

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Paula Prado Carvalho

**Investigação de diferentes abordagens do
Algoritmo Evolutivo aplicado ao problema
de predição de proteínas no modelo HP-3D**

Uberlândia, Brasil

2026

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Paula Prado Carvalho

**Investigação de diferentes abordagens do Algoritmo
Evolutivo aplicado ao problema de predição de
proteínas no modelo HP-3D**

Trabalho de conclusão de curso apresentado à Faculdade de Computação da Universidade Federal de Uberlândia, como parte dos requisitos exigidos para a obtenção título de Bacharel em Sistemas de Informação.

Orientador: Profa. Dra. Christiane Regina Soares Brasil

Universidade Federal de Uberlândia – UFU

Faculdade de Computação

Bacharelado em Sistemas de Informação

Uberlândia, Brasil

2026

Paula Prado Carvalho

Investigação de diferentes abordagens do Algoritmo Evolutivo aplicado ao problema de predição de proteínas no modelo HP-3D

Trabalho de conclusão de curso apresentado à Faculdade de Computação da Universidade Federal de Uberlândia, como parte dos requisitos exigidos para a obtenção título de Bacharel em Sistemas de Informação.

Trabalho aprovado. Uberlândia, Brasil, 19 de março de 2026:

Profa. Dra. Christiane Regina Soares
Brasil
Orientadora

Prof. Dr. Paulo Henrique Ribeiro
Gabriel

Prof. Dr. Renan Gonçalves Cattelan

Uberlândia, Brasil
2026

*“Nós só podemos ver um pouco do futuro, mas o suficiente para perceber que há
muito a fazer.” - Alan Turing*

Resumo

Os Algoritmos Evolutivos (AEs) são meta-heurísticas inspiradas na evolução biológica, amplamente utilizadas na resolução de problemas de otimização complexos, especialmente aqueles que envolvem múltiplos objetivos conflitantes. Este trabalho tem como objetivo investigar e comparar o desempenho de diferentes variantes de AEs, considerando abordagens monoobjetivo e multiobjetivo, aplicadas ao problema de Predição de Estruturas de Proteínas (PSP) no modelo hidrofóbico-polar tridimensional (HP-3D). Para isso, foram implementadas e avaliadas computacionalmente diferentes abordagens, com destaque para as variantes multiobjetivo baseadas em múltiplas tabelas, como o AEMMT. A avaliação dos algoritmos se deu utilizando métricas como aptidão das soluções (*fitness*), número de avaliações e taxa de acerto. Os resultados evidenciam que as abordagens multiobjetivo apresentam desempenho superior em relação às monoobjetivo, especialmente em instâncias mais complexas. Como principal contribuição, este estudo fornece uma análise comparativa sistemática que identifica as configurações algorítmicas mais robustas e eficientes para o problema no modelo HP-3D, corroborando a literatura e reforçando o potencial dos AEs multiobjetivo para esta aplicação.

Palavras-chave: Algoritmos Evolutivos, Otimização Multiobjetivo, AEMT, AEMMT, Predição de Estrutura de Proteínas.

Abstract

Evolutionary Algorithms (EAs) are metaheuristics inspired by biological evolution, widely used to solve complex optimization problems, especially those involving multiple conflicting objectives. This work aims to investigate and compare the performance of different EA variants, considering both single-objective and multi-objective approaches, applied to the Protein Structure Prediction (PSP) problem using the three-dimensional hydrophobic-polar (HP-3D) model. To this end, different approaches were computationally implemented and evaluated, highlighting multiobjective variants based on multiple tables, such as the AEMMT. The algorithms were evaluated using metrics such as solution fitness, number of evaluations, and success rate. The results show that multiobjective approaches outperform single-objective ones, especially in more complex instances. As a main contribution, this study provides a systematic comparative analysis that identifies the most robust and efficient algorithmic configurations for the problem in the HP-3D model, corroborating the literature and reinforcing the potential of multiobjective EAs for this application.

Keywords: Evolutionary Algorithms, Multiobjective Optimization, AEMT, AEMMT, Protein Structure Prediction.

Lista de ilustrações

Figura 1 – Estrutura geral de um aminoácido.	19
Figura 2 – Cruzamento 2-pontos.	22
Figura 3 – Mutação.	22
Figura 4 – Estrutura geral de um aminoácido.	31
Figura 5 – Estrutura primária de uma proteína.	32
Figura 6 – Estrutura secundária de proteínas.	32
Figura 7 – Estrutura terciária de uma proteína.	33
Figura 8 – Estrutura quaternária de uma proteína.	33
Figura 9 – Representação do Modelo Hidrofóbico-Polar (HP)	35
Figura 10 – Representação do modelo Hidrofóbico-Polar (HP) em 3D	36
Figura 11 – Conformação obtida pelo AEMMT 5 para a sequência 3d4.	73
Figura 12 – Conformação obtida pelo AEMMT 5 para a sequência A11.	73

Lista de tabelas

Tabela 1 – Primeiro conjunto de instâncias do modelo HP-3D utilizadas nos experimentos (TRAYKOV et al., 2018; ESPITIA; PANG; GUMBART, 2024; LIN; SU, 2011).	57
Tabela 2 – Segundo conjunto de instâncias do modelo HP-3D utilizadas nos experimentos (MANSOUR; KANJ; KHACHFE, 2011; GUO et al., 2017; ESPITIA; PANG; GUMBART, 2024; GABRIEL; MELO; DELBEM, 2012).	57
Tabela 3 – Resultados do AE monoobjetivo com função de aptidão de n_H para o primeiro conjunto de instâncias.	58
Tabela 4 – Resultados do AE monoobjetivo com função de aptidão de n_H para o segundo conjunto de instâncias.	58
Tabela 5 – Resultados do algoritmo AE monoobjetivo baseado na função ponderada (δ) para o primeiro conjunto de instâncias.	59
Tabela 6 – Resultados do algoritmo AE monoobjetivo baseado na função ponderada (δ) para o segundo conjunto de instâncias.	60
Tabela 7 – Resultados do algoritmo AE monoobjetivo baseado na energia simplificada (E_s) para o primeiro conjunto de instâncias.	61
Tabela 8 – Resultados do algoritmo AE monoobjetivo baseado na energia simplificada (E_s) para o segundo conjunto de instâncias.	61
Tabela 9 – Resultados do AE monoobjetivo com função ponderada (δ) para o primeiro conjunto de instâncias.	63
Tabela 10 – Resultados do AE monoobjetivo com função ponderada (δ) para o segundo conjunto de instâncias.	63
Tabela 11 – Resultados do AEMMT 5 para o primeiro conjunto de instâncias.	65
Tabela 12 – Resultados do algoritmo AEMMT 5 para o segundo conjunto de instâncias.	65
Tabela 13 – Resultados do algoritmo AEMMT 6 para o primeiro conjunto de instâncias.	67

Tabela 14 – Resultados do algoritmo AEMMT 6 para o segundo conjunto de instâncias.	67
Tabela 15 – Comparação geral entre os algoritmos implementados.	70
Tabela 16 – Principais trabalhos relacionados para PSP no modelo HP-3D. .	71
Tabela 17 – Comparação da energia mínima obtida por diferentes algoritmos para o primeiro conjunto de instâncias.	72
Tabela 18 – Comparação de energia mínima obtida por diferentes algoritmos para o segundo conjunto de instâncias.	72

Lista de abreviaturas e siglas

AE	Algoritmo Evolutivo
AEMT	Algoritmo Evolutivo Multiobjetivo em Tabela
AEMMT	Algoritmo Evolutivo Multiobjetivo com Muitas Tabelas
AEMO	Algoritmo Evolutivo Multiobjetivo
AG	Algoritmo Genético
AGMO	Algoritmo Genético Multiobjetivo
ACO	<i>Ant Colony Optimization</i> (Otimização por Colônia de Formigas)
EGA	<i>Enhanced Genetic Algorithm</i>
FFNN-R	<i>Feedforward Neural Network with Reservoir Computing</i>
GA	Genetic Algorithm
HGA-PSO	<i>Hybrid Genetic Algorithm with Particle Swarm Optimization</i>
HP	Hidrofóbico-Polar
HP-3D	Modelo Hidrofóbico-Polar Tridimensional
LSTM	<i>Long Short-Term Memory</i>
MaOEA	Algoritmo Evolutivo com Muitos Objetivos
MOEA/D	<i>Multiobjective Evolutionary Algorithm based on Decomposition</i>
ND	Não Dominância
NP	<i>Nondeterministic Polynomial Time</i>
NSGA-II	<i>Non-Dominated Sorting Genetic Algorithm II</i>

NSGA-III	<i>Non-Dominated Sorting Genetic Algorithm III</i>
PG	Programação Genética
PSO	<i>Particle Swarm Optimization</i>
PSP	Predição de Estruturas de Proteínas
RNM	Ressonância Nuclear Magnética
SPEA-II	<i>Strength Pareto Evolutionary Algorithm II</i>
TPPSO	Tabu-based Particle Swarm Optimization

Sumário

1	INTRODUÇÃO	13
1.1	Contextualização	13
1.2	Objetivos	15
1.2.1	Objetivo Geral	15
1.2.2	Objetivos Específicos	15
1.3	Justificativa	15
1.4	Estrutura do texto	16
2	ALGORITMOS EVOLUTIVOS	18
2.1	AEs e seus operadores genéticos	18
2.1.1	Seleção	20
2.1.2	Cruzamento	21
2.1.3	Mutação	21
2.2	Algoritmos Evolutivos Multiobjetivos	23
2.2.1	Fronteira de Pareto	23
2.2.2	Principais Algoritmos Evolutivos Multiobjetivos	25
2.2.3	Algoritmo Evolutivo Multiobjetivo em Tabela	26
2.2.4	Algoritmo Evolutivo Multiobjetivo com Muitas Tabelas	27
2.3	Considerações finais	28
3	O PROBLEMA DE PREDIÇÃO DE ESTRUTURA DE PROTEÍNA	30
3.1	Fundamentos da predição de estrutura de proteínas	30
3.1.1	Estrutura de Proteínas	31
3.1.2	Métodos Computacionais de Predição	33
3.1.3	Modelo Hidrofóbico-Polar (HP)	34
3.1.4	Funções de Energia no Modelo HP	37
3.2	Trabalhos Relacionados	37
3.3	Considerações finais	41

4	AEMMT PARA O PROBLEMA PSP NO MODELO HP-3D	43
4.1	Representação do indivíduo e do espaço de busca	43
4.2	Avaliação dos indivíduos	44
4.2.1	<i>Pipeline</i> de avaliação e ordem de cálculo	45
4.2.2	Dependência entre métricas	47
4.2.3	Manutenção de soluções inviáveis e penalização	47
4.3	Algoritmos Evolutivos monoobjetivo	48
4.3.1	Fluxo evolutivo dos AEs monoobjetivo	49
4.4	Algoritmos Evolutivos Multiobjetivo com Muitas Tabelas	51
4.4.1	Fluxo evolutivo dos AEMMTs	53
4.5	Considerações Finais	54
5	RESULTADOS	56
5.1	Configuração Experimental	56
5.2	Experimentos com Algoritmos Evolutivos monoobjetivo	58
5.2.1	AE monoobjetivo com maximização de contatos H (n_H)	58
5.2.2	AE monoobjetivo com função ponderada (δ)	59
5.2.3	AE monoobjetivo com energia simplificada (E_s)	60
5.2.4	Análise comparativa entre os AEs monoobjetivo	62
5.2.5	AE monoobjetivo de melhor desempenho (δ)	62
5.3	Experimentos com os AEMMTs	64
5.3.1	AEMMT com 5 Tabelas	64
5.3.2	AEMMT com 6 Tabelas	66
5.3.3	Análise Comparativa entre AEMMT 5 e AEMMT 6	68
5.4	Análise comparativa entre métodos	69
5.4.1	Análise comparativa entre os AEs implementados	70
5.4.2	Análise comparativa dos AEs com trabalhos relacionados	70
5.5	Considerações Finais	74
6	CONCLUSÃO	76
	REFERÊNCIAS	79

1 Introdução

Este capítulo tem como objetivo apresentar o problema abordado neste trabalho, bem como os principais conceitos necessários para sua compreensão. Serão discutidas as abordagens consideradas, a justificativa para sua utilização e, por fim, a organização geral do trabalho.

1.1 Contextualização

Os Algoritmos Evolutivos (AEs) constituem uma classe de métodos computacionais inspirados em princípios da evolução biológica, como seleção natural, recombinação e mutação, sendo amplamente utilizados na resolução de problemas complexos de otimização. Esses algoritmos destacam-se por sua capacidade de explorar espaços de busca extensos, não lineares e de alta dimensionalidade, nos quais métodos determinísticos tradicionais frequentemente apresentam desempenho insatisfatório ([GABRIEL; DELBEM, 2008](#); [EIBEN; SMITH, 2015](#)).

Uma característica fundamental de diversos problemas do mundo real é a presença de múltiplos objetivos (conflitantes ou não), o que inviabiliza a obtenção de uma única solução ótima sob todos os critérios simultaneamente. Nesse contexto, surgiram os Algoritmos Evolutivos multiobjetivo, que buscam aproximar o conjunto de soluções eficientes. Tal conjunto é conhecido como Fronteira de Pareto ([COELLO, 2009](#)). Entretanto, à medida que o número de objetivos aumenta, especialmente acima de quatro, os algoritmos multiobjetivo tradicionais passam a enfrentar limitações relacionadas à perda de pressão seletiva, à manutenção da diversidade e ao aumento do custo computacional ([ISHIBUCHI; TSUKAMOTO; NOJIMA, 2008](#)), sendo necessário trabalhar com os AEs com muitos objetivos nestes casos.

Dentre os diversos problemas que se beneficiam dessas abordagens, a Predição de Estruturas de Proteínas (PSP) destaca-se como um desafio computacional

de alta complexidade, sendo classificado como um problema NP¹. A PSP consiste em determinar a estrutura tridimensional de uma proteína a partir de sua sequência de aminoácidos, sendo essa estrutura diretamente relacionada à sua função biológica. Portanto, descobrindo a conformação estrutural de uma proteína, tem-se a sua funcionalidade. Embora existam métodos experimentais para a determinação estrutural, como a Cristalografia de Raio X e a Ressonância Nuclear Magnética, tais técnicas são extremamente custosas computacionalmente e têm limitações práticas, o que justifica a crescente utilização de métodos de otimização computacional (LAU; DILL, 1989; CUTELLO; NARZISI; NICOSIA, 2005; SANTANA; NAHUM, 2022). Neste problema, os métodos exatos tornam-se inviáveis, tendo os Algoritmos Evolutivos como alternativas eficientes para a obtenção de soluções aproximadas. Para viabilizar a modelagem computacional, modelos simplificados como o Modelo Hidrofóbico-Polar (HP) são amplamente empregados, representando o dobramento proteico em malhas discretas e considerando as interações hidrofóbicas² como principal força motriz.

Diante disso, as abordagens baseadas em otimização multiobjetivo e com muitos objetivos mostram-se particularmente adequadas ao PSP, permitindo considerar simultaneamente múltiplos critérios de qualidade das estruturas preditas. Neste sentido, este trabalho focou na implementação e na comparação de variantes de Algoritmos Evolutivos, com ênfase no AE monoobjetivo (um objetivo) e no AEMMT (Algoritmo Evolutivo Multiobjetivo com Muitas Tabelas) (BRASIL, 2012; BRASIL; DELBEM; SILVA, 2013), aplicadas ao problema de PSP no Modelo HP-3D.

¹ Um problema NP (*Nondeterministic Polynomial time*) é uma classe de problemas de decisão para os quais, dada uma solução candidata, é possível verificar se essa solução é correta em tempo polinomial por um algoritmo determinístico (CORMEN et al., 2022).

² Interações hidrofóbicas são interações que ocorrem entre moléculas ou partes de moléculas apolares quando estão em um meio aquoso, que tendem a se agrupar para evitar contato com a água (NELSON; COX, 2017).

1.2 Objetivos

1.2.1 Objetivo Geral

Este Trabalho de Conclusão de Curso visou investigar o desempenho de variantes distintas de Algoritmos Evolutivos ao problema de Predição de Estruturas de Proteínas no Modelo HP-3D.

1.2.2 Objetivos Específicos

Como objetivos específicos, este trabalho buscou:

- Desenvolver cinco abordagens do AE (três versões monoobjetivo e duas versões do AEMMT).
- Realizar experimentos com os AEs desenvolvidos variando-se os parâmetros do algoritmo, como o tamanho da população, o número de gerações e o tipo de seleção empregado.
- Analisar os resultados obtidos por meio de experimentação controlada, comparando sistematicamente o desempenho das diferentes variantes com base em métricas como a aptidão das soluções e a taxa de acerto.
- Comparar os resultados obtidos a alguns trabalhos da literatura desta área específica com base nos melhores valores de aptidões alcançados.

1.3 Justificativa

Os Algoritmos Evolutivos multiobjetivo são técnicas de otimização inspirados na evolução natural que permitem resolver problemas com múltiplos objetivos conflitantes simultaneamente, algo muito comum no mundo real. Diferentemente dos AEs monoobjetivo que focam em apenas um critério, esses algoritmos buscam um conjunto de soluções (Fronteira de Pareto), permitindo equilibrar diferentes fatores, que podem ser conflitantes ou não.

Na prática, esses algoritmos são aplicados nas mais diversas áreas. Na engenharia, auxiliam no projeto de produtos ao otimizar de modo simultâneo características como peso, resistência e eficiência energética. Na logística, ajudam a planejar rotas equilibrando tempo, custo e emissões. Em sistemas de energia, contribuem para o planejamento de geração e distribuição considerando custo, confiabilidade e impacto ambiental. Além disso, na área de *Machine Learning*, podem otimizar simultaneamente desempenho, complexidade e tempo de treinamento de modelos, tornando-os ferramentas valiosas para resolver problemas reais complexos.

Diante desse contexto, estudar Algoritmos Evolutivos multiobjetivo é fundamental para desenvolver métodos capazes de lidar com a complexidade dos problemas modernos, nos quais diferentes objetivos precisam ser considerados concomitantemente. A compreensão desses algoritmos permite criar soluções mais eficientes, flexíveis e adaptáveis, além de fornecer suporte mais robusto à tomada de decisão em ambientes com múltiplos critérios e restrições. Dessa forma, o estudo desses algoritmos não apenas amplia o conhecimento científico, mas também possibilita o desenvolvimento de soluções mais eficazes para desafios reais e complexos.

