	24
3.7 VALIDAÇÃO ESTRUTURAL E ANÁLISE DE MOTIVOS.....	25
4 MATERIAIS E MÉTODOS.....	26
4.1 OBTENÇÃO DO PROTEOMA E BASES DE DADOS UTILIZADAS.....	26
4.2 ANOTAÇÃO FUNCIONAL BASEADA EM MODELOS HMM (KOFAMSCAN).....	26
4.3 SELEÇÃO DAS CINCO ENZIMAS-CHAVE E MAPEAMENTO EM VIAS KEGG..	27
4.4 VALIDAÇÃO DA FUNÇÃO E IDENTIDADE DAS PROTEÍNAS.....	28
4.5 CARACTERIZAÇÃO FÍSICO-QUÍMICA.....	28
4.6 ANÁLISE DE MOTIVOS E REGULAÇÃO PÓS-TRADUCIONAL.....	28
4.7 MODELAGEM ESTRUTURAL TRIDIMENSIONAL.....	29
4.8 REPRODUTIBILIDADE COMPUTACIONAL.....	29
5 RESULTADOS E DISCUSSÃO.....	30
5.1 PROTEÍNAS SELECIONADAS PARA CARACTERIZAÇÃO FUNCIONAL E ESTRUTURAL.....	30
5.2 ACC (K01961): O PONTO DE CONTROLE LIMITANTE.....	30
5.3 KAS I (K00655): O MÓDULO DE ALONGAMENTO.....	31
5.4 SAD/FAD2-LIKE (K03921): QUALIDADE E INSATURAÇÃO.....	32
5.5 GPAT (K13508): O INÍCIO DA VIA DE KENNEDY.....	32
5.6 DGAT (K13339): A ESTOCAGEM FINAL.....	33
5.7 ANÁLISE DE VIA METABÓLICAS COM KEGG MAPPER.....	34

5.8 CARACTERIZAÇÃO ESTRUTURAL E REGULATÓRIA DAS CINCO PROTEÍNAS-CHAVE.....	42
5.9 MODELAGEM TRIDIMENSIONAL DAS CINCO PROTEÍNAS CHAVE.....	47
5.10 ENZIMAS PLASTIDIAIS: ACC, KAS I E SAD/FAD2 LIKE.....	47
5.11 ENZIMAS DA VIA DE KENNEDY: GPAT E DGAT.....	54
5.12 INTEGRAÇÃO DOS RESULTADOS ESTRUTURAIS.....	60
6 CONCLUSÃO.....	62
7 REFERÊNCIAS.....	63

1 INTRODUÇÃO

O dendezeiro, conhecido cientificamente como *Elaeis guineensis* Jacq., é uma palmeira originária da costa ocidental da África, região onde há registros arqueológicos e etnobotânicos que indicam sua utilização há milhares de anos (ANDRADE, 2012). Essa palmeira foi introduzida no Brasil durante o período colonial e, ao longo dos séculos, adaptou-se de maneira excepcional às condições tropicais úmidas do país, especialmente na Bahia e posteriormente na Amazônia, tornando-se uma das espécies perenes mais expressivas para a economia agrícola nacional (EMBRAPA, 2010; EMBRAPA, 2016). O dendezeiro apresenta um ciclo de vida longo, um porte imponente e uma capacidade singular de produção de frutos oleaginosos, fatores que contribuíram para sua difusão e interesse crescente em diversos setores industriais e científicos.

Figura 1 - *Elaeis guineensis* Jacq.



Fonte: Acervo Horto botânico – Museu Nacional/UFRJ

O fruto dessa palmeira, amplamente conhecido pelo nome popular dendê, consiste em uma drupa formada por um epicarpo fino, um mesocarpo rico em óleo e uma amêndoa no interior. Desse fruto extraem-se dois óleos distintos: o óleo de dendê, obtido do mesocarpo, e o óleo de palmiste, proveniente da amêndoa. Ambos possuem composições químicas particulares e apresentam importância econômica substancial (EMBRAPA, 2023). A estrutura do fruto e sua elevada capacidade de acumular lipídios são fatores amplamente estudados em pesquisas que envolvem oleaginosas tropicais, devido à eficiência metabólica com que o mesocarpo converte fotoassimilados em triacilgliceróis (GUNSTONE, 2011; GUNSTONE *et al.*, 2021). O óleo de dendê, em especial, destaca-se por apresentar uma combinação balanceada de ácidos graxos saturados e insaturados, além de elementos bioativos como carotenoides, tocoferóis e tocotrienóis, substâncias essenciais para a estabilidade química do óleo e para suas propriedades nutricionais e funcionais (YING *et al.*, 2024).

Figura 2 - Frutos do dendê



Fonte: Daniela Collares – EMBRAPA/ageitec

Figura 3 - Dendê

Fonte: Daniela Collares – EMBRAPA/Ageitec

O cultivo do dendezeiro no Brasil intensificou-se principalmente a partir de estudos de zoneamento agroecológico e manejo sustentável conduzidos pela Embrapa, os quais identificaram regiões aptas ao plantio e permitiram a expansão da cultura sem impactos de desmatamento quando integrada a áreas antropizadas ou degradadas (EMBRAPA, 2010; EMBRAPA, 2016). No estado do Pará, a palma de óleo tornou-se uma das culturas mais estratégicas para o desenvolvimento regional, favorecida por fatores como alta luminosidade, elevada umidade atmosférica, solos adequados e disponibilidade de mão de obra especializada. A capacidade da espécie de se integrar a sistemas agroflorestais e de contribuir para a recuperação de áreas anteriormente utilizadas pela pecuária extensiva tem reforçado seu potencial produtivo e ecológico.

A produção de dendê no Brasil ocupa posição de destaque na economia agroindustrial, pois o país é o maior produtor da América Latina e detentor de alto potencial de expansão sustentável (EMBRAPA, 2023). O óleo extraído do fruto representa um dos

insumos vegetais mais utilizados no mundo, sendo encontrado em alimentos processados, margarinas, sorvetes, chocolates, conservantes alimentares, produtos cosméticos, sabonetes, detergentes, tintas, resinas, fármacos e, mais recentemente, em formulações de biocombustíveis e biodiesel (EMBRAPA, 2023; GUNSTONE, 2011). A versatilidade do óleo de dendê resulta da sua estabilidade oxidativa, de seu elevado ponto de fusão e de sua capacidade funcional de conferir textura e propriedades físico-químicas específicas a inúmeros produtos industriais. Além disso, o rendimento extremamente alto da cultura, chegando a superar em múltiplas vezes o rendimento de outras oleaginosas como soja, girassol e canola, reforça sua importância estratégica na matriz oleaginosa global.

A importância do dendê na economia brasileira também se manifesta pela geração de emprego e renda em regiões historicamente afastadas da industrialização. Em áreas amazônicas, por exemplo, a cultura tem sido adotada como alternativa de diversificação produtiva e fonte de dinamização econômica, contribuindo para a fixação de famílias no campo e para o desenvolvimento de cadeias produtivas integradas (EMBRAPA, 2010). A literatura aponta que a palma de óleo representa uma das poucas culturas capazes de unir elevada produtividade, demanda internacional crescente, viabilidade climática e potencial de sustentabilidade quando implantada sob diretrizes técnicas responsáveis (ANDRADE, 2012; EMBRAPA, 2023).

Do ponto de vista bioquímico, o óleo de dendê possui características singulares que ampliam sua utilidade industrial e seu valor biológico. Esse óleo é rico em compostos bioativos com propriedades antioxidantes, anti-inflamatórias e protetoras celulares, incluindo carotenoides como β -caroteno, precursores de vitamina A, e formas de vitamina E como tocoferóis e tocotrienóis (YING *et al.*, 2024). Esses componentes conferem ao óleo elevada estabilidade térmica, além de benefícios fisiológicos documentados na literatura médico-nutricional, incluindo efeitos sobre o estresse oxidativo e o metabolismo lipídico corporal (YING *et al.*, 2024).

Do ponto de vista metabólico, o acúmulo de lipídios no mesocarpo resulta da interação entre a biossíntese de ácidos graxos, predominantemente em compartimentos plastidiais, e sua posterior esterificação em triacilgliceróis pela chamada Via de Kennedy, localizada em membranas do retículo endoplasmático. Nessa via, aciltransferases, incluindo a glicerol-3-fosfato aciltransferase (GPAT) e a diacilglicerol aciltransferase (DGAT) promovem a montagem sequencial do esqueleto de glicerol até TAG, a forma final de armazenamento de óleo em tecidos vegetais (OHLROGGE; BROWSE, 1995). A composição lipídica específica do óleo de palma faz com que ele seja uma matéria-prima essencial não

apenas na alimentação, mas também em produtos farmacêuticos, nutracêuticos e formulações energéticas de alto desempenho (GUNSTONE *et al.*, 2021).

Na culinária brasileira, o óleo de dendê possui relevância cultural profunda, especialmente na Bahia, onde constitui elemento central da gastronomia afro-indígena. Seu aroma e cor característicos são determinantes na elaboração de pratos tradicionais como acarajé, vatapá, caruru e xinxim de galinha. Esse uso cultural está intimamente ligado ao simbolismo religioso e à ancestralidade africana, o que reforça a importância sociocultural do fruto no Brasil contemporâneo (ANDRADE, 2012). Já na indústria, a estabilidade oxidativa e o comportamento térmico do óleo justificam sua ampla utilização como gordura estrutural em processos de fritura, panificação e confeitaria, além de sua crescente adoção na formulação de biodiesel, especialmente em contextos onde há necessidade de combustíveis avançados derivados de matérias-primas renováveis (EMBRAPA, 2023).

Figura 4 - Óleo extraído do dendê



Fonte: EMBRAPA - iStock

A expansão da cultura do dendezeiro, entretanto, requer suporte técnico e científico, especialmente no que diz respeito à genética da espécie, ao entendimento da regulação metabólica responsável pelo acúmulo lipídico e à adaptação da planta frente a estresses ambientais. Nesse sentido, estudos envolvendo genômica, transcriptômica e bioinformática

são essenciais para desvendar os mecanismos moleculares que governam a produção de óleo, a arquitetura fisiológica da planta e suas respostas ambientais. O genoma de *Elaeis guineensis* vem sendo progressivamente estudado por diferentes grupos de pesquisa, com avanços significativos na compreensão dos genes associados à biossíntese lipídica, ao metabolismo primário e secundário e à resposta a estresses bióticos e abióticos. Esses conhecimentos, integrados à biotecnologia moderna, têm potencial para aprimorar programas de melhoramento genético e ampliar a produtividade com sustentabilidade.

Nesse cenário, este trabalho se insere na interseção entre biologia molecular, bioinformática e agronomia tropical, enfocando a caracterização *in silico* de enzimas centrais do metabolismo lipídico em *Elaeis guineensis*. Para isso, o proteoma da espécie foi anotado funcionalmente, as vias metabólicas associadas à biossíntese de ácidos graxos e à montagem de triacilgliceróis foram mapeadas, e um conjunto de proteínas de interesse biotecnológico foi submetido à predição estrutural tridimensional. A compreensão integrada desses elementos representa um passo essencial para o avanço do conhecimento sobre a fisiologia do dendezeiro e para o desenvolvimento de estratégias científicas e tecnológicas voltadas à maximização do potencial produtivo do fruto no Brasil.

2 OBJETIVOS

2.1 Objetivo geral

Caracterizar *in silico*, em nível funcional, estrutural e regulatório, cinco enzimas-chave envolvidas na biossíntese de ácidos graxos e na montagem de triacilgliceróis em *Elaeis guineensis*, com base na anotação por KEGG Orthology e em ferramentas de bioinformática.

2.2 Objetivos específicos

Os objetivos específicos são:

Anotar funcionalmente o proteoma de *Elaeis guineensis* utilizando o KOfamScan, gerando uma lista de KOs de alta confiança e mapeando-os em vias metabólicas no KEGG Mapper.

Identificar e selecionar cinco enzimas representativas das etapas limitantes da biossíntese de ácidos graxos e da Via de Kennedy (ACC, KAS I, SAD/FAD2-like, GPAT e DGAT), com base em critérios estatísticos (*Score* HMM e *E-value*) e na relevância biológica para a lipogênese.

Caracterizar as propriedades físico-químicas, a predição de localização subcelular e a presença de domínios conservados nessas cinco proteínas, empregando ProtParam, SignalP, TMHMM, DeepLoc2 e CDD-Search.

Inferir o potencial regulatório das enzimas por meio da predição de sítios de fosforilação e de motivos conservados, utilizando NetPhos e MEME Suite, relacionando esses padrões com a posição de cada proteína nas rotas lipídicas.

Modelar as estruturas tridimensionais das cinco proteínas selecionadas com AlphaFold/ColabFold, visualizar os modelos em PyMOL e mapear domínios, motivos e sítios

de fosforilação, discutindo as implicações estruturais para a biossíntese e o acúmulo de triacilgliceróis em *E. guineensis*.

3 REFERENCIAL TEÓRICO

3.1 Bioinformática como eixo estruturante da pesquisa em organismos vegetais

A bioinformática consolidou-se como um campo interdisciplinar indispensável para a biotecnologia moderna, atuando na interface entre biologia, computação e matemática. Seu avanço acompanha o ritmo acelerado da geração de dados produzidos pelas tecnologias de alto rendimento. Em espécies vegetais de interesse agrícola, como o dendezeiro (*Elaeis guineensis*), a bioinformática permite explorar a complexidade do genoma, caracterizar padrões de expressão gênica e compreender como processos metabólicos são regulados em diferentes condições fisiológicas e ambientais. Esse conjunto de ferramentas transformou o modo como sistemas biológicos são analisados, superando as abordagens tradicionais baseadas em experimentação isolada e permitindo uma interpretação integrada do organismo (GONÇALVES, 2023; SALGADO, 2023).