1.4 Estrutura do texto

Este Trabalho de Conclusão de Curso foi organizado da seguinte maneira:

- Capítulo 2: discutirá os fundamentos de Algoritmos Evolutivos, incluindo seus operadores genéticos e conceitos relacionados à otimização multiobjetivo.
- Capítulo 3: apresentará os conceitos fundamentais do problema de predição de estruturas de proteínas, abordando o modelo hidrofóbico-polar (HP), funções de energia e trabalhos relacionados.
- Capítulo 4: descreverá as abordagens propostas dos AEs monoobjetivo e dos AEMMTs para o problema de predição de estruturas de proteínas no modelo HP tridimensional, incluindo a representação das soluções, os critérios de avaliação e o fluxo evolutivo dos algoritmos.

- Capítulo 5: apresentará os experimentos realizados, a configuração experimental adotada e a análise comparativa dos resultados obtidos pelos algoritmos implementados.
- Capítulo 6: apresentará as conclusões do trabalho e discutirá possíveis direções para pesquisas futuras.

2 Algoritmos Evolutivos

Este capítulo apresentará os principais conceitos teóricos que fundamentam o desenvolvimento deste Trabalho de Conclusão de Curso. Inicialmente, serão discutidos os aspectos relacionados aos Algoritmos Evolutivos, abordando seus princípios, funcionamento e principais operadores genéticos. Em seguida, serão introduzidos os conceitos de otimização multiobjetivo e com muitos objetivos, destacando suas características, desafios e algoritmos clássicos.

2.1 AEs e seus operadores genéticos

Os Algoritmos Evolutivos (AEs) constituem uma classe de meta-heurísticas de otimização inspiradas em princípios da evolução biológica, como seleção natural, recombinação genética e mutação. Essas técnicas imitam o processo evolutivo observado na natureza, utilizando populações de soluções candidatas que evoluem ao longo de gerações em busca de soluções de melhor qualidade para um determinado problema ([GABRIEL; DELBEM, 2008](#); [COELLO, 2009](#)).

Em um AE, cada solução potencial é representada por um indivíduo, cuja qualidade é avaliada por meio de uma função de aptidão (*fitness*). A partir dessa avaliação, os indivíduos mais bem adaptados possuem maior probabilidade de serem selecionados, dependendo do tipo de seleção, para gerar novos indivíduos por meio da aplicação de operadores genéticos, como o cruzamento e a mutação, se for o caso (uma vez que nem sempre é efetuada a mutação). Esse processo é repetido iterativamente até que um critério de parada seja satisfeito, como o alcance de um número máximo de gerações ou a obtenção da solução ótima.

Além disso, é comum empregar um mecanismo de elitismo, responsável pela seleção direta dos melhores indivíduos da população atual para a próxima geração, sem sofrer alterações pelos operadores evolutivos. Esse procedimento garante que soluções de alta qualidade não sejam perdidas ao longo das iterações, contribuindo para a estabilidade e a convergência do processo evolutivo. O elitismo é geralmente

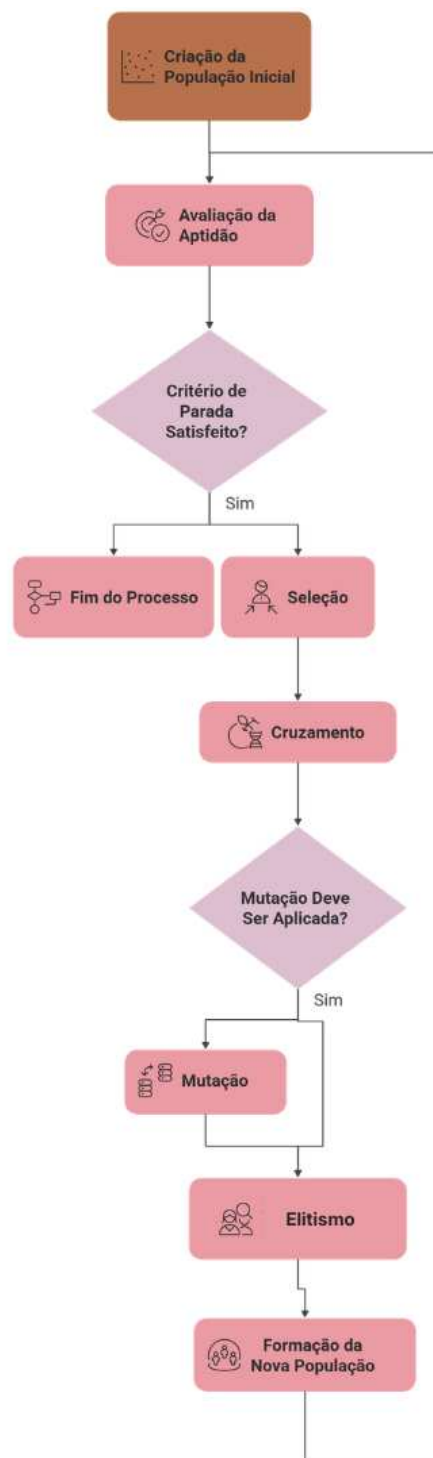


Figura 1 – Fluxograma de um Algoritmo Evolutivo.

controlado por uma taxa ou proporção de indivíduos preservados, cujo valor é sempre maior ou igual a zero. A Figura 1 demonstra uma representação geral de um AE típico.

A seguir, serão detalhados os operadores genéticos que os AE utilizam para mimetizar o processo evolutivo que ocorre na natureza.

2.1.1 Seleção

A seleção é o mecanismo responsável por definir quais indivíduos da população participarão do processo de geração da próxima geração, estabelecendo quais soluções terão a oportunidade de contribuir com seus atributos para os descendentes. Dependendo do método adotado, essa escolha pode ou não ser influenciada diretamente pelos valores de aptidão, podendo incluir desde estratégias baseadas no desempenho até abordagens aleatórias ou parcialmente estocásticas.

Entre os métodos de seleção frequentemente utilizados, destacam-se a seleção por roleta, a seleção elitista e a seleção por torneio de N , descritas brevemente a seguir:

- Na seleção por roleta, cada indivíduo possui uma probabilidade de ser escolhido proporcional ao seu valor de aptidão, de modo que soluções com melhor desempenho tendem a ser selecionadas com maior frequência, embora indivíduos menos aptos ainda possam ser escolhidos estocasticamente. Deste modo, há grande chance de ser escolhidos dois pais com boa qualificação (GABRIEL; DELBEM, 2008).
- A seleção elitista consiste na preservação direta de um número ou proporção dos melhores indivíduos da população atual para a próxima geração, garantindo que as soluções de maior qualidade não sejam perdidas ao longo do processo evolutivo (GABRIEL; DELBEM, 2008).
- Na seleção por torneio de N , um subconjunto de N indivíduos é escolhido aleatoriamente e o indivíduo com melhor aptidão dentro desse subconjunto é selecionado como pai, e esse processo é repetido para selecionar o outro pai.

Este método é amplamente empregado devido à sua simplicidade e ao controle indireto de pressão seletiva por meio do tamanho do torneio (SHUKLA; PANDEY; MEHROTRA, 2015).

2.1.2 Cruzamento

O cruzamento (*crossover*), também denominado de recombinação, é inspirada na reprodução sexuada e tem como objetivo combinar informações de dois (ou mais) indivíduos, produzindo descendentes que herdem partes do material genético de cada progenitor. No contexto dos Algoritmos Genéticos¹ (uma classe dos Algoritmos Evolutivos), esse operador é considerado central, pois promove a troca de características potencialmente úteis entre soluções já exploradas.

Entre os operadores mais comuns, citam-se o cruzamento de um ponto, de dois pontos (Figura 2) e o cruzamento uniforme, os quais diferem principalmente pela forma como definem os pontos de corte e como alternam os segmentos herdados de cada pai (GABRIEL; DELBEM, 2008). A Figura 2 ilustra o cruzamento de dois pontos, no qual são definidos dois pontos de corte nos cromossomos dos pais, delimitando um segmento intermediário. Esse trecho é então trocado entre os progenitores, de modo que cada descendente mantém parte do material genético original e recebe o segmento correspondente do outro pai, resultando em novas combinações de características.

2.1.3 Mutação

A mutação é capaz de introduzir uma variação genética ao modificar aleatoriamente genes de um indivíduo com determinada probabilidade, contribuindo para preservar diversidade e reduzir o risco de convergência prematura. Em codificações binárias, por exemplo, uma forma clássica de mutação consiste na inversão do valor de um gene (troca entre 0 e 1). Outra forma de mutação é ilustrada na Figura 3, em que ocorre por substituição do valor de um gene por outro valor possível em seu domínio, alterando diretamente o conteúdo da posição selecionada

¹ Os Algoritmos Genéticos são uma classe específica dos Algoritmos Evolutivos, que obrigatoriamente utilizam o operador de cruzamento para geração de novos indivíduos, podendo ou não efetuar a mutação (GOLDBERG, 1989).

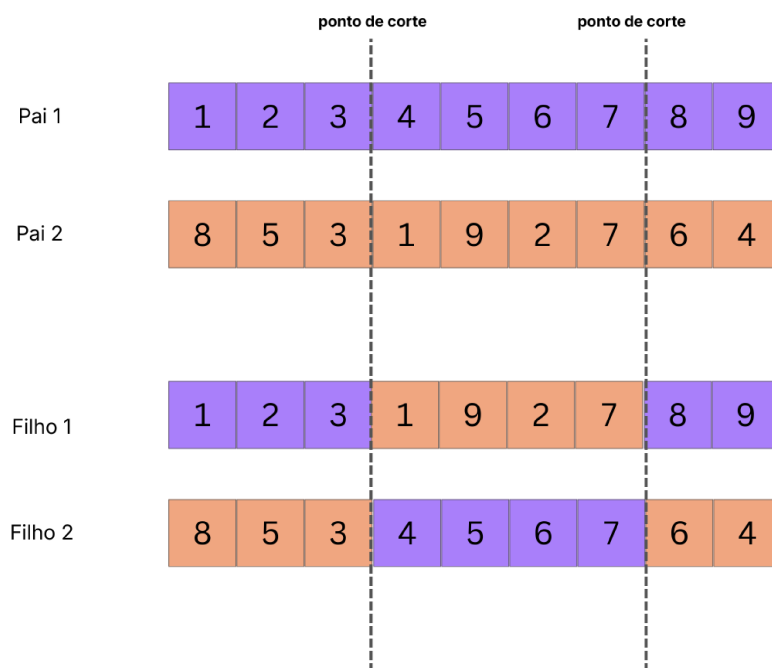


Figura 2 – Cruzamento 2-pontos. Adaptado de Dias (2016)

do cromossomo. A taxa de mutação deve ser adequadamente ajustada, pois valores muito elevados tendem a deteriorar a qualidade das soluções, ao passo que valores muito baixos podem reduzir a exploração do espaço de busca (GOLDBERG, 1989).

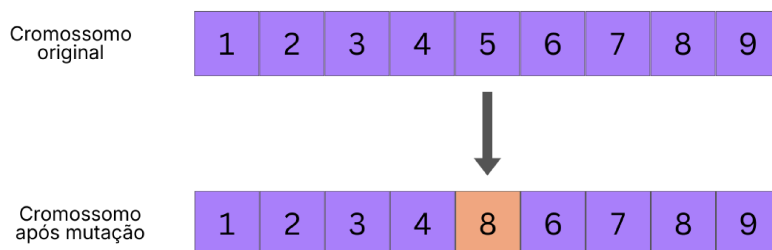


Figura 3 – Mutação. Adaptado de Dias (2016)

A seguir, serão explicados os principais fundamentos dos Algoritmos Evolutivos Multiobjetivos, citando alguns dos principais algoritmos da literatura nesta área de estudo.

2.2 Algoritmos Evolutivos Multiobjetivos

Uma das principais vantagens dos AEs está associada à sua capacidade de explorar espaços de busca amplos, não lineares e altamente complexos, nos quais métodos determinísticos tradicionais tendem a apresentar limitações. Além disso, por trabalharem com um conjunto de soluções simultaneamente, os AEs tendem a apresentar maior robustez frente a ótimos locais, sendo altamente aplicados em problemas de otimização combinatória, engenharia, bioinformática e inteligência artificial (ZHANG; LI, 2007; DIAS, 2016; SANTOS, 2023; SANTOS, 2024; NASCIMENTO; VARGAS; WANNER, 2024).

Em problemas reais de otimização, é frequente a presença de múltiplos objetivos que podem ser conflitantes ou não. Os objetivos conflitantes são aqueles que a melhoria de um critério pode resultar na degradação de outro. Nesse cenário, os Algoritmos Evolutivos Multiobjetivo (AEMOs) configuram-se como uma extensão natural dos algoritmos evolutivos tradicionais, também chamados de monoobjetivos (com um objetivo), permitindo o tratamento simultâneo de múltiplos objetivos por meio de vetores de aptidão e do conceito de dominância de Pareto (BUENO; OLIVEIRA, 2010).

Nas próximas subseções, serão explicados os fundamentos da Fronteira de Pareto.

2.2.1 Fronteira de Pareto

Em problemas de otimização que envolvem múltiplos objetivos, muitas vezes conflitantes entre si, não existe, via de regra, uma única solução que seja a melhor em todos os critérios simultaneamente. Para lidar com esse cenário, introduz-se o conceito de Fronteira de Pareto (VELDHUIZEN, 1998).

Formalmente, a Fronteira de Pareto é constituída pelo conjunto de soluções não dominadas em relação às demais. Uma solução é considerada não dominada quando não existe outra que apresente desempenho superior em todos os objetivos simultaneamente.

Nesse contexto, diz-se que uma solução domina outra quando ela apresenta

desempenho melhor em pelo menos um critério, sem ser pior em nenhum dos demais.

Deste modo, o conjunto de todas as soluções não dominadas forma uma espécie de “curva de equilíbrio” no espaço das soluções, na qual qualquer melhoria em um objetivo implica necessariamente a piora de pelo menos outro. Esse conjunto é conhecido como Fronteira de Pareto.

Em termos matemáticos, esse conceito pode ser compreendido por duas condições fundamentais (COELLO, 2009):

1. Para qualquer solução x na fronteira P , não existe outra solução y tal que todas as funções objetivo de y sejam menores ou iguais às de x .
2. Deve existir pelo menos uma função objetivo em que y seja estritamente melhor que x .

Uma maneira intuitiva de compreender esse conceito é por meio de situações cotidianas que envolvem múltiplos critérios de escolha, como a seleção de um plano de internet. Suponha que uma pessoa deseje contratar um serviço considerando dois objetivos principais: velocidade de download e preço mensal. Planos com maior velocidade costumam ter um custo mais elevado, enquanto opções mais acessíveis tendem a oferecer uma conexão mais lenta. A Fronteira de Pareto, nesse contexto, seria composta pelos planos que proporcionam o melhor equilíbrio entre velocidade e preço, ou seja, aqueles que não são superados simultaneamente por nenhum outro plano nos dois critérios. Cada ponto da fronteira representa uma escolha eficiente, e a decisão final dependerá das prioridades individuais do usuário, uma vez que não existe uma única opção absolutamente superior (MARSON, 2017; VELDHUIZEN, 1998).

Nos Algoritmos Evolutivos Multiobjetivo, o principal desafio é encontrar ou se aproximar dessa fronteira, maximizando a diversidade e a qualidade das soluções ao longo do processo evolutivo. Desta forma, a Fronteira de Pareto não define uma única resposta ótima, mas um conjunto de soluções equivalentes em termos de eficiência, permitindo que a decisão final considere preferências específicas ou restrições contextuais.

2.2.2 Principais Algoritmos Evolutivos Multiobjetivos

Diferentemente das abordagens monoobjetivo (ou seja, com apenas um objetivo a ser tratado), os AEMOs trabalham com a otimização de diferentes objetivos, gerando um conjunto de soluções eficientes que aproximam-se à solução ótima (fronteira de Pareto). Entre os algoritmos mais consolidados destacam-se o NSGA-II (*Non-Dominated Sorting Genetic Algorithm II*), desenvolvido por (DEB, 2001), e o SPEA-II (*Strength Pareto Evolutionary Algorithm II*), proposto por (ZITZLER; LAUMANN; THIELE, 2001). Ambos se destacam pelo desempenho robusto em problemas com dois ou três objetivos. O NSGA-II organiza a população por níveis de não dominância e utiliza uma métrica de *crowding distance* para preservar a diversidade entre as soluções. O SPEA-II adota uma abordagem baseada em distâncias euclidianas entre os indivíduos, combinando dominância e densidade para avaliar a aptidão e conduzir a seleção evolutiva (DEB, 2001; ZITZLER; LAUMANN; THIELE, 2001).

Contudo, estudos indicam que algoritmos baseados exclusivamente em dominância podem apresentar limitações quando o número de objetivos aumenta, resultando em perda de diversidade e seleção ineficiente (DEB, 2001; ZITZLER; LAUMANN; THIELE, 2001). Os problemas que envolvem um número elevado de objetivos, quatro ou mais, são classificados como problemas de otimização com muitos objetivos (*many-objective optimization*) (COELLO, 2009). Nestes cenários, os métodos multiobjetivo tradicionais tornam-se menos eficazes, pois grande parte da população tende a se tornar não dominada, dificultando a distinção entre soluções e comprometendo o processo de seleção.

Os Algoritmos Evolutivos com Muitos Objetivos (MaOEAs, do inglês *Many-Objective Evolutionary Algorithms*) surgiram como uma resposta a essas limitações, incorporando estratégias específicas para lidar com a alta dimensionalidade do espaço de objetivos. Entre essas estratégias destacam-se abordagens baseadas em decomposição, vetores de referência, indicadores de qualidade e subpopulações (GOULART; CAMPELO, 2016).