O surgimento da chamada Era *Omics* foi determinante para essa transformação. Ao possibilitar a análise simultânea de milhares de genes, proteínas e metabólitos, a abordagem ômica deslocou o foco das pesquisas biológicas da observação pontual para a compreensão sistêmica dos organismos. No caso das plantas, essa mudança proporcionou avanços importantes na elucidação de vias metabólicas, na identificação de fatores responsáveis pela resistência a estresses abióticos, na caracterização de redes regulatórias complexas e na otimização de características agronômicas de interesse econômico. Assim, o estudo integrado do genoma, transcriptoma e proteoma tornou-se uma estratégia essencial para o entendimento da biologia molecular de espécies de relevância econômica, como o dendezeiro.

3.2 A cascata omics e sua relevância para a biologia vegetal

A Cascata *Omics* refere-se ao conjunto de níveis biológicos estruturados em camadas interdependentes que descrevem o fluxo de informação do DNA ao fenótipo. Na base encontra-se o genoma, que reúne a totalidade da informação hereditária da espécie. Contudo,

a presença de um gene não implica necessariamente sua expressão; essa decisão é tomada em nível transcritômico, que caracteriza o conjunto de RNAs presentes em uma célula ou tecido em um momento específico. A expressão gênica resulta na produção de proteínas, que constituem o proteoma e desempenham funções estruturais e catalíticas essenciais. Por fim, essas proteínas atuam sobre rotas bioquímicas que originam o metaboloma, conjunto de moléculas responsáveis pelas características fisiológicas observáveis do organismo (NAVA *et al.*, 2023).

Essa abordagem sequencial é particularmente relevante na biologia vegetal, uma vez que plantas respondem continuamente ao ambiente e ajustam sua expressão gênica conforme condições externas, como variação hídrica, luminosidade, salinidade ou ataque patogênico. A interpretação dessa cascata permite compreender como estímulos ambientais são traduzidos em respostas químicas e fisiológicas. No dendezeiro, cuja produtividade de óleo depende da regulação fina de vias metabólicas lipídicas, o entendimento dos processos transcriptômicos é fundamental para determinar quais genes influenciam o acúmulo de triacilgliceróis e como esses genes se comportam em condições adversas (EMBRAPA, 2020; SALGADO, 2023).

3.3 O papel da anotação de ortologia (KO) na caracterização funcional

A anotação de ortologia KEGG (KEGG Orthology, KO), realizada por meio de perfis de Modelos Ocultos de Markov (Hidden Markov Models, HMM), constitui um dos métodos mais robustos para inferência funcional de proteínas, sobretudo em organismos nos quais dados transcriptômicos comparativos ainda não estão disponíveis. Ao fornecer uma classificação universal da função, os KOs atuam como o elo entre a sequência de aminoácidos e a participação da proteína em vias metabólicas específicas (KANEHISA *et al.*, 2006).

No presente trabalho, a priorização das proteínas não se baseou na variação da expressão gênica, mas sim na confiabilidade estatística da anotação funcional fornecida pelo HMM. Métricas como o score HMM e o E-value permitiram a seleção de *hits* de alta confiança, garantindo que as proteínas escolhidas representassem as enzimas centrais e conservadas envolvidas na biossíntese lipídica do dendezeiro.

3.4 Anotação funcional, domínios proteicos e mapeamento metabólico

A anotação funcional constitui a ponte entre a sequência gênica e a interpretação biológica do papel da proteína. Em proteínas, essa anotação é realizada principalmente pela identificação de domínios, unidades funcionais independentes e evolutivamente conservadas. Os domínios definem a estrutura tridimensional, a atividade catalítica e o papel bioquímico de uma proteína, sendo essenciais para inferir sua função mesmo quando não existem dados experimentais diretos (WANG *et al.*, 2021).

A utilização do KOfamScan, que emprega perfis HMM específicos para cada KO, foi o eixo inicial da análise, permitindo atribuir funções enzimáticas com alta precisão. Complementarmente, a ferramenta CDD-Search (NCBI) foi empregada para validar e refinar a identidade funcional das proteínas, sobretudo nos casos em que predições de membrana (TMHMM) e peptídeo sinal (SignalP) apresentaram inconsistências. Essa dupla abordagem garantiu a correta classificação funcional das enzimas selecionadas.

O mapeamento das proteínas nas vias do KEGG permitiu identificar rotas bioquímicas centrais para o metabolismo lipídico. Particularmente relevante é a Via de Kennedy, principal rota de formação de triacilgliceróis, que compreende etapas de acilação e esterificação essenciais para o acúmulo de óleo no mesocarpo (OHLROGGE; BROWSE, 1995). A identificação de cinco enzimas-chave dessa via estruturou todo o processo subsequente de caracterização estrutural e regulatória.

3.5 Caracterização físico-química e regulação pós-traducional

A predição das propriedades físico-químicas das proteínas é indispensável para inferir seu comportamento funcional e localização celular, fornecendo evidências estruturais para a validação das anotações enzimáticas. Ferramentas como o ProtParam avaliam parâmetros como o ponto isoeletrico (pI), que indica a carga líquida da proteína, a média de hidropaticidade (GRAVY), que infere a afinidade por membranas, e o índice de instabilidade, que auxilia na previsão da estabilidade metabólica.

O estudo dessas propriedades é complementado por análises de compartimentalização e de modificações pós-traducionais. A predição de segmentos transmembranares (TMHMM; KROGH *et al.*, 2001) e peptídeos sinal (SignalP; ALMAGRO *et al.*, 2019) é fundamental para enzimas envolvidas na biossíntese lipídica, uma vez que a síntese de ácidos graxos ocorre no plastídio e a montagem do triacilglicerol ocorre no retículo endoplasmático. Enzimas como GPAT e DGAT são tipicamente associadas ao RE, justificando as análises realizadas.

A fosforilação representa um dos mecanismos de regulação mais importantes para enzimas do metabolismo lipídico, especialmente para ACC e DGAT. A predição de sítios de fosforilação via NetPhos 3.1 (BLOM *et al.*, 1999) permitiu identificar potenciais regiões regulatórias relevantes para o controle da biossíntese de lipídios no dendezeiro.

3.6 Predição estrutural de proteínas na análise funcional

A estrutura tridimensional das proteínas é um elemento essencial para compreender seu mecanismo de ação. Métodos tradicionais de determinação estrutural, como cristalografia de raios-X e ressonância magnética nuclear, embora precisos, apresentam limitações técnicas e demandam altos custos e tempo prolongado (DIAS, 2007).

A predição estrutural por inteligência artificial ganhou destaque com o desenvolvimento do AlphaFold2, que demonstrou níveis de acurácia comparáveis aos métodos experimentais (JUMPER *et al.*, 2021). A ferramenta utiliza redes neurais profundas para prever interações entre resíduos, gerar modelos tridimensionais e estimar a confiabilidade estrutural local via pLDDT (The predicted local distance difference test). O uso do ColabFold, que integra o AlphaFold2 a algoritmos eficientes de alinhamento como MMseqs2, torna essas modelagens acessíveis e de alto desempenho (MIRDITA *et al.*, 2022).

No contexto do dendezeiro, a modelagem estrutural das enzimas selecionadas amplia a compreensão de seus mecanismos catalíticos, especificidade de substratos e potenciais interações moleculares, contribuindo diretamente para a interpretação de sua relevância biológica.

3.7 Validação estrutural e análise de motivos

A determinação da estrutura tridimensional de uma proteína representa o elemento mais completo para validar sua função catalítica e seu papel regulatório (WANG *et al.*, 2021). Nesse sentido, a análise de motivos conservados e a modelagem 3D atuam como ferramentas complementares na validação dos achados funcionais.

A análise de motivos realizada pelo MEME Suite (BAILE *et al.*, 2009) permite identificar sequências curtas conservadas associadas a sítios catalíticos, regiões de ligação a cofatores e elementos regulatórios. A detecção desses padrões nas enzimas-chave do dendezeiro contribui para refinar a identificação de sítios ativos e revelar mecanismos compartilhados entre proteínas que atuam em diferentes compartimentos celulares.

A modelagem estrutural obtida via AlphaFold/ColabFold possibilitou a visualização espacial dos domínios identificados pelo CDD e dos resíduos potencialmente fosforiláveis previstos pelo NetPhos, permitindo avaliar sua disposição no contexto tridimensional. A confiabilidade dessa interpretação é sustentada pelo índice pLDDT (JUMPER *et al.*, 2021), assegurando rigor na análise funcional final.

4 MATERIAIS E MÉTODOS

4.1 Obtenção do proteoma e bases de dados utilizadas

O presente estudo utilizou o proteoma de *Elaeis guineensis* disponibilizado pelo National Center for Biotechnology Information (NCBI), associado ao genoma de referência GCF_000442705.2, reconhecido como a montagem mais completa e amplamente empregada para a espécie (SINGH *et al.*, 2013). O arquivo *protein.faa*, contendo todas as sequências proteicas codificadas no genoma, foi a base para a anotação funcional, a seleção das enzimas-chave e as análises físico-químicas, regulatórias e estruturais subsequentes.

Os demais arquivos gerados ao longo do fluxo de análise bioinformática foram organizados e descritos em tabela própria, incluindo o proteoma completo, a saída bruta do KOfamScan, a lista de códigos KO filtrados por critérios estatísticos e o arquivo FASTA contendo as cinco proteínas selecionadas, além dos arquivos produzidos pelos *scripts* utilizados. Essa organização garantiu a rastreabilidade e a reprodutibilidade de todas as etapas computacionais.

4.2 Anotação funcional baseada em modelos HMM (KOfamScan)

A anotação funcional do proteoma completo foi realizada com o *software* KOfamScan, que emprega perfis HMM específicos para cada *ortholog* do banco KEGG Orthology (KANEHISA *et al.*, 2006). O arquivo *protein.faa* foi utilizado como entrada no KOfamScan, que atribuiu códigos KEGG Orthology (KO) às proteínas com base na similaridade aos perfis HMM. Para cada anotação, foram registrados o valor de *E-value*, o *Score* HMM e o limiar (*threshold*) recomendado para o respectivo KO.

Para assegurar alta confiança nas anotações utilizadas nas etapas posteriores, foram retidos apenas os *hits* que apresentaram simultaneamente *Score* HMM superior ao limiar específico de cada KO e *E-value* inferior a 1×10^{-10} , em consonância com critérios de rigor estatístico descritos para análises baseadas em perfis HMM (HUNTER *et al.*, 2012). A partir

desse conjunto filtrado de KOs foi construída a lista de KOs válidos que serviu de base para o mapeamento metabólico e para a seleção das proteínas de interesse.

4.3 Seleção das cinco enzimas-chave e mapeamento em vias KEGG

A partir da lista de códigos KO obtida após a filtragem de alta confiança do KOfamScan, todos os KOs foram projetados em vias metabólicas utilizando o módulo *Reconstruct Pathway* do KEGG Mapper (KANEHISA *et al.*, 2006; KANEHISA *et al.*, 2021). Inicialmente, foram mapeados todos os KOs nas vias globais do KEGG e contabilizado, para cada via, o número de KOs presentes. As vias destacadas para este estudo foram aquelas que apresentaram maior número de KOs válidos associados ao metabolismo lipídico, em especial a biossíntese de ácidos graxos (map00061) e o metabolismo de glicerolipídios (map00561), ambas diretamente relacionadas à síntese de triacilgliceróis em plantas oleaginosas (OHLROGGE; BROWSE, 1995).

Em seguida, os mapas gerados pelo KEGG Mapper foram inspecionados manualmente para identificar códigos KO associados a etapas definidas da rota lipídica, desde a entrada de carbono até a formação de triacilgliceróis. Com base nessa inspeção e em descrições clássicas da lipogênese vegetal, foram selecionadas cinco enzimas associadas a passos-chave: acetil-CoA carboxilase (ACC, K01961), 3-oxoacil-ACP sintase I (KAS I, K00655), esteroil-ACP dessaturase/SAD/FAD2-like (K03921), glicerol-3-fosfato aciltransferase (GPAT, K13508) e diacilglicerol aciltransferase (DGAT, K13339) (OHLROGGE; BROWSE, 1995; BELTRÃO; OLIVEIRA, 2007).

Após a definição desses cinco alvos, foram gerados no KEGG Mapper mapas específicos contendo apenas os KOs K01961, K00655, K03921, K13508 e K13339, de forma a localizar visualmente cada enzima nas vias previamente selecionadas. Nessa etapa, foi identificado o KO K13339 em um mapa adicional associado ao peroxissomo (map04146), que também foi incluído na inspeção por representar contexto subcelular relevante para o metabolismo de lipídios em plantas (OHLROGGE; BROWSE, 1995; BELTRÃO; OLIVEIRA, 2007).

4.4 Validação da função e identidade das proteínas

A identidade enzimática das cinco proteínas selecionadas foi avaliada por meio da identificação de domínios conservados utilizando o Conserved Domain Database (CDD) do National Center for Biotechnology Information, acessado pela ferramenta CDD-Search (MARCHLER-BAUER *et al.*, 2017; LU *et al.*, 2020). As sequências correspondentes aos KOs das proteínas selecionadas foram submetidas individualmente à ferramenta, em formato FASTA, para comparação com modelos de domínios específicos e superfamílias depositados no CDD.

Para cada proteína foram registrados os domínios identificados, os valores de *E-value* e as pontuações em bits associadas aos acertos, bem como a classificação em domínios específicos ou em superfamílias. Na interpretação dos resultados, foram considerados apenas domínios com valores de *E-value* inferiores aos limiares recomendados pelo próprio CDD, de modo a priorizar correspondências estatisticamente robustas (MARCHLER-BAUER *et al.*, 2017).