Como alternativa, foram propostas abordagens como MOEA-D (ZHANG; LI, 2007) e NSGA-III (DEB; JAIN, 2014), a fim de lidar com muitos objetivos

(quatro ou mais). O MOEA-D decompõe o problema multiobjetivo em vários subproblemas escalares otimizados simultaneamente, promovendo a obtenção de soluções distribuídas ao longo da fronteira de Pareto. O NSGA-III estende o NSGA-II ao utilizar pontos de referência para orientar a seleção e manter a diversidade das soluções, sendo especialmente adequado para problemas com grande número de objetivos.

Neste contexto, pode-se destacar um método baseado em tabelas e subpopulações, denominado Algoritmo Evolutivo Multiobjetivo em Tabela (AEMT) (DELBEM, 2002), no qual diferentes subpopulações são mantidas simultaneamente, cada uma associada a um objetivo específico, permitindo maior diversidade e troca de características entre soluções ao longo do processo evolutivo (DELBEM, 2002). Inicialmente, este algoritmo foi proposto para lidar com poucos objetivos. No entanto, essa estratégia inspirou uma extensão para problemas com muitos critérios a serem considerados: o Algoritmo Evolutivo Multiobjetivo com Muitas Tabelas (AEMMT) (BRASIL, 2012; BRASIL; DELBEM; SILVA, 2013). Portanto, este AEMO é voltado a cenários com maior número de objetivos, cuja motivação central é melhorar concomitantemente a convergência à fronteira de Pareto e a diversidade das soluções, tornando esses algoritmos particularmente adequados para problemas complexos e de alta dimensionalidade (BRASIL; DELBEM; SILVA, 2013).

Nas próximas seções, serão aprofundados o Algoritmo Evolutivo Multiobjetivo em Tabela (AEMT), e sua variante, o Algoritmo Evolutivo Multiobjetivo com Muitas Tabelas (AEMMT).

2.2.3 Algoritmo Evolutivo Multiobjetivo em Tabela

O Algoritmo Evolutivo Multiobjetivo em Tabela (AEMT) foi proposto por Delbem (2002) como uma alternativa aos algoritmos evolutivos multiobjetivo baseados exclusivamente em dominância de Pareto, baseada na manutenção de múltiplas subpopulações organizadas em tabelas, em que cada uma é associada a um objetivo específico do problema.

O funcionamento do AEMT baseia-se na decomposição do problema mul-

tiobjetivo em vários problemas monoobjetivo independentes, tratados de forma simultânea. Para cada função objetivo, é criada uma subpopulação (tabela) responsável por armazenar os indivíduos mais bem avaliados em relação àquele critério. O processo de seleção para reprodução pode ocorrer a partir de qualquer uma das subpopulações, permitindo que características vantajosas de uma tabela migrem para outras ao longo das gerações. Essa estratégia favorece a diversidade genética e amplia a exploração do espaço de busca, reduzindo a probabilidade de convergência prematura para ótimos locais (SANTOS et al., 2010; GABRIEL; MELO; DELBEM, 2012).

Outro aspecto relevante do AEMT é o seu mecanismo de inserção e substituição de indivíduos nas tabelas. Um novo indivíduo é incorporado a uma subpopulação caso apresente desempenho superior a pelo menos um dos indivíduos já armazenados naquela tabela, sendo possível que o mesmo indivíduo pertença a múltiplas subpopulações simultaneamente. Essa característica diferencia o AEMT de outros AEMOs tradicionais e contribui para uma representação mais rica das soluções eficientes. Como resultado, o algoritmo apresenta maior capacidade de aproximar a Fronteira de Pareto, especialmente em problemas com múltiplos critérios conflitantes.

2.2.4 Algoritmo Evolutivo Multiobjetivo com Muitas Tabelas

O Algoritmo Evolutivo Multiobjetivo com Muitas Tabelas (AEMMT) surge como uma extensão natural do AEMT (BRASIL, 2012), desenvolvido com o objetivo de lidar com problemas que envolvem um número ainda maior de funções objetivo (quatro ou mais), característicos da otimização com muitos objetivos. Nessas situações, mesmo algoritmos multiobjetivo tradicionais tornam-se insuficientes, exigindo mecanismos adicionais para preservar a diversidade, manter a pressão seletiva e garantir boa convergência. O AEMMT foi inicialmente proposto para a predição *ab initio* de estruturas tridimensionais de proteínas com modelo de representação *full atom*², em que múltiplas funções de energia conflitantes deviam ser minimizadas simultaneamente (BRASIL; DELBEM; SILVA, 2013).

² O modelo *full atom* trata explicitamente todos os átomos da cadeia proteica (CUI; CHEN; WONG, 1998).

A principal inovação do AEMMT reside na organização da população em diversas tabelas, cada uma associada a diferentes combinações de objetivos, além de uma tabela específica para soluções não dominadas. Diferentemente do AEMT, que considera subpopulações ligadas a objetivos individuais, o AEMMT permite que tabelas representem subconjuntos de dois ou mais objetivos ponderados, ampliando a capacidade de capturar características relevantes entre critérios. O critério de inserção dos indivíduos nas tabelas baseia-se em métricas referentes a cada objetivo ou em dominância, dependendo da variante adotada, favorecendo soluções que apresentem bom desempenho em conjunto.

Dessa forma, essa arquitetura possibilita uma exploração mais abrangente do espaço de busca, ao mesmo tempo em que preserva a diversidade das soluções e melhora a aproximação da Fronteira de Pareto em problemas de alta dimensionalidade. Resultados reportados na literatura indicam que o AEMMT é capaz de lidar eficientemente com mais de dez objetivos simultâneos (BRASIL, 2012), superando abordagens clássicas e apresentando desempenho superior em diferentes domínios de aplicação, como no problema de predição de estruturas proteicas com modelo *full atom* (BRASIL; DELBEM; SILVA, 2013), no roteamento de pacotes (LAFETÁ, 2016) e em problemas de escalonamento de tarefas e roteamento de veículos (MACRI; BRASIL, 2025).

Assim, o AEMMT se apresenta como uma abordagem promissora, adequada e escalável para a resolução de problemas multiobjetivo e *many-objective*.

2.3 Considerações finais

Neste capítulo, foram apresentados os principais conceitos teóricos que fundamentam este trabalho, com ênfase nos Algoritmos Evolutivos e em suas extensões voltadas à resolução de problemas de otimização multiobjetivo e com muitos objetivos. Inicialmente, foram discutidos os princípios básicos desses algoritmos, destacando seus mecanismos inspirados em processos naturais de evolução, como seleção, cruzamento e mutação, que permitem a exploração eficiente de espaços de busca complexos. Em seguida, abordou-se a otimização multiobjetivo, com destaque para o conceito de dominância entre soluções e a definição da Fronteira de

Pareto.

Além disso, foram examinadas as principais limitações dos métodos tradicionais de otimização quando aplicados a problemas com múltiplos critérios, especialmente em cenários com grande número de objetivos (quatro ou mais). Entre essas limitações, destacam-se a perda de pressão seletiva e a dificuldade em manter a diversidade das soluções, fatores que podem comprometer a qualidade da aproximação da Fronteira de Pareto. Esses fundamentos teóricos fornecem a base necessária para a metodologia proposta e para as abordagens adotadas nos capítulos subsequentes deste trabalho.

3 O problema de predição de estrutura de proteína

Neste capítulo, serão abordados os principais conceitos relacionados ao problema alvo deste Trabalho de Conclusão de Curso, que reside na área da predição de estruturas de proteínas, incluindo os métodos computacionais de predição, o modelo de representação da proteína, as funções de energia, bem como os trabalhos relacionados ao tema de estudo.

3.1 Fundamentos da predição de estrutura de proteínas

A Predição de Estrutura de Proteínas (PSP, do inglês *Protein Structure Prediction*) é um dos problemas mais desafiadores da Biologia Molecular e da Ciência da Computação (DILL et al., 2007). Trata-se de um problema combinatório de alta complexidade, classificado como NP-Completo, cujo objetivo é determinar a estrutura tridimensional funcional de uma proteína a partir de sua sequência de aminoácidos (BRANDEN; TOOZE, 1999). A resolução deste problema é fundamental, pois a estrutura de uma proteína está diretamente ligada à sua função biológica. Logo, o conhecimento dessas estruturas permite o desenvolvimento de novos fármacos e a compreensão de diversas doenças.

Embora existam métodos laboratoriais para determinar a estrutura proteica, como a Cristalografia de Raio X e a Ressonância Nuclear Magnética (RNM), eles são processos caros, lentos e com limitações técnicas, o que justifica a crescente busca por métodos computacionais de otimização (LIMA, 2006; CUSTODIO, 2008; GABRIEL; MELO; DELBEM, 2012; BRASIL; DELBEM; SILVA, 2013; DIAS, 2016; CAVALCANTI, 2018; ESPITIA; PANG; GUMBART, 2024).

A seguir, serão apresentadas os diferentes tipos de estruturas da proteína e suas definições.

3.1.1 Estrutura de Proteínas

Proteínas são macromoléculas biológicas essenciais, formadas por um conjunto de 20 tipos de aminoácidos ligados entre si por ligações peptídicas (NELSON; COX, 2017). Os aminoácidos compartilham uma estrutura básica formada por um grupo amino, um grupo carboxila, um hidrogênio e um grupo R variável, responsável pelas propriedades químicas de cada aminoácido. A Figura 4 apresenta a estrutura geral de um aminoácido.

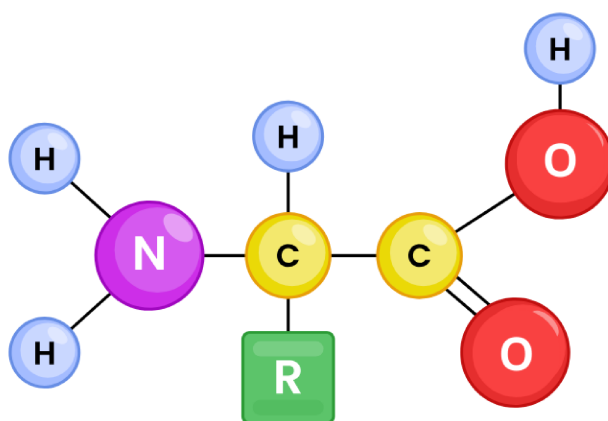


Figura 4 – Estrutura geral de um aminoácido, com destaque para o grupo R, responsável pelas propriedades específicas de cada aminoácido (NELSON; COX, 2017).

A estrutura de uma proteína pode ser classificada em quatro níveis hierárquicos:

- **Estrutura Primária:** é a sequência linear de aminoácidos que compõe a cadeia polipeptídica, sendo a informação mais fundamental, definindo a identidade da proteína. Na Figura 5, pode-se observar um exemplo de uma estrutura primária de uma proteína.
- **Estrutura Secundária:** refere-se aos arranjos locais da cadeia de aminoácidos, estabilizados por pontes de hidrogênio. As formas mais comuns são as alfa-hélices (espirais) e as folhas-beta (segmentos dobrados em zigue-zague). A Figura 6 ilustra as duas estruturas secundárias.

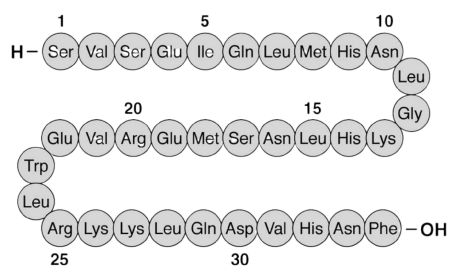
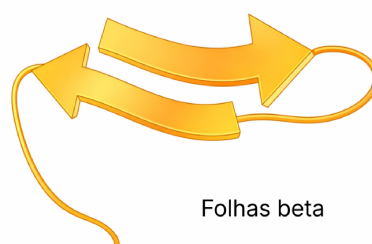


Figura 5 – Estrutura primária de uma proteína.



(a) Folhas-beta



(b) Alfa-hélice

Figura 6 – Estrutura secundária de proteínas: (A) Folhas-beta e (B) Alfa-hélice.

- **Estrutura Terciária:** corresponde ao dobramento completo da cadeia polipeptídica no espaço, formando sua estrutura tridimensional funcional. Esta conformação é determinada por interações de longa distância entre os aminoácidos, especialmente as interações hidrofóbicas, onde os resíduos que têm aversão à água (hidrofóbicos) tendem a se agrupar no núcleo da proteína. A Figura 7 mostra uma proteína na sua estrutura terciária, com o dobramento das suas estruturas.
- **Estrutura Quaternária:** ocorre em proteínas compostas por mais de uma cadeia polipeptídica (subunidades), descrevendo como essas subunidades se

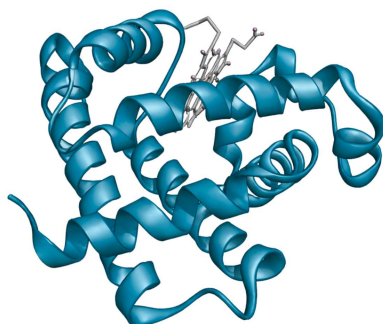


Figura 7 – Estrutura terciária de uma proteína.

organizam e interagem para formar um complexo proteico funcional.



Figura 8 – Estrutura quaternária de uma proteína.

3.1.2 Métodos Computacionais de Predição

Os métodos computacionais de PSP buscam prever a estrutura terciária a partir da primária. A abordagem *ab initio* (ou “por primeiros princípios”) é uma das mais importantes, pois utiliza apenas a sequência de aminoácidos como dado de entrada, sem depender de estruturas previamente conhecidas. Esses métodos baseiam-se na hipótese termodinâmica de que a estrutura nativa de uma proteína corresponde ao seu estado de menor energia livre global. Deste modo, o desafio é mapeado como um problema de otimização: encontrar a conformação que minimiza uma função de energia.

Devido à imensa complexidade e ao vasto espaço de busca conformacional, modelos simplificados são amplamente utilizados para tornar o problema computacionalmente tratável, como o Modelo Hidrofóbico-Polar (LAU; DILL, 1989). Diferentes modelos de representação estrutural têm sido propostos na literatura com o objetivo de reduzir o custo computacional do problema. Entre esses modelos, também destacam-se o *lattice*, o *off-lattice* e o *full-atom*, descritos a seguir:

- O modelo *lattice* (UNGER; MOULT, 1993) é o mais simplificado, representando cada aminoácido como um único ponto posicionado em uma malha discreta, o que reduz significativamente o espaço de busca e preserva interações essenciais entre resíduos adjacentes.
- O modelo *off-lattice* (ROCHA; PINTO; CALIRI, 2004) permite maior liberdade geométrica ao representar as conformações em espaço contínuo, aproximando-se de estruturas mais realistas, porém com aumento considerável do custo computacional.
- Os modelos *full-atom* (CUI; CHEN; WONG, 1998) descrevem explicitamente todos os átomos da cadeia proteica, proporcionando alta fidelidade física, mas tornando sua aplicação inviável em experimentos de larga escala.

Nesse contexto, modelos simplificados como o Hidrofóbico-Polar (HP), formulados sobre representações *lattice*, apresentam-se como uma alternativa eficiente para estudos computacionais, especialmente quando combinados com Algoritmos Evolutivos (BRASIL; DELBEM; SILVA, 2013; SHAKHNOVICH et al., 1991; SCHULZE-KREMER, 1993).

3.1.3 Modelo Hidrofóbico-Polar (HP)

O Modelo Hidrofóbico-Polar (HP), proposto por (LAU; DILL, 1989), é um dos modelos (*lattice*) mais estudados para a PSP. Sua principal simplificação é a redução do alfabeto de 20 aminoácidos para apenas dois tipos: H (hidrofóbico, apolar) e P (polar, hidrofílico). A hidrofobicidade é uma propriedade físico-química associada à tendência de determinadas moléculas ou grupos químicos evitarem o

contato com a água, preferindo interagir com moléculas apolares. No contexto das proteínas, a hidrofobicidade está relacionada principalmente às cadeias laterais dos aminoácidos, que podem apresentar caráter apolar ou polar. Aminoácidos hidrofóbicos tendem a se agrupar no interior da estrutura proteica durante o processo de dobramento, minimizando sua exposição ao solvente aquoso, enquanto resíduos polares e carregados permanecem preferencialmente na superfície da proteína. Esse efeito é considerado uma das principais forças motrizes do dobramento proteico e da estabilidade estrutural das proteínas, desempenhando papel central em modelos simplificados de predição estrutural, como o modelo Hidrofóbico-Polar (HP) (NELSON; COX, 2017). A estrutura da proteína é representada como uma cadeia autoevitante de resíduos H e P dispostos sobre os vértices de uma malha regular, que pode ser bidimensional (quadrática) ou tridimensional (cúbica).

Nesse modelo, o dobramento é guiado principalmente pelas interações entre resíduos hidrofóbicos, que tendem a se agrupar no interior da estrutura formando um núcleo compacto e reduzindo o contato com o solvente aquoso. Como ilustrado na Figura 9.

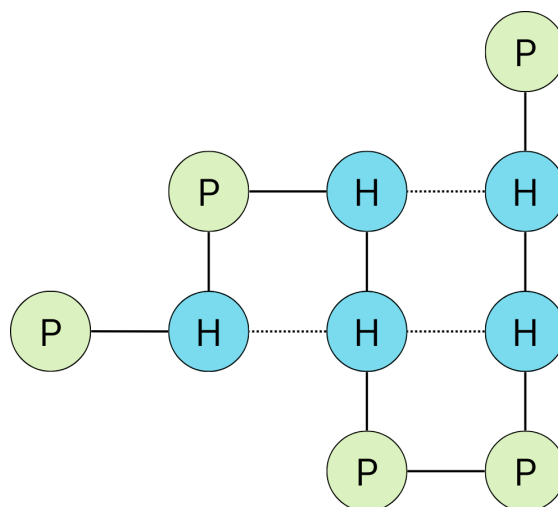


Figura 9 – Representação do Modelo Hidrofóbico-Polar (HP) em uma malha bidimensional, mostrando resíduos hidrofóbicos (H) e polares (P). (DILL, 1985)

Uma extensão natural do modelo HP bidimensional é o modelo HP em três dimensões (HP-3D), no qual a cadeia proteica é representada sobre uma malha

cúbica tridimensional. Nesse caso, cada aminoácido ocupa um vértice da rede e a cadeia deve formar um caminho autoevitante, ou seja, nenhum vértice pode ser ocupado por mais de um resíduo. A representação tridimensional aumenta significativamente o espaço de busca e a complexidade computacional do problema (BERGER; LEIGHTON, 1998). Apesar dessa complexidade, o modelo HP-3D continua sendo amplamente utilizado em estudos computacionais por capturar, de forma simplificada, o efeito hidrofóbico que direciona a formação do núcleo proteico durante o processo de dobramento. A Figura 10 apresenta uma representação do modelo HP em uma malha cúbica tridimensional, na qual resíduos hidrofóbicos (H) e polares (P) ocupam posições discretas nos vértices da rede.