4.5 Caracterização físico-química

As propriedades físico-químicas das cinco proteínas selecionadas foram avaliadas utilizando o ProtParam, disponível no servidor ExPASy (GASTEIGER *et al.*, 2005). Foram obtidos os parâmetros de peso molecular, ponto isoelétrico teórico (pI), índice de instabilidade e média de hidropaticidade (GRAVY).

A predição de segmentos transmembranares foi realizada com o TMHMM v. 2.0, ferramenta baseada em modelos HMM treinados para identificar hélices inseridas em bicamadas lipídicas, a predição de peptídeos sinal foi conduzida pelo SignalP 6.0.

4.6 Análise de motivos e regulação pós-traducional

Para investigar regiões funcionalmente conservadas e potenciais mecanismos de regulação, realizou-se uma análise de motivos utilizando o MEME Suite (BAILEY *et al.*, 2009). Essa ferramenta permitiu identificar sequências curtas altamente conservadas, frequentemente associadas a sítios catalíticos, alvos regulatórios ou elementos estruturais críticos. Em complemento, aplicou-se o *software* NetPhos 3.1 para prever sítios de fosforilação em resíduos de serina, treonina e tirosina (BLOM *et al.*, 2004).

4.7 Modelagem estrutural tridimensional

A modelagem tridimensional das proteínas selecionadas foi realizada com o ColabFold, que integra o algoritmo AlphaFold2 a pipelines otimizados de alinhamento de sequências por MMseqs2 (JUMPER *et al.*, 2021; MIRDITA *et al.*, 2022). As estruturas geradas forneceram modelos de alta confiança, acompanhados dos valores de pLDDT utilizados para avaliar a confiabilidade local de cada região da proteína.

Os modelos tridimensionais foram analisados visualmente nos *softwares* PyMOL. Essa etapa permitiu observar a topologia geral das proteínas, verificar a conformação dos domínios catalíticos e mapear os motivos e resíduos de fosforilação previamente identificados.

4.8 Reprodutibilidade computacional

Todas as etapas computacionais foram realizadas em ambiente Linux padronizado, assegurando consistência entre execuções e facilidade de replicação. Todas as ferramentas utilizadas são de acesso livre ou possuem versões acadêmicas gratuitas, incluindo KOfamScan, TMHMM, SignalP, Protpam, MEME, NetPhos, ColabFold e PyMOL.

Os scripts utilizados para filtragem dos resultados do KOfamScan e extração das sequências FASTA foram armazenados e documentados em repositório próprio. As versões dos *softwares* utilizados foram registradas com o intuito de garantir reprodutibilidade conforme recomendações da literatura em bioinformática computacional (VALENTE, 2013).

5 RESULTADOS E DISCUSSÃO

5.1 Proteínas selecionadas para caracterização funcional e estrutural

Com base na anotação funcional gerada pelo KofamScan, no mapeamento metabólico no KEGG Mapper e na caracterização funcional descritos em Materiais e Métodos, foram priorizadas cinco enzimas diretamente relacionadas à lipogênese em *Elaeis guineensis*. A escolha considerou de forma conjunta a participação em vias centrais da biossíntese de ácidos graxos e de triacilgliceróis, bem como a posição estratégica de cada enzima ao longo do fluxo de carbono, desde a formação de malonil-CoA até a etapa final de estocagem em triacilglicerol (TAG).

O conjunto formado por acetil-CoA carboxilase (ACC, K01961), 3-oxoacil-ACP sintase I (KAS I, K00655), estearoil-ACP dessaturase/SAD/FAD2-like (K03921), glicerol-3-fosfato aciltransferase (GPAT, K13508) e diacilglicerol aciltransferase (DGAT, K13339) representa um recorte funcional da rota lipídica que integra compartimentos plastidiais e membranas do retículo endoplasmático. Essas enzimas cobrem etapas consecutivas de limitação de fluxo, alongamento, dessaturação e montagem de TAG, o que permite discutir a lipogênese do dendezeiro de forma integrada, tanto em nível funcional quanto estrutural.

5.2 ACC (K01961): O ponto de controle limitante

A acetil-CoA carboxilase (ACC), classificada no KEGG sob o código enzimático EC 6.4.1.2, catalisa a carboxilação do acetil-CoA em malonil-CoA, utilizando bicarbonato como doador de carbono e ATP como fonte de energia. Essa reação constitui a etapa de compromisso da via de biossíntese de ácidos graxos, pois o malonil-CoA produzido abastece diretamente o sistema de alongação dependente de ACP no plastídio, servindo como unidade de dois carbonos incorporada ciclicamente ao acil em crescimento (OHLROGGE; BROWSE, 1995).

Do ponto de vista estrutural e funcional, a ACC é uma carboxilase dependente de biotina, composta por módulos que incluem os domínios biotin carboxylase, carboxiltransferase e a região carreadora de biotina, que atua como braço oscilante entre os centros catalíticos. Essa organização modular permite que a enzima execute, de forma coordenada, a ativação do bicarbonato, a transferência do grupo carboxila para a biotina e, em seguida, a carboxilação do acetil-CoA, consolidando a formação de malonil-CoA como intermediário central da lipogênese.

Na rota lipídica do dendezeiro, a ACC ocupa posição estratégica por controlar o fluxo de carbono proveniente de vias geradoras de acetil-CoA, como a glicólise, em direção à síntese de ácidos graxos. A literatura descreve essa etapa como limitante porque variações na atividade da ACC impactam diretamente a disponibilidade de malonil-CoA e, consequentemente, a taxa de elongação das cadeias aciladas (OHLROGGE; BROWSE, 1995). Em espécies oleaginosas, a regulação fina dessa enzima, frequentemente mediada por fosforilação, interação com metabólitos e controle transcricional, está associada à modulação do acúmulo de óleo em tecidos de reserva.

5.3 KAS I (K00655): O módulo de alongamento

A 3-oxoacil-ACP sintase I (KAS I, KO K00655) aparece, no KEGG, associada principalmente à reação EC 2.3.1.51, em que catalisa a condensação entre malonil-ACP e acil-ACP durante a elongação da cadeia de ácidos graxos. Nessa etapa, a enzima promove a formação de 3-oxoacil-ACP com liberação de CO₂, ampliando a cadeia em duas unidades de carbono a cada ciclo. Essa reação integra o sistema de síntese de ácidos graxos do tipo II (FAS II), típico de plastídios de plantas e de procariotos, no qual cada etapa da elongação é executada por uma enzima distinta.

No contexto do metabolismo lipídico vegetal, a KAS I atua de forma reiterada até formar cadeias de 16 e 18 carbonos, que correspondem aos principais ácidos graxos encontrados no óleo de palma. A posição dessa enzima no ciclo elongatório faz com que ela funcione como um módulo estrutural central da rota, conectando diretamente a entrada de malonil-ACP, fornecida pela ACC, à geração dos precursores que serão posteriormente dessaturados e esterificados na Via de Kennedy. Dessa forma, a KAS I não apenas sustenta a

progressão da síntese de ácidos graxos, como também contribui para definir o comprimento de cadeia predominante que irá compor os triacilgliceróis armazenados no mesocarpo de *Elaeis guineensis*.

5.4 SAD/FAD2-like (K03921): qualidade e insaturação

As enzimas agrupadas no KO K03921 incluem dessaturases de ácidos graxos do tipo stearoyl-ACP dessaturase (SAD) e oleate desaturase (FAD2-like), responsáveis por introduzir duplas ligações em cadeias saturadas de 18 carbonos. Em termos bioquímicos, essas dessaturases catalisam reações de oxidorredução dependentes de oxigênio, convertendo estearato (C18:0) em oleato (C18:1) e, em alguns contextos, participando de etapas subsequentes de dessaturação em fosfolipídios de membrana. Essa modificação altera o grau de insaturação dos ácidos graxos e impacta diretamente propriedades como fluidez de membrana, ponto de fusão e estabilidade do óleo acumulado. Em oleaginosas, a atividade de SAD/FAD2-like é decisiva para o perfil de ácidos graxos do triacilglicerol, influenciando a proporção entre frações saturadas e insaturadas e, conseqüentemente, características tecnológicas e nutricionais do óleo produzido (OHLROGGE; BROWSE, 1995).

5.5 GPAT (K13508): O início da via de Kennedy

A glicerol-3-fosfato O-aciltransferase, associada aos códigos enzimáticos EC 2.3.1.15 e EC 2.3.1.198, ocupa uma posição estratégica na interface entre o fornecimento de precursores pela glicólise e a montagem do esqueleto de triacilglicerol na Via de Kennedy. De acordo com a anotação do KEGG, essa enzima catalisa a acilação inicial do glicerol-3-fosfato, formando lisofosfatidato em reação dependente de acil-CoA ou acil-ACP, etapa que inaugura a construção do esqueleto glicerolipídico que posteriormente será convertido em diacilglicerol (DAG) e, por fim, em triacilglicerol (TAG).

Funcionalmente, a GPAT estabelece o ponto de conexão entre o carbono proveniente da glicólise (que gera glicerol-3-fosfato e acetil-CoA) e o pool de ácidos graxos sintetizados

no plastídio, que chegam ao retículo endoplasmático na forma de acil-CoA. Ao acoplar esses dois conjuntos de precursores, a enzima define o início efetivo da rota de esterificação característica da Via de Kennedy, sendo determinante para a taxa com que o glicerol-3-fosfato é canalizado para a formação de TAG em tecidos oleaginosos.

Do ponto de vista fisiológico, a atividade da GPAT influencia diretamente a eficiência com que a planta converte fotoassimilados em lipídios de armazenamento. Em espécies como o dendezeiro, nas quais o mesocarpo acumula grandes quantidades de TAG, uma GPAT funcionalmente competente contribui para sustentar um fluxo contínuo de formação de lisofosfatidato e fosfatidato, intermediários que alimentam tanto a síntese de TAG quanto, em alguns contextos, o metabolismo de fosfolipídios de membrana. Assim, variações na expressão, na atividade catalítica ou na regulação pós-traducional dessa aciltransferase tendem a repercutir no perfil e na quantidade de óleo acumulado, reforçando sua importância como alvo em estudos de lipogênese e melhoramento de oleaginosas.

5.6 DGAT (K13339): A estocagem final

A diacilglicerol O-aciltransferase (DGAT), catalisa a etapa terminal da Via de Kennedy, promovendo a conversão do diacilglicerol (DAG) em triacilglicerol (TAG) por meio da transferência de um grupamento acila a partir de acil-CoA. Essa reação é considerada o ponto de “compromisso final” da rota lipídica, pois direciona irreversivelmente os precursores de glicerol e ácidos graxos para a forma de armazenamento de energia e carbono em tecidos oleaginosos. No contexto do metabolismo vegetal, a DGAT está associada a membranas de retículo endoplasmático e a estruturas envolvidas na formação de gotículas lipídicas, conectando diretamente a síntese de TAG ao seu empacotamento e acúmulo em organelas especializadas.

Do ponto de vista funcional, as enzimas do grupo K13339 integram a família das aciltransferases envolvidas na biossíntese de triacilgliceróis e outros ésteres de acila, participando de vias como “Glycerolipid metabolism” e módulos específicos de “Triacylglycerol biosynthesis” descritos no KEGG. Essa posição metabólica faz com que a DGAT atue como reguladora indireta do teor final de óleo, uma vez que a eficiência da etapa final de esterificação determina o quanto de DAG disponível será efetivamente convertido em

TAG e acumulado. Em espécies oleaginosas, o aumento da atividade ou da expressão de DGAT está frequentemente associado a maior conteúdo de óleo nas sementes ou frutos, reforçando seu papel estratégico em programas de melhoramento voltados à produtividade lipídica.

No dendezeiro, a presença de um KO K13339 de alta confiança indica que o proteoma da espécie contém uma DGAT funcionalmente conservada, capaz de sustentar taxas elevadas de esterificação em tecidos de armazenamento. A combinação entre a posição terminal na Via de Kennedy, a associação a membranas do retículo endoplasmático e a contribuição direta para o acúmulo de TAG justifica a eleição dessa proteína como um dos principais alvos do presente estudo. A caracterização estrutural e regulatória da DGAT fornece subsídios para compreender como o dendezeiro consolida o fluxo de carbono previamente direcionado pela ACC, prolongado pela KAS I, modulado pela SAD/FAD2-like e canalizado para o esqueleto de glicerol pela GPAT, fechando o ciclo biossintético que culmina na estocagem de óleo em *Elaeis guineensis*.

Tabela 1 - Resultados da anotação funcional das enzimas selecionadas pelo KofamScan.

KO_ID	Gene_ID	KO_Function	HMM_Score	Threshold	E_value
K00655	XP_010919015.2	1-acyl-sn-glycerol-3-phosphate acyltransferase	254.5	120.03	4.1999999999999997e-75
K01961	XP_010927174.1	acetyl-CoA carboxylase, biotin carboxylase subunit	874.0	734.07	1.5e-262
K03921	NP_001391894.1	acyl-[acyl-carrier protein] desaturase	717.2	163.7	2.6e-215
K13339	XP_010907043.2	peroxin-6	1000.8	611.47	4.1e-300
K13508	XP_010923095.2	glycerol-3-phosphate acyltransferase	792.9	205.77	4.8000000000000004e-238

Fonte: Autor, 2025.

5.7 Análise de via metabólicas com KEGG Mapper

A reconstrução de vias metabólicas com o KEGG Mapper, a partir da lista de KOs válidos obtida pelo KOfamScan, permitiu avaliar a representatividade funcional do proteoma analisado e identificar rotas com maior cobertura enzimática. Entre os mapas resultantes, duas vias se destacaram por concentrarem a maior parte dos KOs anotados: a biossíntese de ácidos graxos (map00061), o metabolismo de glicerolipídios (map00561).

Esses mapas reforçam a coerência metabólica do conjunto de dados, indicando que o tecido analisado apresenta forte direcionamento de carbono para a lipogênese e para o acúmulo de triacilgliceróis.

Na via map00061, observou-se ampla cobertura do ciclo plastidial de alongação, com presença de enzimas estruturais da rota, como FabD (malonil-CoA:ACP transacilase), FabG (3-oxoacil-ACP redutase), FabZ/FabA (desidratases), FabI (enoil-ACP redutase) e demais etapas do ciclo de alongamento. Esse padrão é típico de tecidos oleaginosos, nos quais o fluxo de carbono é canalizado para a formação de ácidos graxos saturados e insaturados de cadeia longa (C16 e C18), que constituem a base estrutural do óleo vegetal. Dentro desse conjunto, a acetil-CoA carboxilase (ACC) aparece como enzima determinante do passo de entrada, por catalisar a formação de malonil-CoA, molécula que condiciona a velocidade global de síntese de ácidos graxos. A literatura clássica sobre lipogênese em oleaginosas descreve a ACC como a etapa limitante que controla a disponibilidade de unidades de dois carbonos para o ciclo elongatório, sendo portanto decisiva para o rendimento lipídico final (BELTRÃO; OLIVEIRA, 2007).

Além disso, o mapa evidencia a participação da 3-oxoacil-ACP sintase I (KAS I) no núcleo do ciclo elongatório. Essa enzima atua nas condensações repetitivas entre malonil-ACP e acil-ACP, sustentando a extensão progressiva da cadeia até os comprimentos C16 e C18, majoritários no óleo de palma. Em sequência, a presença da dessaturase SAD/FAD2-like sinaliza a capacidade de conversão de cadeias saturadas em insaturadas, processo que define o grau de insaturação do óleo. As dessaturases que introduzem duplas ligações, especialmente aquelas relacionadas à formação de oleato e linoleato, são reconhecidas como componentes essenciais para o perfil lipídico de sementes e frutos oleaginosos, atuando em associação a fosfolipídios de membrana durante o remodelamento dos ácidos graxos (BELTRÃO; OLIVEIRA, 2007).

Para deixar explícita a posição das enzimas selecionadas nesta rota, a inspeção do mapa com apenas as proteínas-chave mostra que: a ACC está localizada no passo inicial de carboxilação do acetil-CoA (EC 6.4.1.2), imediatamente antes da formação de malonil-CoA; a KAS I aparece nas reações centrais do ciclo elongatório de acil-ACP; e a SAD/FAD2-like

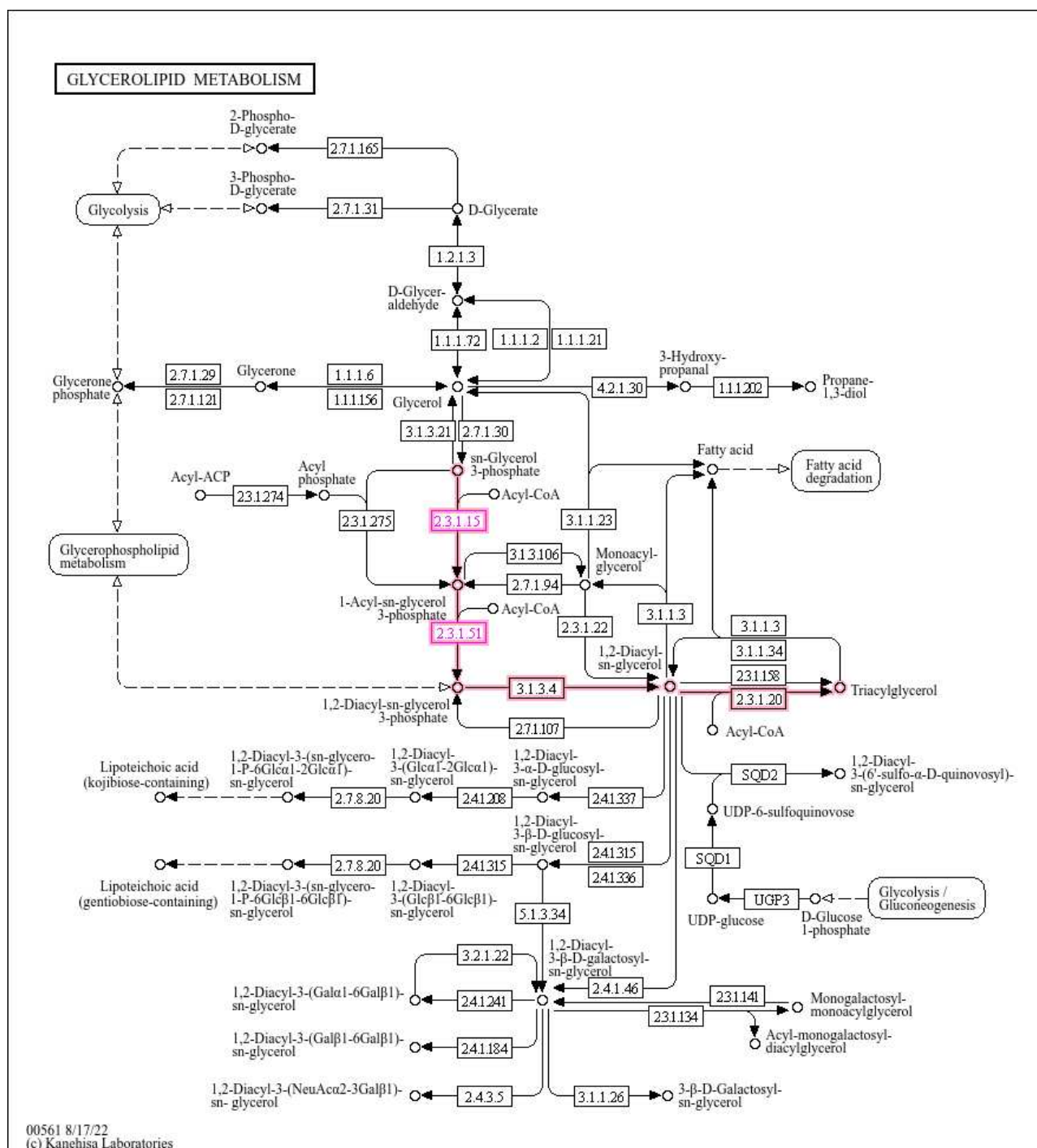
[illegible]

Os números dentro das caixas representam o código da enzima (EC) envolvida na via sendo que cada caixa destacada de vermelho e rosa correspondem às enzimas identificadas e selecionadas para o trabalho. **Fonte:** Autor, 2025.

Em paralelo, a via map00561 evidenciou forte cobertura do metabolismo de glicerolipídios, especialmente nos passos da Via de Kennedy, responsável pela montagem sequencial do triacilglicerol (TAG). O conjunto de KOs válidos marcou enzimas de acilação e remodelamento de glicerol-3-fosfato, incluindo aciltransferases iniciais e enzimas relacionadas à formação de DAG e TAG. Dentro dessa rota, a glicerol-3-fosfato aciltransferase (GPAT; KO K13508) foi identificada no primeiro passo de acilação do glicerol-3-fosfato, formando lisofosfatidato. Já a diacilglicerol O-aciltransferase (DGAT; KO K13339) aparece na etapa terminal, catalisando a conversão de DAG em TAG. Esta reação é descrita como o “compromisso final” da rota lipídica, pois direciona irreversivelmente os precursores para armazenamento em forma de óleo, ocorrendo em complexos enzimáticos associados ao retículo endoplasmático (BELTRÃO; OLIVEIRA, 2007).

O mapa específico com as proteínas selecionadas reforça essa organização: a GPAT está posicionada no início da Via de Kennedy (EC 2.3.1.15), enquanto a DGAT está no último passo de acilação (EC 2.3.1.20). Além delas, a rota apresentou cobertura de enzimas intermediárias como LPAT e fosfatases de fosfatidato, o que indica que o caminho de montagens sucessivas até TAG encontra-se funcionalmente representado no proteoma avaliado.

Figura 6 - Map00561 com as enzimas DGAT e GPAT em evidência.

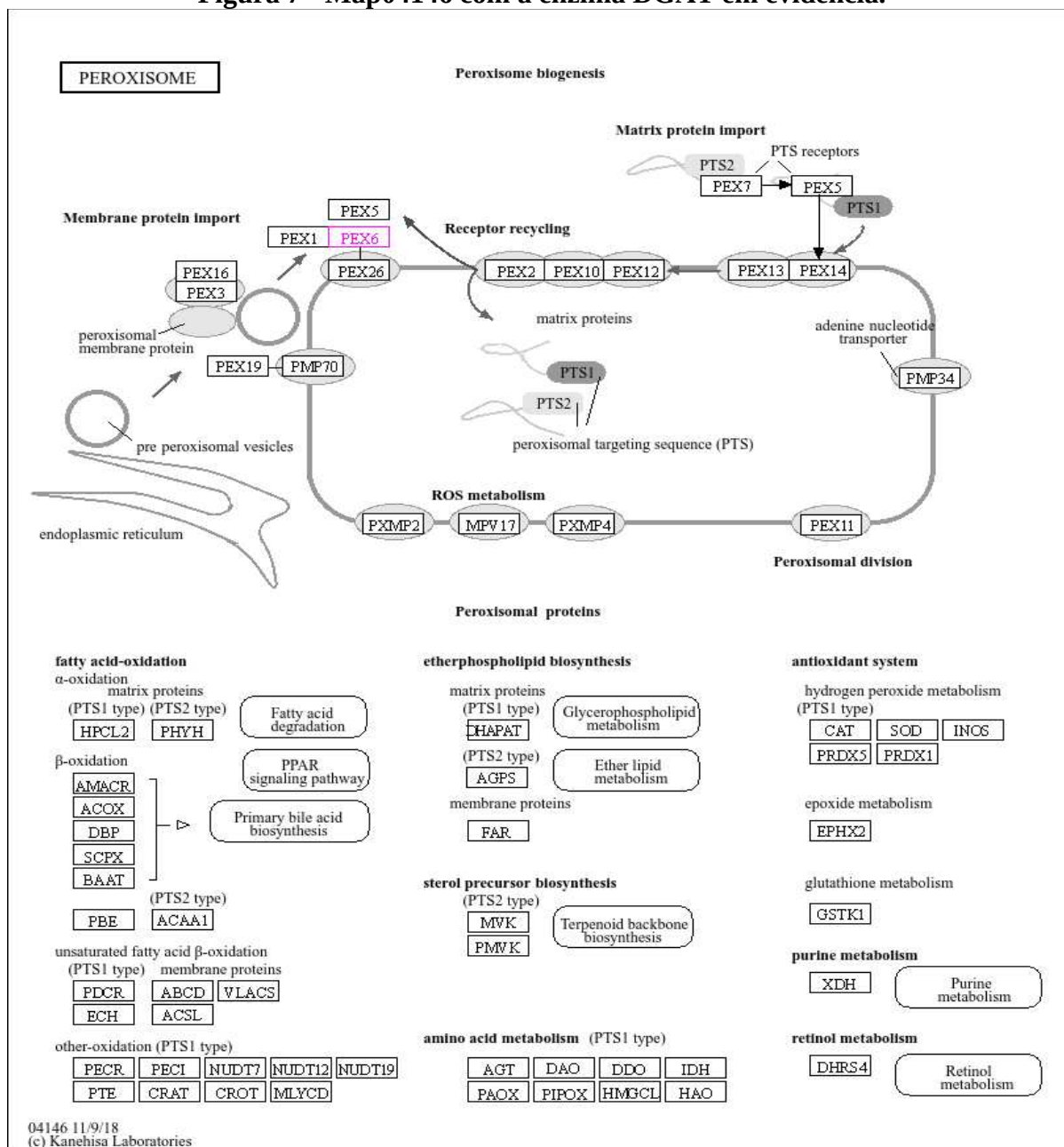


Os números dentro das caixas representam o código da enzima (EC) envolvida na via sendo que cada caixa destacada de vermelho e rosa correspondem às enzimas identificadas e selecionadas para o trabalho. **Fonte:** Autor, 2025.

Por fim, um achado complementar foi a identificação do KO K13339 (DGAT) também em um mapa relacionado a peroxissomos. Essa observação não invalida a seleção realizada. Em plantas, peroxissomos participam ativamente do metabolismo lipídico,

sobretudo na β -oxidação e no controle do balanço entre síntese e degradação de ácidos graxos. A ocorrência de KOs lipídicos em mapas peroxissomais reflete essa integração entre compartimentos e reforça que o metabolismo de TAG em oleaginosas depende de rotas anabólicas no retículo endoplasmático e rotas catabólicas ou de reciclagem associadas a peroxissomos (BELTRÃO; OLIVEIRA, 2007). Dessa forma, a presença de DGAT em diferentes contextos metabólicos é interpretada como evidência de relevância funcional ampla para o manejo de lipídios no dendezeiro.

Figura 7 - Map04146 com a enzima DGAT em evidência.



Os números dentro das caixas representam o código da enzima (EC) envolvida na via sendo que cada caixa destacada de vermelho e rosa correspondem às enzimas identificadas e selecionadas para o trabalho. **Fonte:** Autor, 2025.

Em conjunto, os mapas KEGG confirmam que a biossíntese de ácidos graxos e a Via de Kennedy constituem os eixos mais ativos e biologicamente relevantes no conjunto analisado. A alta cobertura do ciclo plastidial, associada à forte representação dos passos de montagem de TAG, justifica a escolha de ACC, KAS I, SAD/FAD2-like, GPAT e DGAT

como proteínas-chave para caracterização estrutural e regulatória, uma vez que elas cobrem desde a entrada do carbono na rota lipídica até a etapa final de armazenamento do óleo.