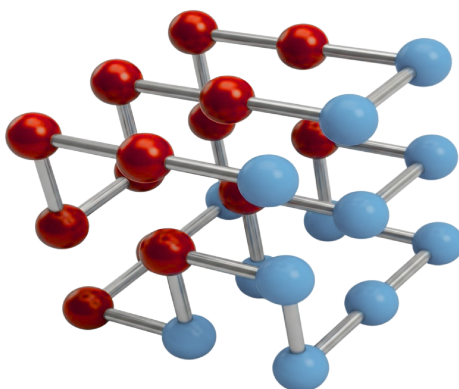


Figura 10 – Representação do modelo Hidrofóbico-Polar (HP) em uma malha cúbica tridimensional (HP-3D). Os resíduos hidrofóbicos (H) são representados pelas esferas vermelhas, enquanto os resíduos polares (P) são representados pelas esferas azuis (BERGER; LEIGHTON, 1998).

Portanto, este modelo destaca-se por oferecer um equilíbrio entre simplicidade computacional e representatividade estrutural. Embora simplifique a complexidade das interações físico-químicas reais ao considerar apenas dois tipos de aminoácidos (hidrofóbicos e polares), o modelo HP-3D permite capturar um dos principais fatores que influenciam o dobramento proteico: as interações hidrofóbicas.

Além disso, ao utilizar uma representação tridimensional em malha discreta, o modelo HP-3D aproxima-se mais da estrutura espacial real das proteínas quando comparado a modelos bidimensionais, possibilitando uma análise mais fiel

das conformações possíveis.

3.1.4 Funções de Energia no Modelo HP

A estabilidade de uma conformação no Modelo HP pode ser avaliada por uma função de energia que quantifica as interações favoráveis. A energia de uma estrutura é calculada com base no número de contatos entre aminoácidos hidrofóbicos (H-H) que não são adjacentes na sequência primária, mas são vizinhos topológicos na malha.

O objetivo da otimização é encontrar a conformação que maximiza o número de contatos H-H, o que equivale a minimizar a energia livre da estrutura. A energia de conformação, conforme a clássica energia de [Lau e Dill \(1989\)](#), pode ser expressa como o negativo do número de contatos H-H.

Recentemente, outras funções de energia foram propostas para refinar a avaliação. A energia simplificada ([ZHANG; WU; WANG, 2013](#)), por exemplo, calcula a distância euclidiana entre todos os aminoácidos hidrofóbicos não consecutivos na malha, buscando conformações mais compactas. Além disso, abordagens multiobjetivo consideram simultaneamente o número de contatos H-H e o grau de compactação da estrutura (medido pela distância entre os resíduos H), permitindo uma avaliação mais robusta das soluções geradas.

Na próxima seção, serão apresentados trabalhos relacionados ao tema de pesquisa deste Trabalho de Conclusão de Curso.

3.2 Trabalhos Relacionados

Os trabalhos correlatos ao tema de estudo deste Trabalho de Conclusão de Curso serão apresentados nesta seção em ordem cronológica, e não por ordem de importância, com uma breve descrição do objetivo de cada um.

Um dos trabalhos relevantes na literatura é o de [Cutello, Narzisi e Nicosia \(2005\)](#), que propõem uma abordagem evolutiva multiobjetivo para o problema de predição da estrutura de proteínas. Nesse estudo, os autores argumentam que a modelagem tradicional do problema como uma otimização monoobjetivo pode

ser limitada, uma vez que diferentes componentes da função de energia podem apresentar comportamentos conflitantes. Assim, o problema é reformulado como um problema de otimização multiobjetivo, permitindo explorar simultaneamente múltiplos critérios energéticos e obter um conjunto de soluções não dominadas que representam diferentes conformações viáveis da proteína. A busca no espaço conformacional é realizada por meio de algoritmos evolutivos, visando aproximar a fronteira de Pareto e capturar a diversidade de estruturas possíveis próximas ao estado nativo.

Em continuidade a essa linha de pesquisa, [Delbem \(2002\)](#) investiga o uso de algoritmos evolucionários para problemas complexos de otimização, explorando estratégias de representação, avaliação e busca adaptadas a domínios com grande espaço de soluções. Neste estudo foi introduzido o AEMT (Algoritmo Multiobjetivo em Tabelas) para aplicações em sistemas reais de distribuição elétrica, ratificando a importância de tratar problemas de alta complexidade com uma abordagem multiobjetivo. Este algoritmo é a primeira abordagem multiobjetivo com tabelas específicas para cada objetivo a ser tratado, estendendo-se posteriormente ao AEMMT.

Seguindo o interesse por abordagens bioinspiradas, [Cutello et al. \(2007\)](#) apresentaram um Algoritmo Imune (IA), inspirado no princípio de seleção clonal, para solucionar o PSP em modelos de grade 2D e 3D. O método utiliza operadores de hipermutação e um mecanismo de envelhecimento para manter a diversidade da população e evitar a convergência prematura. Testado em modelos clássicos (como o *HP model*), o IA demonstrou ser competitivo, indicando que o uso conjunto dos operadores imunológicos contribui para uma exploração mais eficiente do espaço de busca.

No que se refere à otimização com muitos objetivos (quatro ou mais), [Zhang e Li \(2007\)](#) propuseram o algoritmo MOEA/D (*Multiobjective Evolutionary Algorithm based on Decomposition*), que é um algoritmo evolutivo multiobjetivo que resolve o problema dividindo-o em vários subproblemas escalarizados. Cada subproblema corresponde a uma combinação de pesos dos objetivos e é otimizado simultaneamente. As soluções associadas a subproblemas vizinhos compartilham informação durante a evolução, o que melhora a busca.

A utilização de algoritmos evolutivos multiobjetivo em problemas complexos tem se mostrado uma abordagem eficaz, especialmente ao lidar com múltiplos critérios conflitantes. Neste cenário, tem-se o trabalho de [Brasil \(2012\)](#), que propôs uma extensão do AEMT: o Algoritmo Evolutivo Multiobjetivo com Muitas Tabelas (AEMMT) para a predição *ab initio* de estruturas proteicas. A metodologia adota múltiplas subpopulações para otimizar combinações de funções de energia conflitantes, além de uma tabela de não-dominância, obtendo sucesso na predição de estruturas complexas com folhas- β .

Outra importante contribuição para o PSP com o Modelo HP veio do trabalho de [Gabriel, Melo e Delbem \(2012\)](#). Para superar a limitação do modelo que foca apenas em contatos hidrofóbicos, o estudo aplicou o AEMT, otimizando simultaneamente o número de contatos H-H e o grau de compactação da estrutura. Os experimentos demonstraram que a abordagem alcançou resultados equivalentes ou superiores a algoritmos tradicionais, apresentando maior robustez e menor número de avaliações da função objetivo.

[Dias \(2016\)](#) apresentou um estudo comparativo de algoritmos de otimização para a predição de estruturas de proteínas. Este trabalho adotou o Modelo HP com AEs e Algoritmos de Colônia de Formiga, implementando operadores genéticos e mecanismos para tratamento de soluções inviáveis. Os experimentos demonstraram que a utilização de heurísticas evolutivas associadas ao modelo HP proporcionou resultados relevantes na identificação de estruturas mais estáveis.

Diversas abordagens baseadas em metaheurísticas têm sido propostas para resolver o problema de predição da estrutura de proteínas no modelo HP-3D. Nesse contexto, [Mansour, Kanj e Khachfe \(2011\)](#) propuseram um algoritmo genético aprimorado que incorpora heurísticas específicas para reparar soluções inviáveis geradas durante o cruzamento e garantir que as mutações produzam estruturas válidas, guiando a busca para regiões mais promissoras do espaço conformacional.

De forma semelhante, [Lin e Su \(2011\)](#) apresentaram uma abordagem híbrida que combina Algoritmos Genéticos com Otimização por Enxame de Partículas (PSO), utilizando o PSO principalmente no operador de mutação para ajustar dinamicamente as direções de dobramento com base nas melhores soluções

encontradas, além de empregar mecanismos de busca local para refinamento das conformações.

Guo et al. (2017) introduziram um método híbrido que integra PSO, *Tabu Search* e uma estratégia denominada *pulling strategy*, utilizada para evitar convergência prematura e favorecer a formação de estruturas mais compactas. Os resultados apresentados nesses trabalhos indicaram que tais abordagens são capazes de alcançar energias mínimas competitivas para diversas sequências de teste do modelo HP.

Além das metaheurísticas evolutivas, outras estratégias também têm sido exploradas para o problema. Traykov et al. (2018) propuseram um algoritmo heurístico para o dobramento de proteínas em malha cúbica tridimensional que divide a sequência de aminoácidos em segmentos menores, gerando diferentes possibilidades de dobramento e selecionando iterativamente as conformações com maior número de contatos hidrofóbicos. Essa estratégia permite construir progressivamente estruturas completas a partir de soluções parciais promissoras, obtendo resultados competitivos quando comparados a métodos como algoritmos genéticos e Monte Carlo. Mais recentemente, técnicas baseadas em aprendizado profundo também têm sido aplicadas ao problema.

Nesse contexto, Espitia, Pang e Gumbart (2024) apresentaram duas arquiteturas baseadas em *Deep Reinforcement Learning*, incluindo um modelo baseado em *reservoir computing* combinado com redes neurais profundas e uma arquitetura baseada em LSTM, capaz de capturar dependências de longo alcance entre aminoácidos. Os experimentos mostraram que essas abordagens conseguem atingir os melhores valores de energia conhecidos para diferentes sequências do modelo HP, além de apresentar melhorias na eficiência de treinamento em comparação com métodos tradicionais.

Dentre as abordagens mais recentes, Liu e Iba (2025) exploraram o uso de Deep Reinforcement Learning, combinando uma Deep Q-Network com mecanismos de atenção baseados em Transformers para modelar o processo de dobramento como uma sequência de decisões, permitindo capturar dependências espaciais complexas e alcançar resultados competitivos em benchmarks do modelo HP-3D. Além

disso, trabalhos recentes de revisão destacam a crescente integração de diferentes paradigmas computacionais, incluindo heurísticas clássicas, aprendizado por reforço, modelos generativos e até algoritmos quânticos, evidenciando que o modelo HP continua sendo amplamente utilizado como um ambiente de teste para o desenvolvimento e avaliação de novas técnicas (Wang, Boyuan, 2026).

A partir dos trabalhos analisados, é possível perceber a relevância do problema de predição de estruturas de proteínas, tanto do ponto de vista biológico quanto computacional, em razão de sua elevada complexidade e do vasto espaço de busca envolvido. Observa-se também a ampla aplicação de algoritmos evolutivos e de suas diferentes variantes na resolução de problemas complexos de otimização, incluindo aqueles relacionados diretamente ao PSP. Essas abordagens têm se mostrado particularmente adequadas para lidar com a natureza combinatória e multimodal do problema, explorando estratégias diversas de busca, representação e avaliação das soluções.

Os trabalhos de Mansour, Kanj e Khachfe (2011), Lin e Su (2011), Brasil (2012), Gabriel, Melo e Delbem (2012), Guo et al. (2017), Traykov et al. (2018), Espitia, Pang e Gumbart (2024) foram utilizados como base para o desenvolvimento deste trabalho, principalmente as investigações de Brasil (2012) e Gabriel, Melo e Delbem (2012), uma vez que forneceram referências teóricas sobre o problema de predição de estruturas de proteínas e aplicação de algoritmos evolutivos multiobjetivo.

3.3 Considerações finais

Neste capítulo, foi apresentado o problema da Predição de Estrutura de Proteínas, destacando sua relevância na bioinformática e na biologia molecular, assim como os desafios computacionais envolvidos em sua resolução. Enfatizou-se a elevada complexidade do problema, decorrente do grande espaço de busca associado às possíveis conformações que uma proteína pode assumir, o que torna inviável a utilização de abordagens determinísticas. Também foram discutidos os principais conceitos relacionados às proteínas, incluindo sua composição e estrutura, fornecendo o embasamento necessário para a compreensão dos métodos computacionais

aplicados à predição das mesmas.

Adicionalmente, foram descritos diferentes modelos de representação estrutural utilizados na literatura, com destaque para o Modelo Hidrofóbico-Polar (HP), adotado neste trabalho por sua simplicidade e eficiência em capturar aspectos essenciais do dobramento proteico. Os conceitos e fundamentos apresentados estabeleceram a base teórica para a metodologia proposta, bem como para a interpretação das análises e resultados nos capítulos subsequentes. Por fim, realizou-se uma revisão de trabalhos relacionados, destacando as abordagens computacionais mais relevantes que influenciaram diretamente o desenvolvimento deste Trabalho de Conclusão de Curso.

4 AEMMT para o problema PSP no modelo HP-3D

Este capítulo apresentará a descrição dos algoritmos implementados neste Trabalho de Conclusão de Curso, sendo três Algoritmos Evolutivos monoobjetivo e dois Algoritmos Evolutivos Multiobjetivo com Muitas Tabelas (AEMMTs), ambos aplicados ao problema de predição de estruturas de proteínas, utilizando o modelo HP-3D. O objetivo metodológico consistiu em investigar o comportamento de diferentes estratégias evolutivas, comparando essas abordagens, sob o ponto de vista dos valores de aptidão obtidos e a taxa de acerto.

Todos os algoritmos foram desenvolvidos na linguagem Python e compartilham a mesma estrutura básica de representação, operadores genéticos e parâmetros, diferenciando-se principalmente pela formulação das funções objetivo e pelos critérios de seleção e sobrevivência.

O código-fonte completo do projeto (em linguagem Python versão 3), bem como as instruções necessárias para a reprodução dos experimentos, encontram-se disponíveis publicamente no repositório GitHub, acessível por meio do seguinte link: <<https://github.com/paulaprado1904/AlgoritmoEvolutivo>>.

A seguir, serão descritos os tópicos fundamentais de cada algoritmo desenvolvido.

4.1 Representação do indivíduo e do espaço de busca

Os indivíduos foram representados da mesma maneira nos cinco algoritmos implementados. Cada indivíduo da população equivale a uma possível conformação (solução) de uma dada proteína no Modelo HP em malha cúbica (3D). A representação adotada foi a codificação absoluta (KRANE; RAYMER, 2003), na qual o cromossomo é descrito por uma sequência de movimentos unitários no espaço. Essa escolha é amplamente empregada em abordagens evolutivas para modelos em

rede, pois permite mapear diretamente um cromossomo em uma estrutura 3D de forma determinística, com baixo custo computacional, facilitando a aplicação de operadores genéticos e a avaliação de conformações ao longo das gerações.

O cromossomo é composto por $L = n - 1$ genes, em que n corresponde ao número de resíduos da sequência HP. Cada gene assume um valor inteiro no intervalo $[0, 5]$, representando um deslocamento unitário em uma das seis direções possíveis da malha cúbica (por exemplo, cima/baixo, esquerda/direita e frente/trás). A partir dessa sequência, as coordenadas cartesianas dos resíduos são reconstruídas de modo incremental, iniciando-se da origem $(0, 0, 0)$, e cada novo gene define a posição do próximo resíduo na malha.

Uma consequência direta da codificação absoluta é que o cromossomo não impede, a priori, movimentos de retorno ou a formação de trajetórias que resultem em autointerseções (colisões), isto é, quando dois ou mais resíduos ocupam a mesma coordenada. Em vez de eliminar tais indivíduos antecipadamente, este trabalho adota a estratégia de permitir soluções inviáveis na população, contabilizando-se o número de colisões (n_C) durante a reconstrução da conformação. Essa decisão metodológica mantém a diversidade genética e evita restringir excessivamente o espaço de busca, ao custo de introduzir a necessidade de penalização na função objetivo. Assim, soluções com colisões podem ser exploradas nas fases iniciais do processo evolutivo, mas tendem a ser gradualmente não selecionadas devido à redução de aptidão associada a valores elevados de n_C .

4.2 Avaliação dos indivíduos

A avaliação de um indivíduo consiste em reconstruir a conformação tridimensional correspondente ao cromossomo e, a partir dessa estrutura, calcular métricas de qualidade relacionadas à viabilidade geométrica, ao grau de compactação e à estabilidade no Modelo HP-3D (GABRIEL; MELO; DELBEM, 2012). Para garantir consistência entre os diferentes algoritmos implementados, foi adotado um *pipeline* de avaliação, no qual métricas subsequentes dependem diretamente da correta obtenção das coordenadas cartesianas dos resíduos.

4.2.1 Pipeline de avaliação e ordem de cálculo

Dado um indivíduo c (sequência de movimentos), a avaliação segue as etapas:

1. **Reconstrução das coordenadas:** a partir da origem $(0, 0, 0)$, a sequência de movimentos é aplicada incrementalmente para obter as coordenadas de todos os resíduos. Essa etapa produz a lista de posições ocupadas na malha, que é utilizada nas etapas seguintes.
2. **Deteção de colisões e cálculo de inviabilidade:** durante a reconstrução, verifica-se se uma coordenada já foi ocupada anteriormente. O número de colisões $n_C(c)$ é computado como o total de posições com ocupação superior a uma unidade. Essa métrica caracteriza a inviabilidade geométrica da solução.
3. **Cálculo de contatos hidrofóbicos:** Inicialmente avalia-se o número de colisões $n_C(c)$. Caso não haja colisões ($n_C(c) = 0$), calcula-se o número de contatos hidrofóbicos $n_H(c)$, definidos como pares de resíduos do tipo H que não são consecutivos na cadeia e que ocupam posições adjacentes na malha (distância de Manhattan igual a 1). Quando há colisões ($n_C(c) > 0$), o termo $n_H(c)$ é definido como zero. Essa estratégia reduz o custo computacional, evitando o cálculo de contatos hidrofóbicos para conformações inviáveis.