5.8 Caracterização estrutural e regulatória das cinco proteínas-chave

A caracterização estrutural e regulatória das cinco proteínas selecionadas foi conduzida para verificar propriedades físico-químicas, indícios de localização celular, identidade enzimática e potencial de regulação pós-traducional. Essa etapa foi essencial porque a seleção inicial foi baseada em ortologia KEGG de alta confiança, mas a confirmação funcional exige evidências complementares de estrutura e conservação de domínios catalíticos. Assim, as sequências correspondentes aos KOs K01961, K00655, K03921, K13508 e K13339 foram avaliadas em conjunto com ProtParam, TMHMM, SignalP, DeepLoc2, CDD-Search, NetPhos e MEME Suite

Tabela 2: Propriedades físico-químicas previstas pelo ProtParam para as cinco enzimas selecionadas da lipogênese em *Elaeis guineensis*.

Proteína (KO)	Tamanho (aa)	PM (kDa)	pI Teórico	II (Instabilidade)	GRAVY (Hidropaticidade)	Estabilidade
K00655 (KAS I)	335	37	10	49	216	Instável
K01961 (ACC)	535	58	7	41	-133	Instável
K03921 (SAD/FAD2-like)	393	45	6	44	-496	Instável
K13339 (DGAT)	946	103	6	43	-159	Instável
K13508 (GPAT)	528	58	9	48	56	Instável

Fonte: Autor, 2025.

A tabela apresenta, para cada proteína (KO/enzima), o tamanho em resíduos de aminoácidos, o peso molecular (PM, em kDa), o ponto isoelétrico teórico (pI), o índice de instabilidade (II), o índice GRAVY (hidropaticidade média) e a classificação global de estabilidade, permitindo comparar o perfil estrutural básico de ACC, KAS I, SAD/FAD2-like, GPAT e DGAT.

Tabela 3 - Parâmetros físico-químicos e predição de localização (DeepLoc2, TMHMM e SignalP).

Proteína	Função	Localização (DeepLoc2)	Peptídeo Sinal (SignalP)	TMs (TMHMM)	GRAVY (Hidrofobicidade)
ACC	Entrada (Limitante)	Cytoplasm/ Mitochondria	Não	0	-138
KAS I	Elongação	Mitochondria	Não	0	216
SAD	Dessaturação	Cytoplasm	Não	0	-496
GPAT	Montagem Esqueleto	Mitochondria	Não	0	56
DGAT	Finalização TAG	Mitochondria	Não	0	-159

Fonte: Autor, 2025.

Os resultados consolidados indicaram que nenhuma das proteínas apresenta peptídeo sinal clássico segundo o SignalP, o que sugere ausência de direcionamento para secreção ou via endomembranar dependente de sinal N-terminal (TEUFEL *et al.*, 2022). De modo consistente, o TMHMM não detectou hélices transmembranares em nenhuma sequência, inferindo proteínas solúveis ou sem trechos hidrofóbicos longos característicos de proteínas integrais de membrana (KROGH *et al.*, 2001). No ProtParam, os valores de GRAVY reforçaram esse padrão, com hidropaticidade global negativa para ACC, SAD, DGAT e próxima de zero para GPAT, indicando predominância de regiões hidrofílicas e provável solubilidade em meio aquoso (GASTEIGER *et al.*, 2005). A única exceção parcial foi KAS I, com GRAVY levemente positivo, o que pode refletir regiões internas apolares típicas de enzimas elongases, mas ainda insuficiente para caracterizar associação obrigatória à membrana.

Apesar dessa coerência físico-química, as predições de localização subcelular pelo DeepLoc2 apontaram citoplasma ou mitocôndria para todas as proteínas. Esse resultado diverge do modelo biológico esperado para enzimas lipogênicas de plantas, nas quais ACC, KAS I e SAD atuam predominantemente no plastídio e GPAT e DGAT são associadas ao retículo endoplasmático (THUMULURI *et al.*, 2022). Como preditores automáticos podem errar em proteínas vegetais multidomínio ou com sinais de trânsito não canônicos, esses dados foram interpretados como indícios iniciais e não como prova definitiva de localização. Por essa razão, foi necessária a validação da identidade pela presença de domínios catalíticos conservados.

Tabela 4 - Domínios conservados identificados nas enzimas-chave da lipogênese em *Elaeis guineensis* por CDD-Search.

Proteína (ID)	Função Principal	Domínios Conservados (CDD)
XP_010927174.1	ACC (Carboxilase)	Biotin_carboxylase; AccC; AccC_N_core
XP_010919015.2	KAS I (Sintase)	KAS (Ketoacyl_synthase); β -Ketoacyl_Synthase
NP_001391894.1	SAD (Dessaturase)	FA_desaturase (Fatty acid desaturase)
XP_010923095.2	GPAT (Aciltransferase)	Glycerol_acyltransferase; MBOAT_GPAT_dom
XP_010907043.2	DGAT (Aciltransferase)	Diacylglycerol_acyltransferase; DGAT_ACAT

Fonte: Autor, 2025.

A análise pelo CDD-Search confirmou que todas as proteínas contêm os módulos funcionais característicos das enzimas atribuídas pelo KOfamScan, consolidando a identidade enzimática do conjunto (MARCHLER-BAUER *et al.*, 2011). A ACC (K01961) apresentou domínios de biotin carboxilase e regiões AccC típicas de carboxilases dependentes de biotina, validando sua função como passo limitante de entrada de carbono na lipogênese. A KAS I (K00655) exibiu domínios do tipo cetooxil-sintase, compatíveis com sua atuação central no ciclo elongatório de ácidos graxos. Para SAD/FAD2-like (K03921), a detecção do domínio FA_desaturase confirmou a capacidade catalítica de inserir insaturações em cadeias C18, processo determinante para o perfil final do óleo. Já GPAT (K13508) e DGAT (K13339) apresentaram domínios de aciltransferases, incluindo assinaturas típicas de MBOAT/GPAT e DGAT/ACAT, respectivamente, corroborando sua participação direta na esterificação sequencial de glicerol-3-fosfato e na etapa terminal de síntese de TAG. Dessa forma, a evidência estrutural obtida pelo CDD resolveu as contradições de localização apontadas por DeepLoc2 e sustentou a seleção das cinco enzimas como representantes funcionais das etapas críticas da rota lipídica.

A regulação pós-traducional por fosforilação foi então investigada, pois esse mecanismo é central no controle metabólico vegetal e pode explicar diferenças de fluxo na biossíntese de óleo.

Tabela 5 - Sítios de fosforilação preditos nas cinco enzimas-chave da lipogênese em *Elaeis guineensis* pelo NetPhos 3.1.

Enzima	S_Total	T_Total	Y_Total	Total_Sites
ACC	18	10	8	36
GPAT	32	18	9	59
KAS 1	21	9	0	30
DGAT	77	13	5	95
SAD	21	12	4	37

Fonte: Autor, 2025.

Segundo a tabela 4, a predição por NetPhos evidenciou alto potencial regulatório para todas as proteínas, com sítios distribuídos em serina, treonina e tirosina acima do limiar de 0,5, que foi o limiar de confiança adotado (LI *et al.*, 2018). Entre as cinco proteínas, a DGAT (K13339) apresentou o maior número total de sítios, sugerindo um forte potencial de regulação fina na etapa terminal de síntese e estocagem de triacilgliceróis. Em seguida, a GPAT (K13508) exibiu uma densidade elevada de sítios, compatível com seu papel como passo inicial da Via de Kennedy. A ACC (K01961), KAS I (K00655) e SAD (K03921) apresentaram números intermediários de sítios, o que indica que, embora também sejam alvos potenciais de fosforilação, a regulação pós-traducional tende a concentrar-se com maior intensidade nas etapas de montagem e finalização do TAG. Esse padrão é coerente com a lógica metabólica em oleaginosas, na qual o controle do fluxo de carbono não depende apenas do passo de entrada na rota lipídica, mas também de pontos de decisão associados ao acúmulo efetivo de óleo.

Em seguida, a análise de motivos conservados foi aplicada para identificar regiões curtas altamente preservadas associadas a função catalítica ou a pontos de interação regulatória.

Tabela 6 - Motivos conservados compartilhados entre as cinco enzimas-chave da lipogênese em *Elaeis guineensis*, identificados pelo MEME Suite.

Motif	Largura (aa)	Valor E	KOs / Proteínas Encontradas
MEME-4	19	1.5×10 ²	K01961 (ACC) e K03921 (SAD)
MEME-1	7	9.4×10 ⁰	K13339 (DGAT) e K13508 (GPAT)
MEME-5	8	4.3×10 ²	K01961 (ACC) e K13508 (GPAT)
MEME-2	9	1.4×10 ¹	K03921 (SAD), K13339 (DGAT), K00655 (KAS I), K13508 (GPAT)
MEME-3	7	5.5×10 ¹	K03921 (SAD) e K01961 (ACC)
MEME-6	6	6.2×10 ²	K13508 (GPAT) e K13339 (DGAT)
MEME-7	6	1.7×10 ²	K13339 (DGAT) e K03921 (SAD)
MEME-8	6	6.5×10 ²	K00655 (KAS I) e K13508 (GPAT)
MEME-9	6	6.9×10 ²	K01961 (ACC) e K13339 (DGAT)
MEME-10	7	7.4×10 ²	K13508 (GPAT) e K00655 (KAS I)

Fonte: Autor, 2025.

O MEME identificou dez motivos, destacando padrões conservados em diferentes combinações de proteínas, o que fortalece a interpretação de conservação funcional (BAILEY *et al.*, 2009). Entre os motivos mais informativos, observou-se um padrão compartilhado por ACC e SAD, sugerindo conservação de regiões associadas ao ambiente plastidial e possivelmente a elementos comuns de regulação ou reconhecimento de cofatores. Outro motif foi compartilhado por GPAT e DGAT, reforçando assinaturas típicas de aciltransferases envolvidas na montagem final de TAG. Notavelmente, um motivo amplo foi encontrado em quatro das cinco proteínas, incluindo enzimas plastidiais e do retículo endoplasmático. Esse achado sugere a presença de uma assinatura funcional ou regulatória transversal à rota lipídica, potencialmente relacionada à ligação de cofatores universais, à estabilidade estrutural ou a interações coordenadas entre etapas sequenciais. Por fim, a presença de motivos compartilhado entre ACC e GPAT é particularmente relevante, pois conecta diretamente o ponto de entrada de carbono no plastídio ao início da esterificação no retículo endoplasmático, sugerindo possibilidade de *crosstalk* regulatório entre compartimentos.

Em síntese, os resultados estruturais e regulatórios mostram que, embora os preditores de localização e de associação à membrana não tenham indicado plastídio ou retículo endoplasmático para todas as proteínas, a validação por domínios conservados confirmou

inequivocamente a identidade funcional de ACC, KAS I, SAD, GPAT e DGAT. As propriedades físico-químicas indicaram proteínas predominantemente solúveis e intracelulares, coerentes com enzimas catalíticas de metabolismo primário, enquanto a ausência de peptídeo sinal e de hélices transmembranares reforça que nenhuma delas segue via secretória clássica. A análise de fosforilação revelou maior potencial regulatório em ACC e GPAT, destacando o controle fino do fluxo de carbono no início da lipogênese e da Via de Kennedy. Por sua vez, os motivos conservados reforçaram assinaturas catalíticas e sugeriram regulação coordenada entre etapas plastidiais e do retículo endoplasmático. Assim, essa etapa consolidada, em nível estrutural e regulatório, a escolha das cinco proteínas como os melhores marcadores *in silico* para interpretar a biossíntese de óleo em *Elaeis guineensis*.

5.9 Modelagem tridimensional das cinco proteínas chave

A modelagem tridimensional constituiu a etapa de integração final dos resultados obtidos a partir da anotação funcional, da caracterização estrutural primária e das análises de regulação pós traducional. As sequências das cinco enzimas selecionadas, previamente validadas por KOfamScan e CDD Search, foram submetidas à predição estrutural no ColabFold, que implementa o algoritmo AlphaFold2 e fornece modelos com estimativas de confiança residuais via índice pLDDT (JUMPER *et al.*, 2021; MIRDITA *et al.*, 2022).

Os modelos tridimensionais foram visualizados e analisados no *software* de código aberto PyMOL (The PyMOL Molecular Graphics System, Version 3.0 Schrödinger, LLC.), com coloração por pLDDT para destacar regiões de alta e baixa confiança e com mapeamento explícito dos domínios conservados, motivos identificados pelo MEME e sítios de fosforilação preditos pelo NetPhos. Essa etapa permitiu verificar se os elementos funcionais descritos em sequência se organizam em arquiteturas compatíveis com o que é conhecido para carboxilases, sintases, dessaturases e aciltransferases envolvidas na biossíntese de lipídios em plantas.

5.10 Enzimas plastidiais: ACC, KAS I e SAD/FAD2 like

Nos modelos da ACC e da KAS I observou-se um núcleo globular bem organizado, com valores de pLDDT predominantemente elevados no corpo do domínio catalítico e quedas localizadas em alças de superfície, padrão típico de enzimas citosólicas ou plastidiais com domínios bem definidos (JUMPER *et al.*, 2021). A ACC apresentou um domínio central compatível com as carboxilases dependentes de biotina, em consonância com os domínios biotin_carboxylase e AccC identificados no CDD-Search. Na representação tridimensional, esses domínios formam uma cavidade catalítica contínua, em torno da qual se distribuem vários dos sítios de fosforilação preditos, especialmente aqueles agrupados em regiões de superfície.

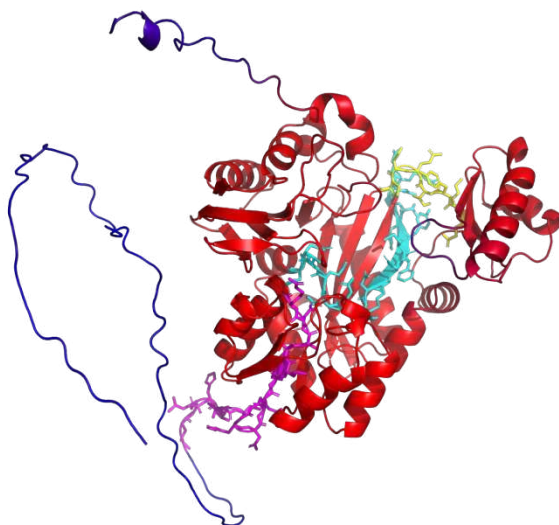
Figura 8 - Modelo 3D da ACC colorido por pLDDT, com o domínio de biotin carboxylase destacado.



Fonte: Autor, 2025.

Modelo tridimensional da ACC com coloração por pLDDT. O domínio catalítico principal, compreendido entre os resíduos 70–526, está realçado em laranja, correspondendo à região de maior confiança e relevância funcional. As extremidades N- e C-terminais permanecem na escala original de pLDDT, em tons mais azulados, indicando menor confiança na predição estrutural nessas regiões.

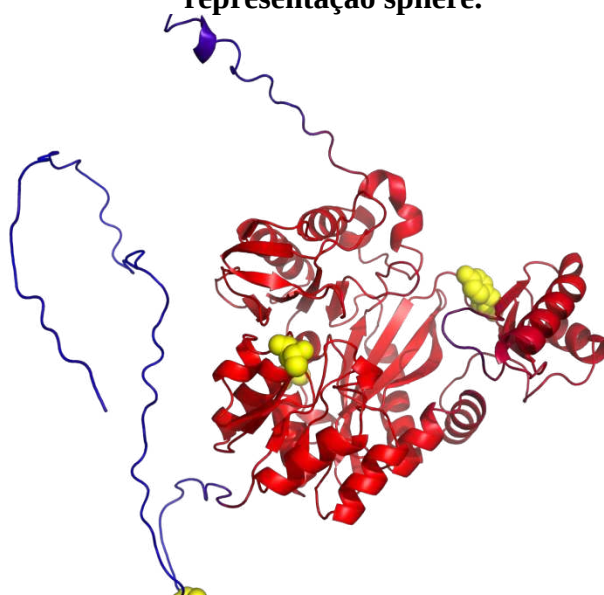
Figura 9 - ACC com motivos selecionados em representação *sticks* e identificação dos resíduos.



Fonte: Autor, 2025.

Na figura 9 estão em destaque os seguintes motivos, em magenta o MEME – 9 com sua região entre os resíduos 60 – 77, em amarelo o MEME – 3 com sua região entre os resíduos 267 – 273 e por ultimo em amarelo o MEME – 4 com sua região compreendida entre os resíduos 346 – 364.

Figura 10 - ACC com sítios de fosforilação selecionados em representação *sphere*.



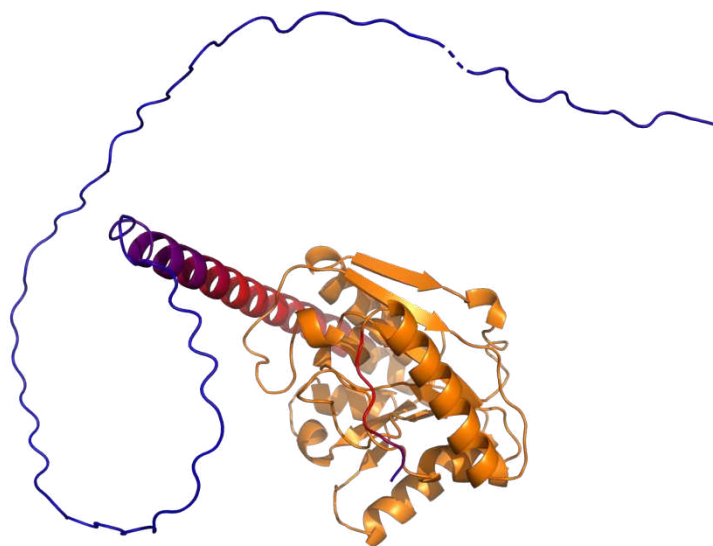
Fonte: Autor, 2025.

A distribuição espacial dos sítios de fosforilação escolhidos para visualização na ACC demonstra que esses resíduos se concentram em alças expostas, posicionadas na periferia do

domínio catalítico. Essa organização é coerente com o papel regulatório já descrito para ACC em outras espécies, nas quais a fosforilação em regiões de superfície altera a conformação relativa entre subdomínios e modula a atividade enzimática sem comprometer o núcleo catalítico (HARWOOD; RUSSELL, 1984; OHLROGGE; BROWSE, 1995). No contexto de *Elaeis guineensis*, essa configuração reforça a interpretação de que a ACC pode ser alvo de regulação fina em resposta ao estado energético celular.

Para a KAS I, o modelo tridimensional evidenciou o dobramento típico de sintases de 3 cetoacil ACP, organizado em torno do domínio ketoacyl synthase descrito no CDD Search. Os motivos selecionados na análise com o MEME se localizaram próximos à região central da cavidade catalítica, muitas vezes em hélices e folhas β internas, o que é compatível com a função de estabilizar o sítio ativo e de coordenar a ligação do ACP e do substrato acilado. Os sítios de fosforilação priorizados para visualização concentraram-se em alças de superfície posicionadas nas bordas do domínio, sugerindo que a modulação da KAS I por quinases, caso ocorra, tende a atuar sobre a dinâmica da entrada e saída de substratos, mais do que sobre o núcleo catalítico propriamente dito.

Figura 11 - Modelo 3D da KAS I colorido por pLDDT, com o domínio ketoacyl synthase destacado.

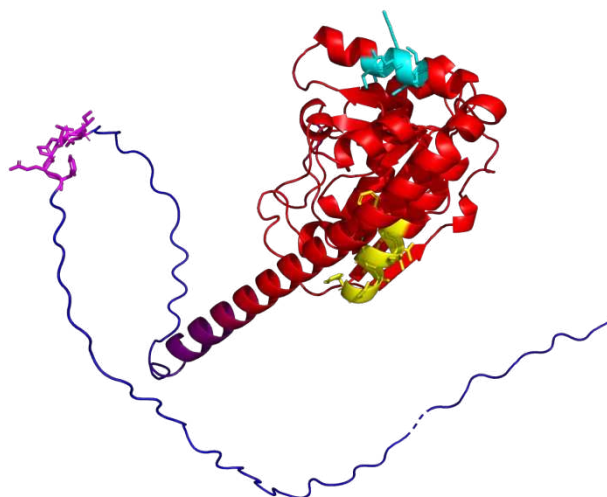


Fonte: Autor, 2025.

Modelo tridimensional da KAS I com coloração por pLDDT. O domínio catalítico, compreendido entre os resíduos 115–328, foi realçado em laranja, correspondendo à região de maior interesse funcional. As porções N- e C-terminais, mantidas na escala original de

pLDDT, aparecem em tons mais azulados, indicando menor confiança na predição estrutural nessas extremidades.

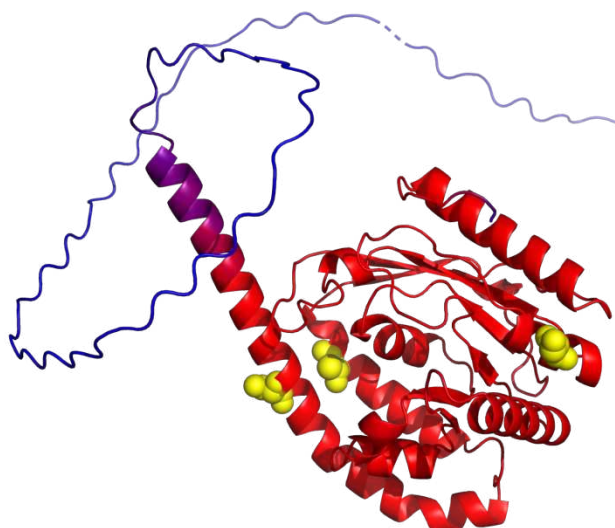
Figura 12 - KAS I com motivos selecionados em representação *sticks*.



Fonte: Autor, 2025.

Na figura 12 estão em destaque os seguintes motivos da enzima KAS 1, em azul ciano está o MEME – 10 com sua região entre os resíduos 48 – 54 e em amarelo o MEME – 2 com sua região compreendida entre os resíduos 140 – 148.

Figura 13 - KAS I com sítios de fosforilação selecionados em representação *sphere*.

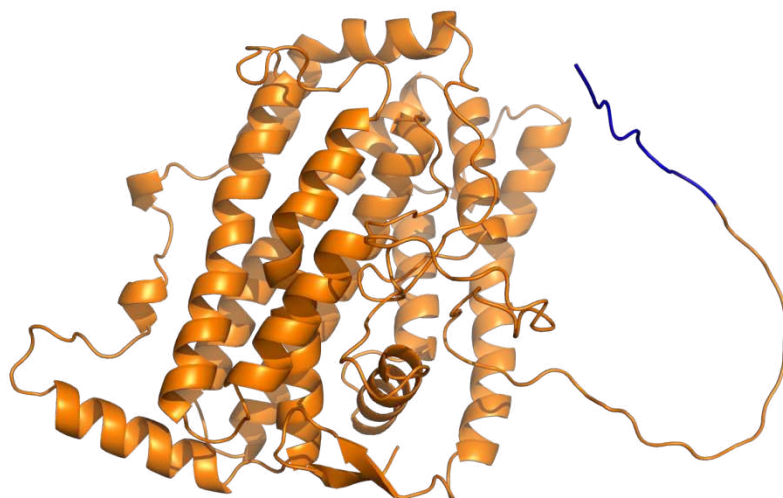


Fonte: Autor, 2025.

Na Figura 13, a estrutura tridimensional da KAS I é apresentada com o domínio catalítico em vermelho (regiões de alta confiança pelo pLDDT) e as extremidades menos confiáveis do modelo em azul. Os sítios de fosforilação selecionados para destaque estão representados em esferas amarelas e se distribuem principalmente em alças expostas na superfície do domínio catalítico, ocupando posições que circundam o provável canal de acesso do substrato ao sítio ativo. Essa disposição espacial indica que modificações por fosforilação nessas regiões têm potencial para alterar o microambiente ao redor do centro catalítico ou a estabilidade conformacional da proteína, atuando como pontos de ajuste fino da atividade elongatória da KAS I na biossíntese de ácidos graxos.

No modelo da SAD/FAD2 like, o domínio FA_desaturase identificado no CDD Search organizou-se como um núcleo compacto, cercado por alças flexíveis que apresentaram valores de pLDDT mais baixos. Essa arquitetura é consistente com dessaturases de ácidos graxos, nas quais o núcleo coordena o cofator metálico e o substrato hidrofóbico, enquanto regiões periféricas contribuem para o reconhecimento de membranas e interação com outros componentes do complexo lipídico (SHANKLIN; CAHOON, 1998). Os motivos conservados selecionados para visualização se posicionaram em regiões internas do núcleo, sugerindo participação direta na coordenação de cofatores ou na estabilização da cavidade hidrofóbica. Já os sítios de fosforilação escolhidos situaram-se em alças voltadas para o solvente, em locais potencialmente acessíveis a quinases, o que reforça a hipótese de regulação pós-traducional do grau de insaturação dos ácidos graxos.

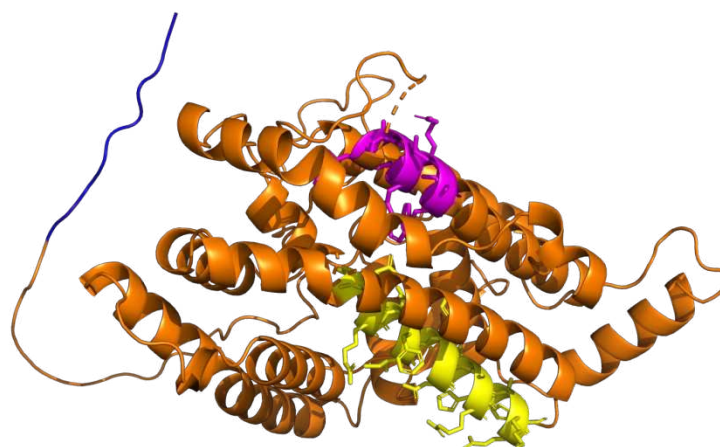
Figura 14 - Modelo 3D da SAD/FAD2 like colorido por pLDDT, com o domínio FA_desaturase destacado.



Fonte: Autor, 2025.

O domínio catalítico da proteína, compreendido entre os resíduos 11 e 393, está representado em laranja, formando o núcleo de hélices característico das dessaturases de ácidos graxos. A região terminal com menor confiança de predição pelo pLDDT aparece em azul, destacando um segmento estruturalmente mais flexível e menos conservado.

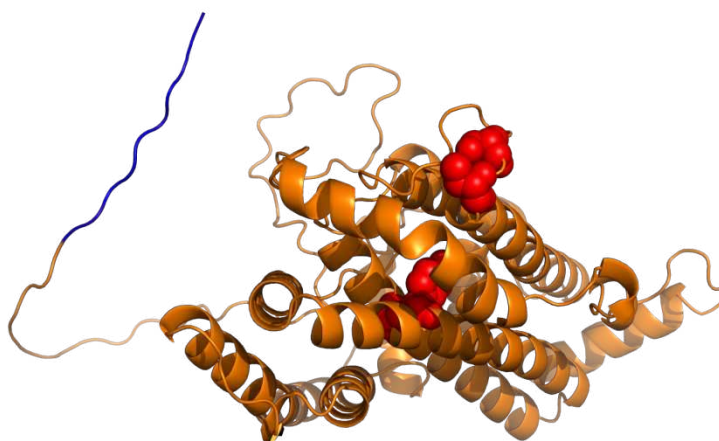
Figura 15 - SAD/FAD2 like com motivos selecionados em sticks.