O valor $n_H(c)$ é utilizado como medida de qualidade estrutural, pois conformações mais compactas tendem a apresentar maior número de contatos H-H. Assim, a aptidão do indivíduo c é definida por:

$$f(c) = n_H(c) - \rho \cdot n_C(c), \quad (4.1)$$

onde $n_H(c)$ representa o número de contatos hidrofóbicos da conformação c , $n_C(c)$ corresponde ao número de colisões detectadas durante a reconstrução da estrutura, e ρ é um parâmetro de penalização positivo. Neste trabalho, adotou-se $\rho = 1$.

Como o objetivo físico do modelo HP é minimizar a energia da conformação, enquanto a função de aptidão foi definida para maximizar a qualidade

estrutural, estabelece-se uma relação inversa entre essas grandezas. Assim, a energia livre associada à conformação é definida como o oposto do valor da aptidão, conforme a Equação 4.2.

$$E(c) = -f(c) \quad (4.2)$$

Dessa forma, conformações com maior número de contatos hidrofóbicos e menor número de colisões apresentam maior aptidão e, conseqüentemente, menor energia, caracterizando estruturas mais estáveis.

4. **Cálculo de métricas geométricas de compactação:** independentemente de colisões, são calculadas distâncias entre resíduos hidrofóbicos não consecutivos, permitindo estimar o grau de compactação. A partir do conjunto de distâncias euclidianas entre pares H–H, são obtidas a distância máxima $d_{\max}(c)$ e a distância média $d_{\text{avg}}(c)$.
5. **Energia simplificada:** além da função baseada no número de contatos hidrofóbicos, foi adotada uma função de energia simplificada, originalmente proposta na literatura para o Modelo HP, com o objetivo de avaliar o grau de compactação estrutural de forma contínua.

Sejam $H = \{r_1, r_2, \dots, r_k\}$ os resíduos hidrofóbicos da cadeia, representados por suas coordenadas cartesianas tridimensionais $r_i = (x_i, y_i, z_i)$. A energia simplificada da conformação c é definida como a soma das distâncias euclidianas entre todos os pares de resíduos hidrofóbicos não consecutivos na sequência primária:

$$E_s(c) = \sum_{i < j} d(r_i, r_j), \quad (4.3)$$

onde

$$d(r_i, r_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2} \quad (4.4)$$

corresponde à distância euclidiana entre os resíduos r_i e r_j .

6. **Cálculo de $\delta(c)$ (função ponderada):** utiliza-se a medida $\delta(c)$ para relacionar aptidão e compactação geométrica. Para isso, combina-se o valor de aptidão derivado do indivíduo com uma métrica de distância ($d_{\max}$ ou d_{avg}), podendo a métrica ativa ser alterada ao longo das gerações conforme o comportamento populacional, definida por:

$$\delta(c) = \frac{f(c)}{d(c)}, \quad (4.5)$$

em que $f(c)$ corresponde à aptidão do indivíduo, calculada conforme a Equação 4.1, e $d(c)$ representa uma medida de compactação geométrica da conformação. Neste trabalho, $d(c)$ pode assumir o valor de $d_{\max}(c)$ ou $d_{\text{avg}}(c)$, dependendo da métrica ativa na geração corrente.

A escolha dinâmica entre $d_{\max}(c)$ e $d_{\text{avg}}(c)$ é realizada com base na variação populacional observada entre gerações consecutivas, intensificando a pressão seletiva sobre o critério estrutural que melhor diferencia os indivíduos naquele momento, ou seja, aquele que mais variou.

4.2.2 Dependência entre métricas

As métricas adotadas possuem dependências naturais. O cálculo de $n_H(c)$, por exemplo, depende de uma conformação **factível** (sem colisões), pois a presença de sobreposição de resíduos compromete a interpretação física dos contatos H–H. Por outro lado, as métricas geométricas ($d_{\max}(c)$ e $d_{\text{avg}}(c)$) e a energia simplificada dependem da reconstrução das coordenadas, mas podem ser computadas mesmo quando há colisões, sendo utilizadas como informação auxiliar para guiar a busca e manter pressão seletiva sobre compactação.

4.2.3 Manutenção de soluções inviáveis e penalização

Diferentemente de abordagens que eliminam indivíduos inviáveis ou aplicam mecanismos explícitos de correção de colisões, neste trabalho adotou-se a estratégia de manter soluções inviáveis na população, aplicando uma penalização proporcional ao número de colisões observadas. Essa escolha tem duas motivações

principais: (i) preservar a diversidade genética, evitando a redução prematura do espaço explorado; e (ii) reduzir o custo computacional associado a procedimentos de reparo, mantendo a avaliação consistente entre variantes monoobjetivo e multiobjetivo.

A penalização atua diretamente sobre a aptidão do indivíduo, reduzindo sua probabilidade de seleção em métodos baseados em torneio. Consequentemente, indivíduos com valores elevados de $n_C(c)$ tendem a ser descartados ao longo das gerações, enquanto indivíduos factíveis são favorecidos. Ao mesmo tempo, permitir a presença temporária de soluções inviáveis pode auxiliar na exploração de regiões promissoras do espaço de busca nas fases iniciais, contribuindo para evitar convergência prematura e permitindo que operadores genéticos recombinaem partes de cromossomos potencialmente úteis até que a população se concentre em soluções factíveis e mais compactas.

4.3 Algoritmos Evolutivos monoobjetivo

Foram implementadas três variantes de Algoritmos Evolutivos monoobjetivo, todas baseadas na mesma estrutura evolutiva, diferenciando-se pela função objetivo adotada:

1. **AE baseado em contatos hidrofóbicos** (n_H), no qual a aptidão do indivíduo corresponde ao número de contatos hidrofóbicos H–H, sendo esse valor maximizado ao longo das gerações ($f(c)$).
2. **AE baseado em energia simplificada**, no qual a função objetivo busca minimizar a soma das distâncias euclidianas entre resíduos hidrofóbicos não consecutivos, incorporando penalizações em casos de colisão estrutural ($E_s(c)$).
3. **AE híbrido**, que considera simultaneamente o número de contatos hidrofóbicos e métricas geométricas de compactação estrutural, mantendo a avaliação do indivíduo expressa por um único valor escalar de aptidão ($\delta(c)$).

4.3.1 Fluxo evolutivo dos AEs monoobjetivo

Em todos os algoritmos evolutivos avaliados, o processo de busca segue um ciclo iterativo de gerações. Cada indivíduo é representado por um cromossomo composto por uma sequência de movimentos, que define a conformação 3D da proteína no reticulado. A cada geração, os indivíduos são avaliados pela função objetivo do método, e então são aplicados operadores de seleção, recombinação, mutação e sobrevivência para compor a população seguinte.

Inicialização e avaliação

A população inicial é construída de forma aleatória, gerando-se $|\mathcal{P}|$ indivíduos com cromossomos válidos (sequência de movimentos). Em seguida, cada indivíduo é avaliado, obtendo-se suas medidas estruturais e seu valor de aptidão (por exemplo, energia, n_H , ou métricas de distância). Nos métodos baseados em critérios dinâmicos (ex.: δ), as estatísticas populacionais necessárias são atualizadas a cada geração para orientar a avaliação relativa entre indivíduos.

Seleção de pais

Para gerar descendentes, os pais são selecionados por torneio a partir da população corrente. Em cada seleção, são amostrados aleatoriamente k indivíduos da população e o indivíduo com melhor valor de aptidão é escolhido como pai (ou mãe). Esse procedimento é repetido para selecionar o segundo progenitor. Neste trabalho, a comparação no torneio é feita com base no critério principal do algoritmo (por exemplo, δ no caso do AE- δ).

Cruzamento

O cruzamento é aplicado com probabilidade p_c por meio de um operador de *crossover multipontos*. Nesse operador, o cromossomo é dividido em n segmentos por meio da definição de múltiplos pontos de corte ao longo da sequência de movimentos. O número de pontos de corte é definido proporcionalmente ao tamanho da sequência de resíduos, sendo dado por $n = \text{round}(|s|/10)$, onde $|s|$ representa o

comprimento da sequência. Esse valor é limitado de forma a garantir pelo menos um ponto de corte e no máximo $L - 1$, onde L é o tamanho do cromossomo.

Após a definição dos pontos de corte, os filhos são gerados alternando os segmentos herdados do pai e da mãe, produzindo dois descendentes com combinações distintas de material genético. Caso o operador de recombinação não seja aplicado, os descendentes são gerados como cópias diretas dos pais.

Mutação

Após a geração dos descendentes, aplica-se o operador de mutação por gene sobre cada indivíduo gerado. Nesse operador, cada posição do cromossomo possui uma probabilidade p_m de ser alterada para um novo movimento aleatório. A mutação é utilizada como mecanismo de exploração do espaço de busca, permitindo introduzir variações que não seriam obtidas apenas por recombinação.

Entretanto, neste trabalho a mutação é aplicada somente quando o descendente gerado não apresenta desempenho superior ao melhor indivíduo global encontrado até o momento. Caso o indivíduo já represente uma melhoria em relação ao melhor global, ele é mantido sem modificação.

Avaliação dos descendentes

Após crossover e mutação, cada descendente é avaliado novamente, de forma consistente com a função objetivo do algoritmo em questão, produzindo o valor de aptidão utilizado na etapa de sobrevivência.

Seleção de sobreviventes (estratégia $\mu + \lambda$)

A população da geração seguinte é formada pela combinação da população corrente (μ indivíduos) com os descendentes gerados (λ indivíduos), caracterizando uma estratégia do tipo $\mu + \lambda$. Neste trabalho, utiliza-se elitismo, preservando-se diretamente uma taxa dos melhores indivíduos. Em seguida, todos esses indivíduos são avaliados em conjunto, e apenas os μ melhores são selecionados para compor a próxima geração. Esse mecanismo permite que soluções de alta qualidade sejam

preservadas mesmo quando os descendentes apresentam desempenho inferior, caracterizando um modelo com maior pressão seletiva (GABRIEL; DELBEM, 2008).

Critérios de parada e repetição dos experimentos

O processo evolutivo se repete até que um critério de parada seja satisfeito, como: (i) atingir o limite máximo de avaliações da função objetivo; ou (ii) atingir um valor-alvo de energia/aptidão.

4.4 Algoritmos Evolutivos Multiobjetivo com Muitas Tabelas

Além das abordagens monoobjetivo, foram implementados duas abordagens do AEMMT (BRASIL, 2012). Nessas abordagens, o processo de seleção considera simultaneamente múltiplos critérios de avaliação, sem a necessidade de agregá-los em uma única função. As duas versões são diferenciadas pelo número de subpopulações (tabelas) utilizadas e pelos critérios de armazenamento dos indivíduos.

Diante dessas características, a adoção do AEMMT neste trabalho justifica-se pela necessidade de tratar simultaneamente diferentes aspectos estruturais relevantes para a qualidade das conformações proteicas no modelo HP tridimensional. A predição estrutural envolve critérios conflitantes – como maximização de contatos hidrofóbicos, aumento da compactação geométrica e minimização de energia – cuja agregação em uma única função escalar pode introduzir vieses decorrentes da escolha de pesos e reduzir a capacidade de exploração do espaço de busca.

Nesse contexto, a organização da população em subpopulações especializadas permite direcionar a pressão seletiva para diferentes propriedades estruturais de forma independente e complementar. Essa estratégia favorece a manutenção de diversidade, reduz a perda de soluções potencialmente promissoras e possibilita uma exploração mais equilibrada do espaço conformacional. Além disso, a estrutura baseada em tabelas oferece flexibilidade para incorporar novos critérios de avaliação sem necessidade de reformular completamente o processo evolutivo, o que é particularmente relevante em problemas complexos e com muitas soluções como a predição de estruturas proteicas (DELBEM, 2002; ISHIBUCHI; TSUKAMOTO; NOJIMA, 2008; SANTOS et al., 2010).

Assim, a utilização do AEMMT neste estudo permite investigar de maneira mais abrangente a interação entre múltiplos critérios estruturais e avaliar como sua consideração simultânea influencia a qualidade das soluções obtidas, a dinâmica evolutiva e a convergência do processo de busca.

AEMMT com 5 tabelas

Na primeira variante, a população é organizada em cinco tabelas, cada uma associada a um critério específico de avaliação:

- Tabela 1: maximização do número de contatos hidrofóbicos $n_H(c)$;
- Tabela 2: minimização da métrica de compactação $d_{\max}(c)$;
- Tabela 3: minimização da métrica de compactação $d_{\text{avg}}(c)$;
- Tabela 4: maximização da função ponderada $\delta(c)$;
- Tabela 5: tabela de não-dominância.

Essa configuração permite intensificar simultaneamente diferentes aspectos estruturais, combinando aptidão clássica baseada em contatos hidrofóbicos com métricas geométricas de compactação.

AEMMT com 6 tabelas

Na segunda variante, foi adicionada uma tabela dedicada exclusivamente à energia simplificada $E_s(c)$, resultando na seguinte estrutura:

- Tabela 1: maximização de $n_H(c)$;
- Tabela 2: minimização de $d_{\max}(c)$;
- Tabela 3: minimização de $d_{\text{avg}}(c)$;
- Tabela 4: maximização da função ponderada $\delta(c)$;
- Tabela 5: minimização da energia simplificada $E_s(c)$.

- Tabela 6: tabela de não-dominância.

A inclusão desta tabela visa ampliar a pressão seletiva sobre critérios relacionados à compactação contínua da estrutura, permitindo que soluções que apresentem boa distribuição espacial dos resíduos hidrofóbicos, mesmo sem número máximo de contatos discretos, sejam preservadas e propagadas ao longo das gerações.

4.4.1 Fluxo evolutivo dos AEMMTs

Nos algoritmos AEMMT, o processo evolutivo segue uma dinâmica baseada em múltiplas tabelas de avaliação, cada uma associada a um critério estrutural específico. Essas tabelas atuam como subpopulações especializadas, permitindo explorar simultaneamente diferentes características das soluções durante o processo evolutivo.

Inicialização da população

Inicialmente, a população é gerada de forma aleatória, produzindo um conjunto de indivíduos representados por sequências de movimentos no reticulado 3D. Cada indivíduo é então avaliado, sendo calculados seus atributos estruturais, como número de contatos hidrofóbicos (n_H), colisões (n_C) e demais métricas utilizadas pelos critérios das tabelas.

Após a avaliação, os indivíduos são inseridos nas tabelas correspondentes aos critérios considerados pelo algoritmo. Cada tabela mantém um número fixo de indivíduos. A inserção ocorre somente quando o indivíduo apresenta desempenho superior ao pior indivíduo atualmente presente na tabela para o critério associado. Nesse caso, o novo indivíduo é inserido e o pior elemento da tabela é removido.

Seleção de pais

A geração de descendentes ocorre por meio da seleção de dois pais. O primeiro progenitor é escolhido aleatoriamente dentre todos os indivíduos presentes nas tabelas. O segundo progenitor é selecionado por meio de torneio baseado no

critério δ , considerando o conjunto de indivíduos disponíveis em todas as tabelas. Nesse torneio, são amostrados k indivíduos aleatoriamente, sendo escolhido aquele com maior valor de δ .

Essa estratégia combina exploração e intensificação do processo evolutivo, permitindo que indivíduos promissores tenham maior probabilidade de reprodução, sem restringir a diversidade populacional.

Cruzamento

O operador de recombinação segue o procedimento descrito nas versões monoobjetivo.

Mutação

O operador de mutação segue o mesmo procedimento descrito nas versões monoobjetivo.

Atualização das tabelas

Após *crossover* e possível mutação, os descendentes são avaliados e podem ser inseridos nas tabelas, seguindo o mesmo critério utilizado na inicialização. Se o valor do atributo correspondente à tabela for melhor que o pior indivíduo armazenado, o novo indivíduo é inserido e o pior elemento é removido, mantendo-se constante o tamanho da tabela.

Esse processo é repetido iterativamente até que seja atingido o critério de parada definido para o experimento, como o número máximo de avaliações da função objetivo ou a obtenção de um valor alvo de energia.

4.5 Considerações Finais

Neste capítulo, foi apresentada a metodologia adotada para a modelagem computacional do problema de predição de estruturas de proteínas no modelo HP tridimensional. Foram definidos os elementos centrais do processo evolutivo,

incluindo a representação dos indivíduos, o procedimento de reconstrução das conformações e o tratamento de soluções inviáveis por meio de penalização.

Também foi descrito o processo de avaliação das soluções, contemplando o cálculo das principais métricas estruturais utilizadas no trabalho, como contatos hidrofóbicos, energia, medidas de compactação geométrica e energia simplificada, bem como suas relações de dependência dentro do processo de avaliação.

Além disso, foram caracterizadas diferentes estratégias evolutivas, incluindo abordagens monoobjetivo e métodos multiobjetivo baseados em tabelas, organizados em subpopulações associadas a distintos critérios de avaliação. Essas estratégias permitem analisar o comportamento do processo evolutivo sob diferentes configurações de otimização estrutural.

A organização metodológica apresentada estabelece a base para a realização dos experimentos e para a comparação sistemática das abordagens propostas. No próximo capítulo, serão descritos os experimentos conduzidos e analisados os resultados obtidos a partir da aplicação dos algoritmos desenvolvidos.

5 Resultados

Neste capítulo serão apresentados os resultados obtidos a partir dos experimentos realizados com diferentes abordagens de Algoritmos Evolutivos (três versões do AE monoobjetivo e duas do AEMMT), as quais foram aplicadas ao problema de predição de estruturas de proteínas no modelo HP-3D.

Nestes experimentos, cada implementação do algoritmo foi executada 50 vezes para cada instância do problema. O desempenho foi avaliado com base na melhor energia obtida, energia média, desvio padrão e taxa de acerto em relação à energia mínima conhecida. Para a execução dos experimentos, utilizou-se a plataforma de computação em nuvem Google Colaboratory (Google Colab), que permite a execução de códigos em ambiente remoto sem a necessidade de configuração local.