Fonte: Autor, 2025.

Na figura 15 estão em destaque os seguintes motivos da enzima SAD, em magenta está o MEME – 2 com sua região entre os resíduos 162 – 170 e em amarelo o MEME – 4 com sua região compreendida entre os resíduos 225 – 243.

Figura 16 - SAD/FAD2 like com sítios de fosforilação selecionados em *spheres*.



Fonte: Autor, 2025.

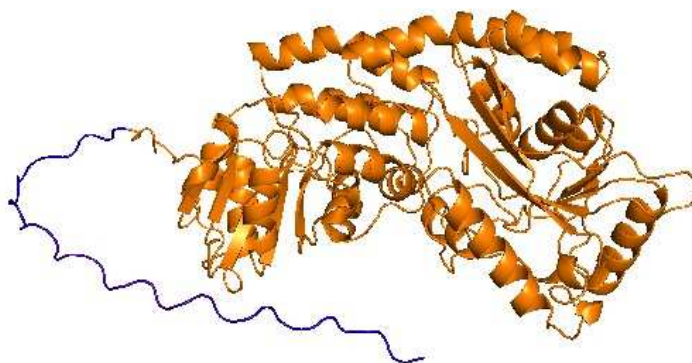
Os sítios de fosforilação preditos para SAD/FAD2-like, destacados em esferas vermelhas na figura 16, localizam-se ao longo do domínio FA_desaturase (em laranja), distribuídos em hélices e alças expostas na superfície da proteína. Muitos desses resíduos circundam a região central do dobramento, próxima ao provável sítio ativo e às hélices envolvidas na acomodação da cadeia acil. Essa organização espacial indica que eventos de fosforilação nessas posições podem alterar o ambiente eletrostático e a flexibilidade local do domínio catalítico, modulando diretamente a eficiência da dessaturação de ácidos graxos e, consequentemente, o grau de insaturação dos triacilgliceróis acumulados no dendezeiro.

5.11 Enzimas da Via de Kennedy: GPAT e DGAT

Os modelos de GPAT e DGAT apresentaram dobras compatíveis com aciltransferases envolvidas na montagem de triacilgliceróis, com núcleos organizados em torno dos domínios Glycerol_acyltransferase/MBOAT e DGAT/ACAT previamente inferidos por anotação funcional. Apesar de TMHMM e SignalP não terem identificado segmentos transmembrana ou peptídeos sinal clássicos, os modelos tridimensionais revelaram superfícies extensas com bolsões hidrofóbicos internos, compatíveis com a interação com membranas ou com substratos apolares. Essa observação sugere que a associação ao retículo endoplasmático pode depender de regiões anfipáticas e de interação proteína membrana não detectadas por preditores lineares, o que tem sido descrito em aciltransferases vegetais (LARDIZÁBAL *et al.*, 2001; CAO; HUANG, 2016).

Na GPAT, os motivos conservados visualizados em PyMOL se localizaram na vizinhança da cavidade onde se espera a ligação do glicerol 3 fosfato e do acil doador, reforçando o papel desses motivos como assinaturas catalíticas e de reconhecimento de substrato. O conjunto reduzido de sítios de fosforilação escolhido para análise tridimensional concentrou-se em alças próximas à superfície do domínio, frequentemente em regiões que fazem a transição entre blocos estruturais mais rígidos. Essa organização espacial é compatível com um modelo em que a fosforilação modularia a abertura da cavidade de acilação ou a afinidade por parceiros de membrana, ajustando o fluxo de carbono no início da Via de Kennedy.

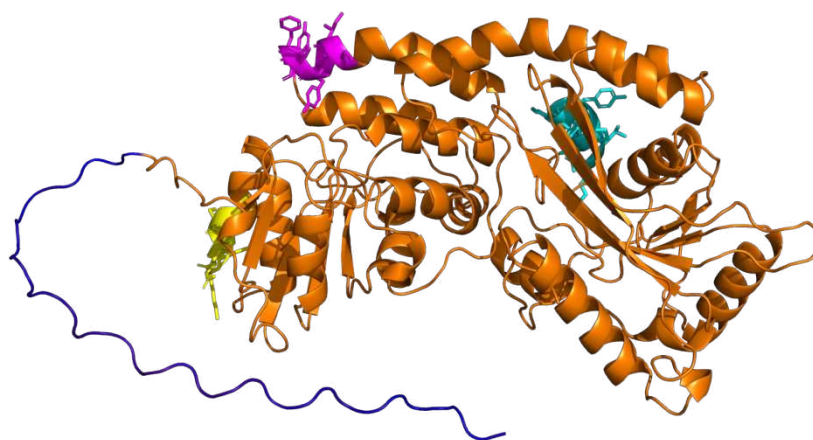
Figura 17 - Modelo 3D da GPAT colorido por pLDDT, com domínio aciltransferase destacado.



Fonte: Autor, 2025.

O domínio catalítico da GPAT está compreendido entre os resíduos 36–528, representado em laranja na figura. As regiões N- e C-terminais fora desse intervalo aparecem como cadeias mais flexíveis em azul, sugerindo segmentos desordenados que não fazem parte do núcleo enzimático principal.

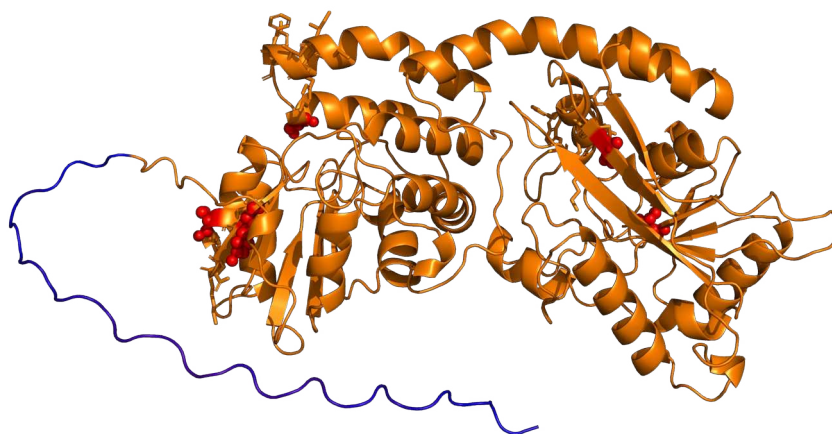
Figura 18 - GPAT com motivos selecionados em *sticks*.



Fonte: Autor, 2025.

Na figura 18, estão em destaque os seguintes motivos da enzima GPAT, em azul ciano está o MEME – 2 com sua região entre os resíduos 448 – 456 e em amarelo o MEME – 5 com sua região compreendida entre os resíduos 236 – 243 por ultimo o MEME – 1 com sua região compreendida entre os resíduos 99 -105.

Figura 19 - GPAT com sítios de fosforilação selecionados em *spheres*.



Fonte: Autor, 2025.

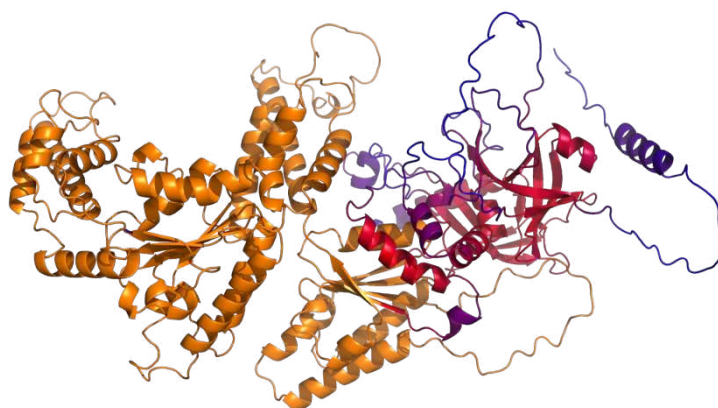
Na figura 19, os sítios de fosforilação previstos (esferas vermelhas na figura) formam dois agrupamentos principais ao longo do domínio catalítico: um na região N-terminal, próximo ao provável sítio de ligação ao glicerol-3-fosfato, e outro em hélices e alças expostas na porção C-terminal. A localização desses resíduos em superfícies acessíveis e adjacentes ao sulco catalítico indica que quinases podem modular diretamente a conformação e a dinâmica da enzima, ajustando de forma fina a etapa inicial da Via de Kennedy, em que o glicerol-3-fosfato é acilado e direcionado para a formação de triacilgliceróis no dendezeiro.

Para a DGAT, o modelo de AlphaFold2 mostrou um núcleo bem definido e valores de pLDDT elevados ao longo da maior parte da cadeia, com regiões de menor confiança concentradas em extremidades N e C terminais e em algumas alças longas expostas ao solvente. Os motivos selecionados pelo MEME que são compartilhados entre DGAT e GPAT apareceram na interface entre o núcleo catalítico e regiões de superfície, o que sugere que essas assinaturas podem estar relacionadas à coordenação de substrato comum ou à organização de um possível complexo multiproteico de montagem de TAG.

Os sítios de fosforilação priorizados na DGAT, escolhidos entre um conjunto muito amplo de resíduos potencialmente modificáveis, foram aqueles localizados em alças superficiais próximas ao suposto sítio catalítico ou a interfaces entre subdomínios. Essa seleção evitou a poluição visual das figuras e, ao mesmo tempo, destacou resíduos que, se

modificados, poderiam influenciar diretamente a eficiência da esterificação de diacilglicerol em TAG. A distribuição desses sítios em regiões de boa confiança estrutural pelo pLDDT reforça a interpretação de que se trata de posições estruturalmente bem definidas e, portanto, plausíveis como pontos de controle regulatório.

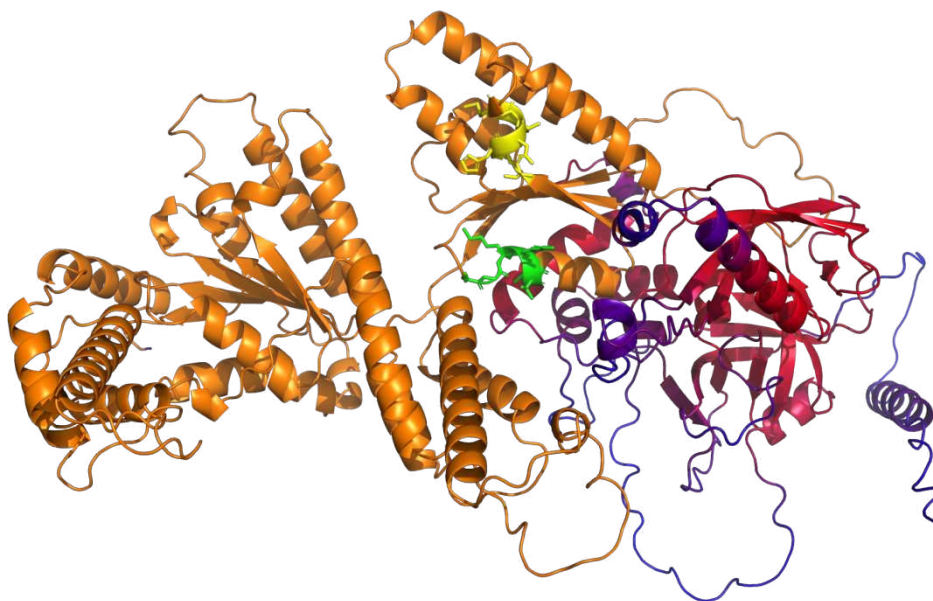
Figura 20 - Modelo 3D da DGAT colorido por pLDDT, com domínio aciltransferase destacado.



Fonte: Autor, 2025.

A figura 20 corresponde a estrutura predita por AlphaFold/ColabFold, colorida por pLDDT da proteína DGAT. A região em laranja corresponde ao domínio catalítico da DGAT, compreendido entre os resíduos 380–945, enquanto os trechos em azul e rosa representam regiões terminais mais flexíveis e de menor confiança estrutural.

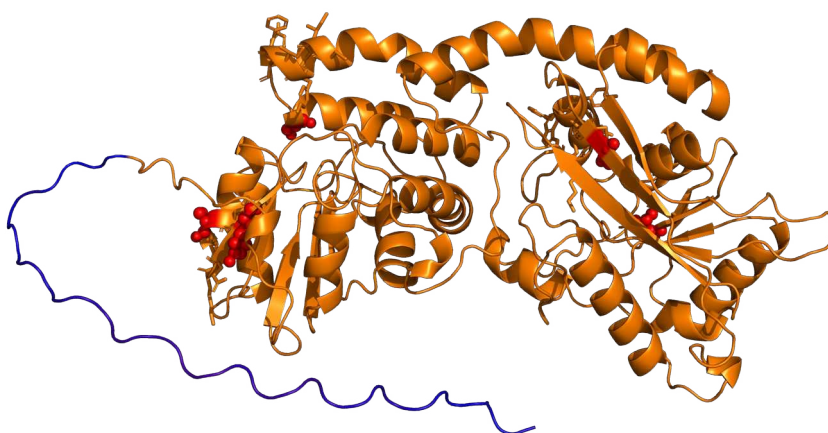
Figura 21 - DGAT com motivos selecionados em *sticks*.



Fonte: Autor, 2025.

Na figura 12 estão em destaque os seguintes motivos da enzima DGAT, em verde está o MEME – 6 com sua região entre os resíduos 388 – 393 e em amarelo o MEME – 9 com sua região compreendida entre os resíduos 409 – 414.

Figura 22 - DGAT com sítios de fosforilação selecionados em *spheres*.



Fonte: Autor, 2025.

Na figura 22 os sítios de fosforilação estão apresentados no formato sphere e coloridos em vermelho. A disposição dos sítios de DGAT sugere que a regulação por fosforilação, tende a ocorrer em áreas acessíveis às quinases, potencialmente modulando tanto a estabilidade conformacional quanto o acesso dos substratos ao sítio ativo da DGAT, o que é coerente com o papel dessa enzima como etapa terminal e altamente regulada na formação de triacilgliceróis.

5.12 Integração dos resultados estruturais

Quando analisados em conjunto, os modelos tridimensionais das cinco proteínas chave reforçam a coerência do quadro funcional delineado nas seções anteriores. Em primeiro lugar, a presença de domínios catalíticos bem definidos e organizados em núcleos de alta confiança estrutural em todas as enzimas confirma, em nível espacial, as anotações de ortologia fornecidas pelo KOfamScan e pelo CDD Search. Em segundo lugar, a localização dos motivos conservados em regiões que circundam ou compõem os sítios ativos sugere que essas assinaturas desempenham papel direto na catálise ou na estabilidade das enzimas, o que é particularmente claro em KAS I, SAD e nas aciltransferases GPAT e DGAT.

Do ponto de vista regulatório, a projeção dos sítios de fosforilação em 3D mostra um padrão consistente com o que foi inferido a partir da sequência: ACC e GPAT apresentam agrupamentos de resíduos fosforiláveis em alças acessíveis e próximas a regiões funcionais sensíveis, o que as posiciona como pontos privilegiados de controle na entrada da biossíntese de ácidos graxos e no início da montagem de TAG. DGAT, por sua vez, apresenta extensa superfície potencialmente modulável por fosforilação, o que é coerente com seu papel terminal na estocagem lipídica. KAS I e SAD exibem menor densidade de sítios, mas ainda assim com distribuição compatível com um ajuste fino de suas atividades de elongação e dessaturação.

Por fim, a integração entre modelagem tridimensional, domínios conservados, motivos e fosforilação reforça a visão de que as cinco proteínas selecionadas não apenas ocupam posições estratégicas nas vias descritas pelo KEGG, como também formam um eixo estruturalmente robusto e regulatoriamente sofisticado para a biossíntese de óleo em *Elaeis*

guineensis. A ACC, KAS I e SAD organizam o fornecimento e a modificação da cadeia de ácidos graxos no contexto plastidial, enquanto GPAT e DGAT coordenam a montagem sequencial do triacilglicerol no retículo endoplasmático. Os modelos 3D fecham o ciclo de análise iniciado na anotação funcional, oferecendo uma base espacial para futuras investigações experimentais e para possíveis estratégias de engenharia metabólica voltadas ao aumento de produtividade e à modulação do perfil lipídico em dendezeiro.

6 CONCLUSÃO

Este trabalho teve como objetivo caracterizar, por abordagens *in silico*, cinco enzimas-chave envolvidas na biossíntese de triacilgliceróis em *Elaeis guineensis*, articulando anotação funcional baseada em ortologia, análise estrutural e modelagem tridimensional. A partir do proteoma oficial da espécie, a utilização do KOfamScan permitiu selecionar, com alto rigor estatístico, as proteínas ACC, KAS I, SAD/FAD2-like, GPAT e DGAT, que representam etapas centrais da entrada de carbono, alongamento, dessaturação e montagem final de TAG.

As análises físico-químicas e de domínios conservados confirmaram a identidade enzimática de todas as proteínas e sustentaram sua participação nas vias de biossíntese de ácidos graxos e de glicerolipídios. Mesmo diante de previsões automáticas de localização celular em alguns casos divergentes do modelo biológico clássico, a presença de domínios catalíticos específicos e de motivos conservados reforçou a coerência funcional das enzimas selecionadas. A predição de múltiplos sítios de fosforilação, especialmente em DGAT e GPAT e padrão intermediário para ACC, evidenciou um elevado potencial de regulação pós-traducional em pontos estratégicos do fluxo lipídico.

A reconstrução de vias metabólicas no KEGG Mapper demonstrou que a biossíntese de ácidos graxos e a Via de Kennedy constituem os principais eixos metabólicos associados ao metabolismo de óleo no dendezeiro, e que as cinco proteínas estudadas ocupam posições-chave nesse encadeamento. A modelagem 3D por AlphaFold/ColabFold, seguida de visualização em PyMOL, permitiu integrar domínios, motivos e sítios de fosforilação em um contexto espacial, oferecendo um panorama estrutural consistente com o papel central dessas enzimas na lipogênese.

Em conjunto, os resultados obtidos mostram que a estratégia de combinação entre anotação por HMM, análise de domínios, predição de regulação e modelagem estrutural é eficaz para priorizar alvos moleculares em espécies oleaginosas. Embora se trate de um estudo inteiramente *in silico*, os achados fornecem uma base sólida para experimentos futuros de validação funcional, como análises de expressão em diferentes condições, ensaios enzimáticos e mutagênese dirigida em resíduos e motivos identificados como potenciais pontos de controle da síntese de óleo em *Elaeis guineensis*.

7 REFERÊNCIAS

- ALTSCHUL, S. F. et al. **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Research*, v. 25, n. 17, p. 3389-3402, 1997.
- ANDRADE, E. A. **Dendzeiro: uma planta que promete progresso para a Amazônia.** Ipades, 2012. Disponível em: <https://www.ipades.com.br/publicacoes/dendzeiro-Uma-planta-que-promete-progresso-para-a-Amazonia.pdf>. Acesso em: 11 nov. 2025.
- ARAMAKI, T. et al. **KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold.** *Bioinformatics*, v. 36, n. 7, p. 2251-2252, 2020.
- ASHBURNER, M. et al. **Gene Ontology: tool for the unification of biology.** *Nature Genetics*, v. 25, n. 1, p. 25-29, 2000.
- ASTARI, R. P. et al. **The protein model of somatic embryogenesis genes of oil palm (Elaeis guineensis Jacq.) using Phyre 2 and Swiss-model.** *IOP Conference Series: Earth and Environmental Science*, v. 1101, p. 012003, 2024.
- BAILEY, T. L. et al. **MEME Suite: tools for motif discovery and searching.** *Nucleic Acids Research*, v. 37, n. suppl. 2, p. W202-W208, 2009.
- BELTRÃO, N. E. de M.; OLIVEIRA, M. I. M. de. **Dendzeiro (Elaeis guineensis Jacq.).** In: BELTRÃO, N. E. de M.; OLIVEIRA, M. I. M. de (org.). *Oleaginosas no Brasil: potencial de produção de óleos vegetais para biodiesel*. Brasília: Embrapa, 2007. p. 341-388.
- BRASILHO, M. B. et al. **O dendzeiro (Elaeis guineensis Jacq.) – revisão.** *Bioenergia em Revista: Diálogos*, v. 11, n. 1, p. 27-38, 2015.
- CAMACHO, C. et al. **BLAST+: architecture and applications.** *BMC Bioinformatics*, v. 10, p. 421, 2009.
- COSTA-SILVA, C.; DOMINGUES, D.; LOPES, F. M. **RNA-Seq differential expression analysis: an extended review and a software tool.** *Frontiers in Genetics*, v. 8, p. 1-10, 2017.
- DIAS, U. M. **Predição da função das proteínas sem alinhamentos usando máquinas de vetor de suporte.** 2007. Dissertação (Mestrado em Ciências da Computação) – Universidade Federal de Alagoas, Maceió, 2007.
- EMBRAPA. **Dendê.** Portal Embrapa, 2011. Disponível em: <https://www.embrapa.br/agencia-de-informacao-tecnologica/tematicas/agroenergia/biodiesel/materias-primas/dende>. Acesso em: 08 nov. 2025.
- EMBRAPA. **Genes highly overexpressed in salt-stressed young oil palm (Elaeis guineensis) plants.** *Pesquisa Agropecuária Brasileira*, v. 55, p. e01234, 2020.

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

GONG, F. et al. **Structural insights into eukaryotic DNA polymerases.** *Frontiers in Molecular Biosciences*, v. 5, p. 8, 2018.

GUNSTONE, F. D. **Vegetable Oils in Food Technology: composition, properties and uses.** 2. ed. Oxford: Wiley-Blackwell, 2011.

HASSAN, A. H. et al. **Transposable elements: multifunctional players in the plant genome.** *Frontiers in Plant Science*, v. 14, p. 1330127, 2024.

HUNTER, S. et al. **InterPro in 2011: new developments in protein families classification and functional annotation.** *Nucleic Acids Research*, v. 40, n. D1, p. D306-D312, 2012.

JUMPER, J. et al. **Highly accurate protein structure prediction with AlphaFold.** *Nature*, v. 596, n. 7873, p. 583-589, 2021.

KANEHISA, M. et al. **KEGG as a reference resource for gene and protein annotation.** *Nucleic Acids Research*, v. 44, n. D1, p. D457-D462, 2016.

KANEHISA, M.; FURUMICHI, M.; SATO, Y. **KEGG: integrating viruses and cellular organisms.** *Nucleic Acids Research*, v. 49, n. D1, p. D545-D551, 2021.

KROGH, A. et al. **Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes.** *Journal of Molecular Biology*, v. 305, n. 3, p. 567-580, 2001.

LEWIS, T. E. et al. **Genome3D: a UK collaborative project to annotate genomic sequences with predicted 3D structures based on SCOP and CATH domains.** *Nucleic Acids Research*, v. 40, n. D1, p. D483-D489, 2012.

LIU, M. et al. **Three differential expression analysis methods for RNA sequencing: limma, edgeR, DESeq2.** *Journal of Visualized Experiments*, n. 175, e62528, 2021.

LOMSADZE, A. et al. **Gene identification in novel eukaryotic genomes by self-training algorithm.** *Nucleic Acids Research*, v. 33, n. 20, p. 6494-6506, 2005.

LOPES, B. T. S. **Predição de genes ab initio combinada com informações de alinhamento.** 2019. Dissertação (Mestrado em Ciências da Computação) – Universidade de São Paulo, São Paulo, 2019.

LOVE, M. I.; HUBER, W.; ANDERS, S. **Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.** *Genome Biology*, v. 15, n. 12, p. 550, 2014.

MARCHLER-BAUER, A. et al. **CDD: a conserved domain database for interactive domain family analysis.** *Nucleic Acids Research*, v. 35, n. suppl. 1, p. D237-D240, 2007.

- MARCHLER-BAUER, A. et al. **CDD: a Conserved Domain Database for the functional annotation of proteins**. *Nucleic Acids Research*, v. 39, n. D1, p. D225-D229, 2011.
- MITCHELL, A. L. et al. **The InterPro protein families database: the classification resource after 15 years**. *Nucleic Acids Research*, v. 43, n. D1, p. D213-D221, 2015.
- MIRDITA, M. et al. **ColabFold: making protein folding accessible to all**. *Nature Methods*, v. 19, n. 6, p. 679-682, 2022.
- NAVA, A. A. et al. **The omics era: a nexus of untapped potential for Mendelian chromatinopathies**. *Human Genetics*, v. 143, n. 4, p. 475-495, 2023.
- NEXO JORNAL. **De onde vem o dendê e onde ele é produzido**. 2022. Disponível em: <https://www.nexojornal.com.br/grafico/2022/07/19/de-onde-vem-o-dende-e-onde-ele-e-produzido>. Acesso em: 11 nov. 2025.
- OHLROGGE, J.; BROWSE, J. **Lipid biosynthesis**. *The Plant Cell*, v. 7, n. 7, p. 957-970, 1995.
- ROBINSON, M. D.; MCCARTHY, D. J.; SMYTH, G. K. **edgeR: a Bioconductor package for differential expression analysis of digital gene expression data**. *Bioinformatics*, v. 26, n. 1, p. 139-140, 2010.
- ROOT, D. D. **PyMOL: molecular graphics system**. Version 1.8. Schrödinger, LLC, 2007.
- SALGADO, F. F. **Predição e anotação de miRNAs e o papel dos fatores de transcrição na resposta de *Elaeis guineensis* Jacq. aos estresses abióticos de seca e salinidade**. 2023. Dissertação (Mestrado em Genética e Melhoramento de Plantas) – Universidade Federal de Lavras, Lavras, 2023.
- SILVA, C. S. L. **Análise da produção do dendezeiro (*Elaeis guineensis* Jacq.) e suas implicações socioeconômicas**. 2017. Dissertação (Mestrado em Geografia) – Universidade Federal de Pernambuco, Recife, 2017.
- SINGH, R. et al. **Oil palm genome sequence reveals divergence of interfertile species in old and new worlds**. *Nature*, v. 500, n. 7462, p. 335-339, 2013.
- TEUFEL, F. et al. **SignalP 6.0 predicts all five types of signal peptides using protein language models**. *Nature Biotechnology*, v. 40, p. 1023-1025, 2022.
- THUMULURI, V. et al. **DeepLoc 2.0: multi-label protein subcellular localization prediction using protein language models**. *Nucleic Acids Research*, v. 50, n. W1, p. W228-W234, 2022.
- TSUYUZAKI, K. et al. **MeSH-based functional analysis of omics data**. *PLoS One*, v. 10, n. 4, e0123719, 2015.
- VALENTE, G. T. **Predição de estruturas de proteínas e suas aplicações**. 2013. Apresentação (Disciplina de Bioinformática Estrutural) – Instituto de Biociências de Botucatu, Universidade Estadual Paulista, 2013.

WAGNER, G. P.; KIN, K.; LYNCH, V. J. **Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples.** *Genome Biology*, v. 13, n. 3, p. R22, 2012.

WANG, Y. et al. **Protein domain identification methods and online resources.** *Computational and Structural Biotechnology Journal*, v. 19, p. 1108-1116, 2021.