5.1 Configuração Experimental

Os experimentos foram conduzidos utilizando dois conjuntos de instâncias de *benchmarks* do modelo Hidrofóbico-Polar tridimensional (HP-3D), ambas empregadas na literatura para avaliação de algoritmos evolutivos aplicados ao problema de predição de estruturas de proteínas (MANSOUR; KANJ; KHACHFE, 2011; LIN; SU, 2011; GUO et al., 2017; TRAYKOV et al., 2018; ESPITIA; PANG; GUMBART, 2024; GABRIEL; MELO; DELBEM, 2012). Foram consideradas instâncias de menor tamanho, uma vez que experimentos preliminares indicaram elevado custo computacional, entre 10 e 15 minutos por instância, o que limitou a utilização de sequências maiores nos estágios de análise.

Tabela 1 – Primeiro conjunto de instâncias do modelo HP-3D utilizadas nos experimentos (TRAYKOV et al., 2018; ESPITIA; PANG; GUMBART, 2024; LIN; SU, 2011).

Identificação	Tamanho	Sequência HP-3D
3d1	20	HPHPPHHHPHPHPHHPPHPH
3d2	24	HHPHPHPHPHPHPHPHPHPHH
3d3	25	PPHPPHHPPPPHHPPPPHHPPPH
3d4	36	PPPHHPHHPPPPPHHHHHHHHPHHPPPPHHPPHP

Tabela 2 – Segundo conjunto de instâncias do modelo HP-3D utilizadas nos experimentos (MANSOUR; KANJ; KHACHFE, 2011; GUO et al., 2017; ESPITIA; PANG; GUMBART, 2024; GABRIEL; MELO; DELBEM, 2012).

Identificação	Tamanho	Sequência HP
A1	27	PHPHPHHHPPHPHPPPPPPPPPHHP
A2	27	PHHPPPPPPPPPHHPPHHPPHPHPH
A3	27	HHHHPPPPPHPPPPHHHPPPPPPPPH
A4	27	HHHPHHHHPPPHHPHPHHPPHPPHH
A5	27	HHHHPPPPHPPHHPPHHPPPPPPPPP
A6	27	HPPPPPHPHHHPPHHPPHPPPPHPH
A7	27	HPPHPHHPPHPPPPPHPHHHPHPHH
A8	27	HPPPPPPPPPPPHHPPPPPPPPHHH
A9	27	PPPPPPHHHPPPHPHHPPHPPHPPP
A10	27	PPPPHHHPHPHPHPHPPHHHPHPPP
A11	27	HPPPPHHHHPPHPHPHHHPHPPHPPH

Os trabalhos de Gabriel, Melo e Delbem (2012), Espitia, Pang e Gumbart (2024) e Mansour, Kanj e Khachfe (2011) realizaram experimentos apenas com as dez primeiras instâncias do conjunto. Por essa razão, na Tabela 18 não há valores reportados para a sequência A11 nesses trabalhos.

A seguir, serão relatados os experimentos e seus respectivos resultados referentes aos AEs monoobjetivo.

5.2 Experimentos com Algoritmos Evolutivos monoobjetivo

Nas Tabelas 3, 4, 5, 6, 7 e 8 são mostrados os resultados obtidos nos testes realizados para análise do comportamento do Algoritmo Evolutivo monoobjetivo. Nesta etapa, para os três algoritmos seguintes desenvolvidos, foram utilizados os seguintes parâmetros: tamanho da população de 120, torneio de 2, número de máximo de avaliações de 1.500.000, taxa de mutação de 3% e taxa de *crossover* de 80%.

5.2.1 AE monoobjetivo com maximização de contatos H (n_H)

A seguir, serão apresentados os resultados obtidos a partir dos dois conjuntos distintos de instância para maximização de (n_H).

Tabela 3 – Resultados do AE monoobjetivo com função de aptidão de n_H para o primeiro conjunto de instâncias.

Seq.	Tamanho	Energia Mínima	AE n_H	Média	DP	Acertos (%)
3d1	20	-11	-11	-10.76	0.43	76
3d2	24	-13	-13	-11.82	0.91	32
3d3	25	-9	-9	-7.96	0.80	24
3d4	36	-18	-16	-13.10	1.37	0
Média						44

Tabela 4 – Resultados do AE monoobjetivo com função de aptidão de n_H para o segundo conjunto de instâncias.

Seq.	Tamanho	Energia Mínima	AE N_H	Média	DP	Acertos (%)
A1	27	-9	-9	-7.36	1.00	2
A2	27	-10	-10	-8.46	0.97	16
A3	27	-8	-8	-7.28	0.75	44
A4	27	-15	-15	-13.78	0.95	26
A5	27	-8	-8	-6.00	0.67	2
A6	27	-11	-11	-9.60	0.72	8
A7	27	-13	-13	-11.54	0.97	14
A8	27	-4	-4	-3.86	0.35	86
A9	27	-7	-7	-5.50	0.70	6
A10	27	-11	-10	-8.54	0.93	0
A11	27	-14	-15	-13.74	0.66	78
Média						25.64

Para o AE baseado na maximização do número de contatos hidrofóbicos (n_H), observa-se, conforme apresentado na Tabela 3, que o método foi capaz de atingir a energia mínima conhecida nas sequências 3d1, 3d2 e 3d3, enquanto para a sequência 3d4 a solução ótima não foi obtida. A taxa média de acerto foi de 44%, com aumento da variabilidade nas instâncias de maior comprimento.

Para as sequências com comprimento 27, cujos resultados estão apresentados na Tabela 4, o algoritmo apresentou menor taxa de acerto, com média de 25,64% comparado ao experimento anterior. Pode-se notar que a energia mínima foi atingida em 9 das 11 instâncias, e que houve maior dispersão nos resultados entre execuções independentes.

Por outro lado, observa-se um resultado relevante na sequência A11, na qual o algoritmo foi capaz de encontrar uma solução com energia $E = -15$, valor inferior ao mínimo reportado na literatura ($E = -14$).

5.2.2 AE monoobjetivo com função ponderada (δ)

A seguir, serão apresentados os resultados obtidos a partir dos dois conjuntos de instâncias com função ponderada (δ).

Tabela 5 – Resultados do algoritmo AE monoobjetivo baseado na função ponderada (δ) para o primeiro conjunto de instâncias.

Seq.	Tamanho	Energia Mínima	AE δ	Média	DP	Acertos (%)
3d1	20	-11	-11	-11.00	0.00	100
3d2	24	-13	-13	-13.00	0.00	100
3d3	25	-9	-9	-9.00	0.50	72
3d4	36	-18	-18	-16.04	1.10	16
Média						72

Tabela 6 – Resultados do algoritmo AE monoobjetivo baseado na função ponderada (δ) para o segundo conjunto de instâncias.

Seq.	Tamanho	Energia Mínima	AE δ	Média	DP	Acertos (%)
A1	27	-9	-9	-8.18	0.43	20
A2	27	-10	-10	-9.30	0.61	38
A3	27	-8	-8	-7.82	0.43	84
A4	27	-15	-15	-14.80	0.45	82
A5	27	-8	-8	-6.88	0.62	12
A6	27	-11	-12	-10.24	0.59	6
A7	27	-13	-13	-12.58	0.60	64
A8	27	-4	-4	-4.00	0.00	100
A9	27	-7	-7	-6.10	0.46	16
A10	27	-11	-11	-10.12	0.77	34
A11	27	-14	-16	-14.10	0.36	2
Média						41.64

Para o AE baseado na função ponderada (δ), observa-se, conforme indicado na Tabela 5, um desempenho superior ao teste anterior para as instâncias do conjunto testado. A energia mínima conhecida foi atingida em todas as sequências, com convergência consistente e baixa variabilidade nas execuções, ainda que a taxa de acerto tenha sido menor para a sequência de maior comprimento (3d4). A taxa média de acerto foi de 72%.

Para as sequências com comprimento 27, apresentadas na Tabela 6, o algoritmo apresentou convergência para a energia mínima em 9 de 11 instâncias com taxa média de acerto de 41,64%, indicando melhoria dos resultados comparando ao experimento anterior com as mesmas instâncias.

Para este algoritmo, do mesmo modo que o experimento anterior, também observa-se na sequência A11, que o algoritmo foi capaz de encontrar uma solução com energia $E = -16$, valor inferior ao mínimo reportado na literatura ($E = -14$).

5.2.3 AE monoobjetivo com energia simplificada (Es)

A seguir, serão apresentados os resultados obtidos a partir dos dois conjuntos de instâncias com energia simplificada (Es).

Tabela 7 – Resultados do algoritmo AE monoobjetivo baseado na energia simplificada (Es) para o primeiro conjunto de instâncias.

Seq.	Tamanho	Energia Mínima	AE Es	Média	DP	Acertos (%)
3d1	20	-11	-11	-10.92	0.27	92
3d2	24	-13	-13	-12.68	0.71	82
3d3	25	-9	-9	-8.40	0.67	50
3d4	36	-18	-16	-13.92	1.22	0
Média						56

Tabela 8 – Resultados do algoritmo AE monoobjetivo baseado na energia simplificada (Es) para o segundo conjunto de instâncias.

Seq.	Tamanho	Energia Mínima	AE Es	Média	DP	Acertos (%)
A1	27	-9	-9	-7.64	0.69	2
A2	27	-10	-10	-8.48	0.88	16
A3	27	-8	-8	-7.52	0.86	72
A4	27	-15	-15	-14.24	0.98	58
A5	27	-8	-8	-6.72	0.83	20
A6	27	-11	-12	-9.90	0.84	2
A7	27	-13	-13	-11.54	0.99	16
A8	27	-4	-4	-4.00	0.00	100
A9	27	-7	-7	-5.84	0.61	12
A10	27	-11	-11	-7.74	1.44	4
A11	27	-14	-14	-13.82	0.66	82
Média						34.91

Para o AE baseado na energia simplificada (Es), os resultados apresentados na Tabela 7 indicam que a energia mínima conhecida foi atingida nas sequências 3d1, 3d2 e 3d3, enquanto para a sequência 3d4 não houve convergência para a solução ótima. A taxa média de acerto foi de 56%, com aumento da variabilidade nas instâncias de maior comprimento. Esses valores mostraram que esta versão do AE mono obteve melhor resultado que com a função n_H e pior que com a função δ , que obteve em todas as instâncias a solução ótima.

Para as sequências com comprimento 27, cujos resultados estão descritos na Tabela 8, o algoritmo atingiu a energia mínima em 11 de 12 instâncias, porém com maior dispersão entre execuções. A taxa média de acerto foi de 34,91%. No-

vamente, esses valores indicaram uma superioridade da versão E_s sobre a função H e inferioridade em relação à função δ .

Note que na sequência A6, o algoritmo encontrou uma solução com energia $E = -12$, valor inferior ao mínimo reportado na literatura ($E = -11$).

5.2.4 Análise comparativa entre os AEs monoobjetivo

A comparação entre as três abordagens monoobjetivo evidencia diferenças significativas de desempenho entre as funções de avaliação utilizadas.

A função ponderada (δ) apresentou a maior taxa média de acerto e maior estabilidade entre execuções, indicando melhor capacidade de guiar o processo de busca em direção a soluções de baixa energia. A energia simplificada (E_s) apresentou desempenho intermediário, com melhoria em relação à maximização direta de contatos hidrofóbicos, porém com maior custo computacional e menor consistência em instâncias mais complexas.

Por sua vez, a maximização direta de contatos hidrofóbicos (n_H) apresentou as menores taxas de acerto e maior variabilidade nos resultados, especialmente para sequências de maior comprimento.

De modo geral, os resultados indicam que cada função objetivo contribui de forma distinta para o processo de busca, influenciando diferentes aspectos da convergência e da estabilidade das soluções obtidas. Essas diferenças sugerem que cada critério de avaliação captura características específicas da qualidade das conformações, impactando de maneira particular a eficiência do algoritmo evolutivo.

5.2.5 AE monoobjetivo de melhor desempenho (δ)

Com base nos resultados obtidos nos experimentos monoobjetivo apresentados anteriormente, foi realizado um novo teste utilizando o AE monoobjetivo com a função ponderada δ , que apresentou o melhor desempenho em média entre as abordagens analisadas até o momento. O objetivo desse experimento foi investigar se o aumento do tamanho da população poderia contribuir para a obtenção de melhores soluções pelo AE monoobjetivo, especialmente para proteínas mais

complexas.

Para isso, o tamanho da população foi ampliado para 400 com o intuito de aumentar a diversidade da população, permitindo uma exploração mais ampla do espaço de busca. Os resultados obtidos são apresentados nas Tabelas 9 e 10.

Tabela 9 – Resultados do AE monoobjetivo com função ponderada (δ) para o primeiro conjunto de instâncias.

Seq	Tamanho	Energia M	AE δ	Média	DP	Acertos (%)
3d1	20	-11	-11	-11	0	100
3d2	24	-13	-13	-12.98	0.14	98
3d3	25	-9	-9	-9	0	100
3d4	36	-18	-18	-16.66	0.98	26
Média						81,00

Tabela 10 – Resultados do AE monoobjetivo com função ponderada (δ) para o segundo conjunto de instâncias.

Seq	Tamanho	Energia M	AE δ	Média	DP	Acertos (%)
A1	27	-9	-9	-8.48	0.50	48
A2	27	-10	-10	-9.32	0.47	32
A3	27	-8	-8	-7.68	0.47	68
A4	27	-15	-15	-14.88	0.38	90
A5	27	-8	-8	-7.30	0.46	30
A6	27	-11	-12	-10.68	0.71	14
A7	27	-13	-13	-12.88	0.32	84
A8	27	-4	-4	-4	0	100
A9	27	-7	-7	-6.34	0.47	34
A10	27	-11	-11	-10.32	0.55	36
A11	27	-14	-16	-14.04	0.19	2
Média						48,91

Observa-se que, para a maioria das proteínas mais simples, houve melhora nos resultados obtidos. Entretanto, para algumas sequências mais complexas, como a proteína 3d4 e a sequência A11, o desempenho não apresentou melhorias significativas. Esses resultados indicaram que, embora o aumento da população possa favorecer a diversidade e melhorar o desempenho em casos mais simples, ele não é suficiente para lidar com instâncias mais complexas do problema, sendo necessária

a utilização de abordagens de busca mais eficientes. Por esse motivo, nos experimentos seguintes foram investigadas abordagens baseadas em algoritmos evolutivos multiobjetivo, uma vez que estas utilizam o critério de não-dominância para alcançar soluções que considerem simultaneamente muitos critérios de avaliação.

A seguir, serão descritos os resultados obtidos a partir das abordagens do AEMMT implementadas neste trabalho.

5.3 Experimentos com os AEMMTs

Nas Tabelas 11, 12, 13 e 14 são apresentados os resultados obtidos nos experimentos realizados para avaliar o comportamento dos algoritmos evolutivos multiobjetivo com muitas tabelas (AEMMT). Os testes foram conduzidos utilizando os mesmos conjuntos de instâncias empregados nos experimentos com os algoritmos monoobjetivo.

Para ambos os algoritmos foram utilizados os seguintes parâmetros: torneio de tamanho 2, número máximo de avaliações de 1.500.000, taxa de mutação de 10% e taxa de *crossover* de 95%.

5.3.1 AEMMT com 5 Tabelas

Nesta configuração foi utilizada uma população de tamanho 400, sendo dividida em cinco subpopulações de tamanho 80 cada. Cada tabela é responsável por priorizar um dos critérios considerados no processo de otimização:

- Tabela 1 – maximização de $n_H(c)$;
- Tabela 2 – minimização de $d_{\max}(c)$;
- Tabela 3 – minimização de $d_{\text{avg}}(c)$;
- Tabela 4 – maximização de $\delta(c)$;
- Tabela 5 – tabela de não-dominância.

A distribuição uniforme entre as tabelas busca equilibrar a influência de cada critério durante o processo evolutivo, promovendo diversidade estrutural nas soluções geradas e permitindo que diferentes características da conformação da proteína sejam exploradas simultaneamente. A quinta tabela corresponde a uma estrutura de não-dominância, responsável por armazenar indivíduos que não são dominados considerando simultaneamente os diferentes critérios avaliados pelo algoritmo.

Tabela 11 – Resultados do AEMMT 5 para o primeiro conjunto de instâncias.

Seq.	Tamanho	Energia Mínima	AEMMT 5	Média	DP	Acertos (%)
3d1	20	-11	-11	-11.00	0.00	100
3d2	24	-13	-13	-13.00	0.00	100
3d3	25	-9	-9	-8.96	0.19	96
3d4	36	-18	-18	-17.02	1.00	48
Média						86

Tabela 12 – Resultados do algoritmo AEMMT 5 para o segundo conjunto de instâncias.

Seq.	Tamanho	Energia Mínima	AEMMT 5	Média	DP	Acertos (%)
A1	27	-9	-9	-8.48	0.50	48
A2	27	-10	-10	-9.60	0.53	62
A3	27	-8	-8	-7.90	0.30	90
A4	27	-15	-15	-15.00	0.00	100
A5	27	-8	-8	-7.38	0.49	38
A6	27	-11	-12	-10.90	0.86	32
A7	27	-13	-13	-12.88	0.32	88
A8	27	-4	-4	-4.00	0.00	100
A9	27	-7	-7	-6.62	0.49	62
A10	27	-11	-11	-10.62	0.53	64
A11	27	-14	-16	-15.44	0.64	50
Média						67

Os resultados obtidos pelo algoritmo AEMMT 5 indicam elevado desempenho para o primeiro conjunto de instâncias. A energia mínima conhecida foi atingida em todas as sequências, com taxas de acerto de 100% para as instâncias 3d1 e 3d2, e 96% para 3d3. Para a instância 3d4, de maior tamanho, observa-se

redução na taxa de acerto (48%) neste experimento, indicando maior dificuldade de convergência para conformações ótimas em sequências mais longas.

Para o conjunto de instâncias com tamanho 27, o algoritmo também apresentou desempenho consistente em diversas sequências, atingindo a energia mínima conhecida em instâncias como A3, A4 e A8, com taxas de acerto elevadas. Destacam-se, em particular, as instâncias A4 e A8, nas quais o algoritmo apresentou taxa de acerto de 100% e desvio padrão nulo, evidenciando elevada estabilidade.

Observa-se ainda que, para as sequências A6 e A11, o algoritmo foi capaz de encontrar soluções com energia $E = -12$ e $E = -16$, respectivamente, valores inferiores aos mínimos reportados na literatura para essas instâncias ($E = -11$ e $E = -14$). Esse resultado sugere a obtenção de conformações energeticamente mais favoráveis do que as previamente documentadas para essas sequências.

Por outro lado, algumas instâncias apresentaram maior variabilidade nos resultados, como A5 e A6, refletindo a maior dificuldade do processo de otimização para determinadas configurações estruturais. A taxa média geral de acerto para esse conjunto de instâncias foi de aproximadamente 67%.

5.3.2 AEMMT com 6 Tabelas

Nesta configuração foi utilizada uma população de tamanho 480, dividida em seis subpopulações de tamanho 80 cada. As quatro primeiras tabelas seguem os mesmos critérios utilizados no algoritmo anterior, enquanto as duas tabelas adicionais introduzem novos mecanismos de avaliação das soluções.

A quinta tabela incorpora o critério baseado na energia simplificada $E_s(c)$, enquanto a sexta tabela corresponde a uma estrutura de não-dominância, responsável por armazenar indivíduos que não são dominados considerando simultaneamente os diferentes critérios do problema.

- Tabela 1 – maximização de $n_H(c)$;
- Tabela 2 – minimização de $d_{\max}(c)$;
- Tabela 3 – minimização de $d_{\text{avg}}(c)$;

- Tabela 4 – maximização de $\delta(c)$;
- Tabela 5 – minimização da energia simplificada $E_s(c)$;
- Tabela 6 – tabela de não-dominância.

Tabela 13 – Resultados do algoritmo AEMMT 6 para o primeiro conjunto de instâncias.

Seq.	Tamanho	Energia Mínima	AEMMT 6	Média	DP	Acertos (%)
3d1	20	-11	-11	-11.00	0.00	100
3d2	24	-13	-13	-12.98	0.14	98
3d3	25	-9	-9	-9.00	0.00	94
3d4	36	-18	-18	-17.04	1.06	52
Média						86

Tabela 14 – Resultados do algoritmo AEMMT 6 para o segundo conjunto de instâncias.

Seq.	Tamanho	Energia Mínima	AEMMT 6	Média	DP	Acertos (%)
A1	27	-9	-9	-8.42	0.49	42
A2	27	-10	-10	-9.48	0.57	52
A3	27	-8	-8	-7.82	0.38	82
A4	27	-15	-15	-14.94	0.23	94
A5	27	-8	-8	-7.38	0.53	40
A6	27	-11	-12	-10.76	0.84	26
A7	27	-13	-13	-12.86	0.35	86
A8	27	-4	-4	-4.00	0.00	100
A9	27	-7	-7	-6.60	0.49	60
A10	27	-11	-11	-10.62	0.53	64
A11	27	-14	-16	-15.68	0.63	56
Média						64

Os resultados obtidos pelo algoritmo AEMMT 6 para o primeiro conjunto de instâncias são apresentados na Tabela 13. Observa-se que a energia mínima conhecida foi atingida em todas as sequências analisadas. Para as instâncias de menor tamanho, como 3d1, 3d2 e 3d3, o algoritmo apresentou desempenho elevado, com taxas de acerto de 100%, 98% e 94%, respectivamente, além de baixos valores de desvio padrão.

Para a sequência 3d4, de maior tamanho, observa-se maior dificuldade de convergência, refletida na taxa de acerto de 52% e no desvio padrão mais elevado (1,06) neste experimento. Ainda assim, a energia mínima conhecida foi atingida, indicando que o algoritmo foi capaz de encontrar conformações ótimas, embora com menor frequência quando comparado às instâncias menores. A taxa média de acerto para esse conjunto foi de aproximadamente 86%.

Os resultados para o segundo conjunto de instâncias, apresentados na Tabela 14, mostram maior variabilidade no desempenho do algoritmo. Algumas sequências apresentaram elevadas taxas de acerto e grande estabilidade, como A8, que atingiu 100% de acerto e desvio padrão nulo, e A4, com taxa de acerto de 94% e baixo desvio padrão.

Em outras instâncias, entretanto, observa-se maior dificuldade de convergência, como em A1, A2 e A5, cujas taxas de acerto foram inferiores a 60%. Destaca-se ainda que, para as sequências A6 e A11, o algoritmo foi capaz de encontrar soluções com energia $E = -12$ e $E = -16$, respectivamente, valores inferiores aos mínimos reportados na literatura para essas instâncias ($E = -11$ e $E = -14$). Esse resultado sugere a possibilidade de existência de conformações energeticamente mais favoráveis para essas sequências.

De forma geral, a taxa média de acerto obtida para esse conjunto de instâncias foi de aproximadamente 64%, evidenciando maior dificuldade do problema à medida que aumenta o tamanho e a complexidade das sequências analisadas.

5.3.3 Análise Comparativa entre AEMMT 5 e AEMMT 6

A comparação entre as configurações AEMMT 5 e AEMMT 6 permite avaliar o impacto do aumento do número de tabelas no desempenho do algoritmo.

Para o primeiro conjunto de instâncias, ambas as configurações apresentaram desempenho semelhante, com taxa média de acerto de aproximadamente 86%. Entretanto, para o segundo conjunto de instâncias observa-se que o AEMMT 5 apresentou desempenho superior, com taxa média de acerto de 67%, enquanto o AEMMT 6 obteve 64%.

Observa-se ainda que, para a sequência A11, ambos os algoritmos foram

capazes de encontrar soluções com energia $E = -16$, valor inferior ao mínimo reportado na literatura consultada ($E = -14$). Esse resultado pode indicar a obtenção de conformações energeticamente mais favoráveis.

Esse comportamento pode estar relacionado ao aumento da diversidade de soluções introduzido pela inclusão de tabelas adicionais, em especial pela presença da tabela de não-dominância. Essa estrutura tende a preservar indivíduos que representam diferentes atributos entre os objetivos considerados, ampliando o conjunto de soluções mantidas ao longo do processo evolutivo.

Embora esse mecanismo contribua para aumentar a diversidade estrutural da população, ele também pode reduzir a pressão seletiva do algoritmo em direção ao ótimo energético. Como consequência, a busca pode tornar-se mais dispersa, favorecendo indivíduos que são competitivos em alguns critérios, mas que não necessariamente conduzem mais rapidamente à convergência para as melhores soluções.

Além disso, é possível que alguns dos critérios considerados nas diferentes tabelas não estejam fortemente correlacionados entre si. Nesse caso, a otimização simultânea desses objetivos pode direcionar a busca para regiões distintas do espaço de soluções, dificultando a convergência para conformações energeticamente ótimas.

Assim, para o conjunto de instâncias analisado neste trabalho, a configuração com cinco tabelas apresentou desempenho ligeiramente superior e suficiente para as instâncias analisadas nesse trabalho, indicando que um equilíbrio entre diversidade e pressão seletiva é fundamental para o bom desempenho do algoritmo.

5.4 Análise comparativa entre métodos

Nesta seção, serão apresentadas uma análise comparativa entre os AEs implementados neste trabalho, bem como uma relação destes com diferentes métodos encontrados na literatura nesta área de pesquisa.

5.4.1 Análise comparativa entre os AEs implementados

Conforme apresentado na Tabela 15, os algoritmos multiobjetivo apresentaram desempenho superior aos métodos monoobjetivo, obtendo maiores taxas médias de acerto para ambos os conjuntos de proteínas.

Entre as abordagens monoobjetivo, a função ponderada (δ) apresentou o melhor desempenho geral. Entre as configurações com muitas tabelas, o AEMMT 5 apresentou o melhor desempenho geral, enquanto o AEMMT 6 não apresentou melhoria significativa, apesar da inclusão de uma tabela adicional.

Tabela 15 – Comparação geral entre os algoritmos implementados.

Algoritmo	Acertos 3d (%)	Acertos 27 (%)	Avaliações médias	Observação
AE N_H	44.00	25.64	153.310	menor taxa de convergência entre todos
AE δ	72.00	41.64	116.835	melhor desempenho dos monoobjetivos
AE Es	56.00	34.91	207.750	maior custo computacional entre os mono-objetivos
AEMMT 5	86.00	66.73	148.277	melhor desempenho geral
AEMMT 6	85.50	63.82	191.859	tabela adicional sem melhoria significativa

5.4.2 Análise comparativa dos AEs com trabalhos relacionados

Diversas abordagens têm sido propostas para o PSP no modelo HP 3D, incluindo algoritmos evolutivos, métodos híbridos e, mais recentemente, técnicas baseadas em aprendizado profundo.

A Tabela 16 apresenta uma síntese dos principais trabalhos relacionados que abordam o problema de predição de estruturas de proteínas (PSP) no modelo HP-3D. Observa-se que diferentes estratégias têm sido empregadas na literatura, incluindo entre esses métodos o *Enhanced Genetic Algorithm* (EGA) proposto por Mansour, Kanj e Khachfe (2011), o algoritmo híbrido HGA-PSO de Lin e Su (2011), o método baseado em PSO com estratégias adicionais proposto por Guo et al. (2017), bem como abordagens mais recentes baseadas em redes neurais

profundas, como LSTM-A e FFNN-R.

A Tabela 17 apresenta a comparação dos métodos para o primeiro conjunto de instâncias. Observa-se que todos os algoritmos avaliados foram capazes de alcançar os valores da energia mínima (EM) conhecida para essas instâncias. A Tabela 18 apresenta a comparação dos valores mínimos de energia obtidos por diferentes algoritmos para o segundo conjunto de instâncias. Observa-se que, neste caso, a maioria dos algoritmos avaliados foi capaz de atingir o valor de referência (EM). Entretanto, em alguns casos específicos, não foram obtidos os valores mínimos conhecidos, como na sequência A6 em que foi alcançado uma energia menor que a conhecida em 5 dos 6 métodos considerados; e na A11, em que não há valor reportado em alguns métodos. Isto ocorre porque os respectivos autores não realizaram o experimento para essa instância.

O algoritmo AEMMT com cinco tabelas, proposto neste trabalho, demonstrou desempenho consistente, atingindo os melhores valores de energia reportados para todas as sequências analisadas.

De forma geral, os resultados demonstram que o AEMMT apresenta desempenho competitivo em relação aos métodos existentes, sendo capaz de atingir os valores mínimos de energia conhecidos para todas as instâncias avaliadas. Além disso, a abordagem multiobjetivo baseada em tabelas permite explorar diferentes critérios de avaliação simultaneamente, contribuindo para um processo de busca mais equilibrado entre exploração e intensificação no espaço de soluções.

Tabela 16 – Principais trabalhos relacionados para PSP no modelo HP-3D.

Autor	Algoritmo	Tipo de otimização
Mansour, Kanj e Khachfe (2011)	Enhanced GA (EGA)	Algoritmo evolutivo
Lin e Su (2011)	HGA-PSO	Algoritmo evolutivo + PSO
Guo et al. (2017)	PSO + Tabu + Pulling (TPPSO)	Metaheurística híbrida
Traykov et al. (2018)	Extended Heuristic Algorithm (EHA)	Algoritmo heurístico
Espitia, Pang e Gumbart (2024)	FFNN-R	Aprendizado profundo
Espitia, Pang e Gumbart (2024)	LSTM-A	Aprendizado profundo
Gabriel, Melo e Delbem (2012)	AEMT	AE multiobjetivo
Este trabalho	AEMMT 5	AE multiobjetivo

Tabela 17 – Comparação da energia mínima obtida por diferentes algoritmos para o primeiro conjunto de instâncias.

Seq.	Tamanho	EM	EHA	LSTM-A	HGA-PSO	AEMMT 5
3d1	20	-11	-11	-11	-11	-11
3d2	24	-13	-13	-13	-13	-13
3d3	25	-9	-9	-9	-9	-9
3d4	36	-18	-18	-18	-18	-18

Tabela 18 – Comparação de energia mínima obtida por diferentes algoritmos para o segundo conjunto de instâncias.

Seq.	Tamanho	EM	TPPSO	FFNN-R	EGA	AEMT	AEMMT 5
A1	27	-9	-9	-9	-9	-9	-9
A2	27	-10	-10	-10	-10	-10	-10
A3	27	-8	-8	-8	-8	-8	-8
A4	27	-15	-15	-15	-15	-15	-15
A5	27	-8	-8	-8	-8	-8	-8
A6	27	-11	-12	-11	-12	-12	-12
A7	27	-13	-13	-13	-13	-13	-13
A8	27	-4	-4	-4	-4	-4	-4
A9	27	-7	-7	-7	-7	-7	-7
A10	27	-11	-11	-11	-11	-11	-11
A11	27	-14	-16	–	–	–	-16

Nos experimentos realizados neste estudo, o AEMMT 5 demonstrou desempenho equivalente a diversos métodos reportados na literatura, atingindo os valores mínimos conhecidos para todas as instâncias avaliadas.

Além da análise quantitativa dos resultados, foi realizada uma análise inspirada na literatura (LIN; SU, 2011; ESPITIA; PANG; GUMBART, 2024) das conformações geradas pelo algoritmo para algumas das proteínas mais complexas avaliadas neste trabalho. As Figuras 11 e 12 apresentam exemplos de estruturas obtidas pelo AEMMT 5 para as sequências 3d4 e A11, respectivamente.

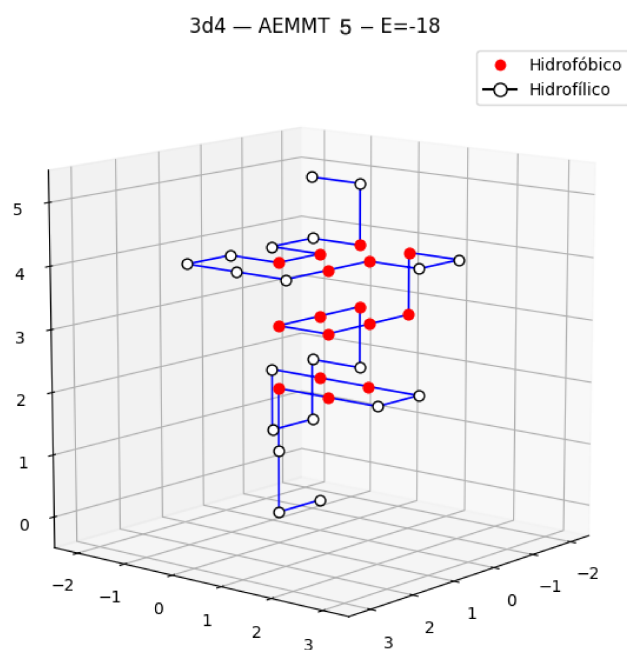


Figura 11 – Conformação obtida pelo AEMMT 5 para a sequência 3d4.

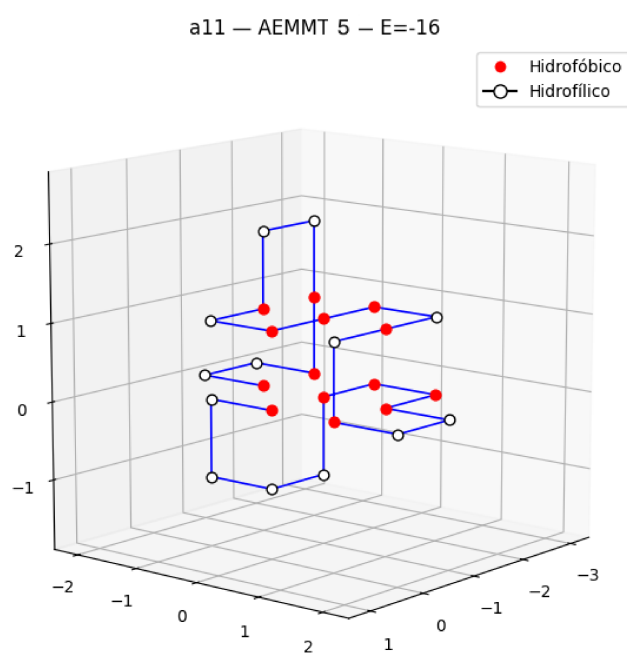


Figura 12 – Conformação obtida pelo AEMMT 5 para a sequência A11.

Observa-se que as estruturas geradas apresentam características compatíveis com os princípios bioquímicos descritos pelo modelo Hidrofóbico-Polar (HP). Em ambas as conformações, os aminoácidos hidrofóbicos tendem a se agrupar no interior da estrutura, formando um núcleo hidrofóbico compacto, enquanto os aminoácidos polares permanecem predominantemente expostos na superfície. Esse padrão reproduz o comportamento bioquímico observado em proteínas reais, nas quais resíduos hidrofóbicos tendem a evitar o contato com o solvente, ao passo que resíduos polares apresentam maior afinidade com o meio aquoso (NELSON; COX, 2017).

Dessa forma, os resultados obtidos indicam que o algoritmo AEMMT 5 não apenas alcança bons valores de energia, mas também gera estruturas coerentes com os conceitos bioquímicos associados ao dobramento de proteínas, corroborando o comportamento esperado para esse tipo de modelo.

5.5 Considerações Finais

Neste capítulo foram apresentados e analisados os resultados obtidos a partir dos experimentos realizados com diferentes algoritmos evolutivos aplicados ao problema de predição de estruturas de proteínas no modelo HP tridimensional. Inicialmente, foram avaliadas três abordagens monoobjetivo, baseadas nas funções de aptidão n_H , energia simplificada e função ponderada δ . De modo geral, observou-se que a utilização de funções objetivo mais informativas contribui para melhorar a qualidade das soluções obtidas e a manutenção do processo evolutivo.

Em seguida, foram analisadas as abordagens multiobjetivo baseadas no Algoritmo Evolutivo Multiobjetivo com Muitas Tabelas (AEMMT). Os resultados demonstraram que essa estratégia foi capaz de alcançar desempenho superior às abordagens monoobjetivo na maioria das instâncias avaliadas, obtendo maiores taxas de acerto e mantendo diversidade no processo evolutivo.

Entre as configurações avaliadas, o AEMMT com cinco tabelas apresentou o melhor desempenho geral para as instâncias consideradas neste trabalho, atingindo os valores mínimos de energia conhecidos para todas as sequências avaliadas, e superando algumas delas. A versão com seis tabelas, embora tenha apresentado

maior variabilidade evolutiva, não trouxe ganhos consistentes em termos de taxa de acerto para as instâncias analisadas.

Além disso, a comparação com trabalhos correlatos demonstrou que os resultados obtidos pelo algoritmo proposto são competitivos em relação aos métodos reportados na literatura, reforçando o potencial da abordagem multiobjetivo baseada em tabelas para o tratamento de problemas complexos de otimização combinatória, como o problema de predição de estruturas de proteínas no modelo HP tridimensional.

6 Conclusão

A predição da estrutura tridimensional de proteínas permanece como um problema desafiador da bioinformática computacional. Mesmo em modelos simplificados, como o modelo HP tridimensional, o espaço de busca cresce rapidamente com o tamanho da sequência, tornando a exploração eficiente das conformações possíveis um problema computacional complexo.

Neste trabalho, foram investigados diferentes Algoritmos Evolutivos para o problema de predição de estruturas de proteínas no modelo HP-3D. Inicialmente, foram avaliadas três abordagens do AE monoobjetivo, utilizando diferentes funções de aptidão baseadas em contatos hidrofóbicos, energia simplificada e métricas estruturais. Posteriormente, foram exploradas duas versões multiobjetivo por meio do Algoritmo Evolutivo Multiobjetivo com Muitas Tabelas (AEMMT), que permite considerar simultaneamente diferentes critérios estruturais durante o processo evolutivo.

Os experimentos realizados demonstraram que os algoritmos monoobjetivo são capazes de encontrar soluções de boa qualidade, porém com uma taxa de acerto bem menor que as outras abordagens. Os AEs baseados em múltiplos critérios mostraram-se capazes de manter maior variabilidade no processo evolutivo, favorecendo a exploração do espaço de soluções.

Entre as abordagens avaliadas, o AEMMT com cinco tabelas apresentou o melhor desempenho geral, sendo capaz de atingir os valores mínimos de energia conhecidos para todas as instâncias analisadas. Embora a versão com seis tabelas também tenha demonstrado comportamento evolutivo interessante, com maior variabilidade ao longo das gerações, os resultados indicaram que, para as instâncias consideradas neste trabalho, a configuração com cinco tabelas mostrou-se suficiente para alcançar soluções de alta qualidade.

Entretanto, a maior variabilidade observada na versão com seis tabelas sugere que essa abordagem pode se tornar particularmente interessante para instâncias ainda maiores, nas quais a exploração mais ampla do espaço de busca pode

desempenhar um papel mais relevante na descoberta de soluções de alta qualidade.

Diversos trabalhos têm investigado o problema de predição de estruturas de proteínas no modelo HP-3D utilizando diferentes estratégias, incluindo algoritmos evolutivos, métodos híbridos e abordagens baseadas em aprendizado profundo. De modo geral, observa-se que essas técnicas são capazes de atingir os valores mínimos de energia conhecidos para diversas instâncias do problema. A comparação realizada neste trabalho demonstrou que o AEMMT apresentou desempenho equivalente aos métodos reportados na literatura, ao alcançar os melhores valores de energia para todas as instâncias analisadas. Além disso, as conformações estruturais obtidas revelaram-se consistentes com aquelas reportadas nos estudos previamente levantados.

Esses resultados sugerem que algoritmos evolutivos multiobjetivo baseados em tabelas constituem uma alternativa promissora para problemas complexos de otimização combinatória, como o PSP no modelo HP-3D. A possibilidade de considerar múltiplos critérios estruturais de forma simultânea permite equilibrar exploração e intensificação da busca, favorecendo a obtenção de soluções competitivas mesmo com estruturas algorítmicas relativamente simples.

Como perspectivas de trabalhos futuros, diferentes estratégias evolutivas podem ser exploradas para ampliar o desempenho do método proposto. Entre elas destacam-se a investigação de operadores de mutação mais sofisticados, a inclusão de novos critérios de aptidão não diretamente relacionados a métricas de distância, bem como a integração com outras estratégias heurísticas ou híbridas. Além disso, a avaliação do algoritmo em conjuntos de instâncias maiores e mais complexos pode contribuir para compreender melhor o impacto da abordagem multiobjetivo baseada em tabelas em cenários de maior escala, investindo no aumento na quantidade de tabelas com a combinação de critérios que o AEMMT possibilita. Por fim, sugere-se a realização de estudos mais aprofundados sobre métricas de avaliação específicas para problemas com muitos objetivos, como o hipervolume, de forma a permitir uma análise mais robusta da qualidade das soluções obtidas.

Essas investigações futuras podem contribuir para ampliar o entendimento sobre o papel de AE multiobjetivo na predição de estruturas de proteínas e para

o desenvolvimento de métodos evolutivos ainda mais eficientes para problemas complexos de bioinformática computacional, além de fomentar a aplicação dos AEMMTs em diversas outras áreas de pesquisa.

Referências

BERGER, B.; LEIGHTON, T. Protein folding in the hydrophobic-hydrophilic (hp) model is np-complete. **Journal of Computational Biology**, v. 5, n. 1, p. 27–40, 1998. Citado na página 36.

BRANDEN, C.; TOOZE, J. **Introduction to Protein Structure**. 2. ed. New York, NY, USA: Garland Publishing, Inc., 1999. 410 p. Citado na página 30.

BRASIL, C. R. S. **Algoritmo evolutivo de muitos objetivos para predição ab initio de estrutura de proteínas**. Tese (Tese (Doutorado em Ciências de Computação e Matemática Computacional)) — Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, SP, 2012. Disponível em: <<https://doi.org/10.11606/T.55.2012.tde-20072012-163056>>. Citado 7 vezes nas páginas 14, 26, 27, 28, 39, 41 e 51.

BRASIL, C. R. S.; DELBEM, A. C. B.; SILVA, F. L. B. d. Multiobjective evolutionary algorithm with many tables for purely *ab initio* protein structure prediction. **Journal of Computational Chemistry**, Wiley Online Library, v. 34, n. 20, p. 1719–1734, 2013. Citado 6 vezes nas páginas 14, 26, 27, 28, 30 e 34.

BUENO, M. L.; OLIVEIRA, G. M. Multicast flow routing: Evaluation of heuristics and multiobjective evolutionary algorithms. In: **IEEE Congress on Evolutionary Computation**. IEEE, 2010. p. 1–8. Disponível em: <<https://doi.org/10.1109/CEC.2010.5585942>>. Citado na página 23.

CAVALCANTI, D. M. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação), **Algoritmos de otimização de inteligência de enxame aplicados ao problema de predição de proteína**. Uberlândia, MG, Brasil: [s.n.], 2018. 43 p. Disponível em: <<https://repositorio.ufu.br/handle/123456789/24120>>. Citado na página 30.

COELLO, C. A. C. Evolutionary multi-objective optimization: some current research trends and topics that remain to be explored. **Frontiers of Computer Science in China**, Springer, v. 3, n. 1, p. 18–30, 2009. Citado 4 vezes nas páginas 13, 18, 24 e 25.

CORMEN, T. H.; LEISERSON, C. E.; RIVEST, R. L.; STEIN, C. **Introduction to Algorithms**. 4. ed. Cambridge, MA: MIT Press, 2022. ISBN 9780262046305. Citado na página 14.

CUI, Y.; CHEN, R. S.; WONG, W. H. Protein folding simulation with genetic algorithm and supersecondary structure constraints. **Proteins: Structure, Function, and Genetics**, v. 31, n. 3, p. 247–257, maio 1998. Citado 2 vezes nas páginas 27 e 34.

CUSTODIO, F. L. Tese (Doutorado em Modelagem Computacional), **Genetic Algorithms for Ab Initio Protein Structure Prediction**. Petrópolis, RJ: [s.n.], 2008. 221 p. Disponível em: <<https://tede.lncc.br/handle/tede/89>>. Citado na página 30.

CUTELLO, V.; NARZISI, G.; NICOSIA, G. A multi-objective evolutionary approach to the protein structure prediction problem. **Journal of The Royal Society Interface**, The Royal Society, v. 3, n. 6, p. 139–151, 2005. Citado 2 vezes nas páginas 14 e 37.

CUTELLO, V.; NICOSIA, G.; PAVONE, M.; TIMMIS, J. An immune algorithm for protein structure prediction on lattice models. **IEEE Transactions on Evolutionary Computation**, v. 11, n. 1, p. 101–117, 2007. Citado na página 38.

DEB, K. **Multi-Objective Optimization Using Evolutionary Algorithms**. [S.l.]: Wiley, 2001. 518 p. ISBN 9780471873396. Citado na página 25.

DEB, K.; JAIN, H. An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part i: Solving problems with box constraints. **IEEE Transactions on Evolutionary Computation**, v. 18, n. 4, p. 577–601, 2014. Citado na página 25.

DELBEM, A. C. B. **Restabelecimento de energia em sistemas de distribuição por algoritmo evolucionário associado a cadeias de grafos**. Tese (Tese (Doutorado em Engenharia Elétrica)) — Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2002. Acesso em: 7 fev. 2026. Disponível em: <<https://doi.org/10.11606/T.18.2002.tde-09102015-160912>>. Citado 3 vezes nas páginas 26, 38 e 51.

DIAS, J. M. Trabalho de Conclusão de Curso (Graduação em Ciência da Computação), **Comparação do algoritmo de colônia de formiga e algoritmo evolutivo aplicado ao problema de predição de estruturas de proteínas**. Uberlândia, MG: [s.n.], 2016. 53 p. Citado 4 vezes nas páginas 22, 23, 30 e 39.

DILL, K. A. Theory for the folding and stability of globular proteins. **Biochemistry**, v. 24, n. 6, p. 1501–1509, 1985. PMID: 3986190. Disponível em: <<https://doi.org/10.1021/bi00327a032>>. Citado na página 35.

DILL, K. A.; OZKAN, S. B.; WEIKL, T. R.; CHODERA, J. D.; VOELZ, V. A. The protein folding problem: when will it be solved? **Current Opinion in Structural Biology**, v. 17, n. 3, p. 342–346, 2007. ISSN 0959-440X. Nucleic acids / Sequences and topology. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0959440X07000772>. Citado na página 30.

EIBEN, A.; SMITH, J. **Introduction to Evolutionary Computing**. [S.l.]: Springer, 2015. (Natural Computing Series). Geburtenis: 2nd edition. ISBN 9783662448731. Citado na página 13.

ESPITIA, G.; PANG, Y. T.; GUMBART, J. C. **Protein Structure Prediction in the 3D HP Model Using Deep Reinforcement Learning**. 2024. Disponível em: <https://arxiv.org/abs/2412.20329>. Citado 8 vezes nas páginas 7, 30, 40, 41, 56, 57, 71 e 72.

GABRIEL, P. H. R.; DELBEM, A. C. B. **Fundamentos de algoritmos evolutivos**. [S.l.]: ICMC-USP, 2008. Citado 5 vezes nas páginas 13, 18, 20, 21 e 51.

GABRIEL, P. H. R.; MELO, V. V. d.; DELBEM, A. C. B. Algoritmos evolutivos e modelo hp para predição de estruturas de proteínas. **Revista de Controle e Automação**, 2012. Citado 9 vezes nas páginas 7, 27, 30, 39, 41, 44, 56, 57 e 71.

GOLDBERG, D. E. **Genetic Algorithms in Search, Optimization, and Machine Learning**. Boston, MA, USA: Addison-Wesley Pub. Co., 1989. 412 p. ISBN 0201157675. Citado 2 vezes nas páginas 21 e 22.

GOULART, F.; CAMPELO, F. Preference-guided evolutionary algorithms for many-objective optimization. **Information Sciences**, v. 329, p. 236–255, 2016. ISSN 0020-0255. Special issue on Discovery Science. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0020025515006696>. Citado na página 25.

GUO, Y.; TAO, F.; WU, Z.; WANG, Y. Hybrid method to solve hp model on 3d lattice and to probe protein stability upon amino acid mutations. **BMC Systems Biology**, v. 11, n. Suppl 4, p. 93, 2017. Citado 7 vezes nas páginas 7, 40, 41, 56, 57, 70 e 71.

ISHIBUCHI, H.; TSUKAMOTO, N.; NOJIMA, Y. Evolutionary many-objective optimization: A short review. In: IEEE. **2008 IEEE congress on evolutionary computation (IEEE world congress on computational intelligence)**. [S.l.], 2008. p. 2419–2426. Citado 2 vezes nas páginas 13 e 51.

KRANE, D. E.; RAYMER, M. L. **Fundamental Concepts of Bioinformatics**. San Francisco, CA: Benjamin Cummings, 2003. 314 p. Citado na página 43.

LAFETÁ, T. F. d. Q. **Algoritmos evolutivos many objectives aplicados ao problema de roteamento multicast com qualidade de serviço**. 146 p. Dissertação (Dissertação (Mestrado em Ciência da Computação)) — Universidade Federal de Uberlândia, Uberlândia, MG, 2016. Disponível em: <<https://doi.org/10.14393/ufu.di.2016.58>>. Citado na página 28.

LAU, K. F.; DILL, K. A. A lattice statistical mechanics model of the conformational and sequence spaces of proteins. **Macromolecules**, ACS Publications, v. 22, n. 10, p. 3986–3997, 1989. Citado 3 vezes nas páginas 14, 34 e 37.

LIMA, T. W. d. **Algoritmos Evolutivos para Predição de Estruturas de Proteínas**. Dissertação (Dissertação (Mestrado em Ciências de Computação e Matemática Computacional)) — Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2006. Acesso em: 7 fev. 2026. Disponível em: <<https://doi.org/10.11606/D.55.2006.tde-30092006-110850>>. Citado na página 30.

LIN, C.-J.; SU, S.-C. Protein 3d hp model folding simulation using a hybrid of genetic algorithm and particle swarm optimization. **International Journal of Fuzzy Systems**, v. 13, p. 140–147, 06 2011. Citado 8 vezes nas páginas 7, 39, 41, 56, 57, 70, 71 e 72.

LIU, P.; IBA, H. **Enhancing Reinforcement learning in 3-Dimensional Hydrophobic-Polar Protein Folding Model with Attention-based layers**. 2025. Disponível em: <<https://arxiv.org/abs/2504.15634>>. Citado na página 40.

MACRI, A.; BRASIL, C. Uma análise de múltiplos critérios do problema de roteamento de veículos usando o aemmt. **Anais do Computer on the Beach**, v. 16, p. 134–139, 05 2025. Citado na página 28.

MANSOUR, N.; KANJ, F.; KHACHFE, H. **Enhanced genetic algorithm for protein structure prediction based on the HP model**. [S.l.]: INTECH Open Access Publisher, 2011. Citado 7 vezes nas páginas 7, 39, 41, 56, 57, 70 e 71.

MARSON, G. A. Trabalho de Conclusão de Curso (Graduação em Ciência da Computação), **Análise comparativa dos algoritmos genéticos many-objective em problemas de otimização discreta**. Uberlândia, MG: [s.n.], 2017. 45 p. Citado na página 24.

NASCIMENTO, P. L. d.; VARGAS, D. E. C.; WANNER, E. F. Performance evaluation of a hybrid nsga-iii for multi and many-objective optimization in real-world problems. In: **Anais do XV Workshop de Sistemas de Informação (WSIS)**. Rio Paranaíba, MG, Brasil: Sociedade Brasileira de Computação, 2024. p. 80–85. Disponível em: <<https://doi.org/10.5753/wsis.2024.33677>>. Citado na página 23.

NELSON, D. L.; COX, M. M. **Lehninger Principles of Biochemistry**. 7. ed. [S.l.]: W. H. Freeman, 2017. Citado 4 vezes nas páginas 14, 31, 35 e 74.

ROCHA, L. F. O.; PINTO, M. E. T.; CALIRI, A. The water factor in the protein-folding problem. **Brazilian Journal of Physics**, Sociedade Brasileira de Física, v. 34, n. 1, p. 90–101, mar. 2004. ISSN 0103-9733. Disponível em: <<https://doi.org/10.1590/S0103-97332004000100013>>. Citado na página 34.

SANTANA, K.; NAHUM, C. **Métodos de modelagem da estrutura tridimensional de proteínas: Um guia teórico e prático**. [S.l.: s.n.], 2022. ISBN 9786525805825. Citado na página 14.

SANTOS, A.; DELBEM, A.; JR, J. L.; BRETAS, N. Node-depth encoding and multiobjective evolutionary algorithm applied to large-scale distribution system reconfiguration. **Power Systems, IEEE Transactions on**, v. 25, p. 1254 – 1265, 09 2010. Citado 2 vezes nas páginas 27 e 51.

SANTOS, J. F. **Algoritmos evolutivos multiobjetivo baseados em tabelas para escalonamento de tarefas em ambientes multiprocessados**. 111 p. Dissertação (Dissertação (Mestrado em Ciência da Computação)) — Universidade Federal de Uberlândia, Uberlândia, MG, 2023. Disponível em: <<http://doi.org/10.14393/ufu.di.2023.163>>. Citado na página 23.

SANTOS, M. R. Trabalho de Conclusão de Curso (Graduação em Sistemas de Informação), **Aplicação do NSGA-III ao problema de escalonamento multiobjetivo de tarefas**. Uberlândia, MG: [s.n.], 2024. 57 p. Citado na página 23.

SCHULZE-KREMER, S. Genetic algorithms for protein tertiary structure prediction. In: SPRINGER. **European Conference on Machine Learning**. [S.l.], 1993. p. 262–279. Citado na página 34.

SHAKHNOVICH, E.; FARZTDINOV, G.; GUTIN, A. M.; KARPLUS, M. Protein folding bottlenecks: A lattice monte carlo simulation. **Phys. Rev. Lett.**, American Physical Society, v. 67, p. 1665–1668, Sep 1991. Disponível em: <<https://link.aps.org/doi/10.1103/PhysRevLett.67.1665>>. Citado na página 34.

SHUKLA, A.; PANDEY, H. M.; MEHROTRA, D. Comparative review of selection techniques in genetic algorithm. In: **2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE)**. [S.l.: s.n.], 2015. p. 515–519. Citado na página 21.

TRAYKOV, M.; YANEV, N.; MAVREVSKI, R.; YURUKOV, B. Algorithm for protein folding problem in 3d lattice hp model. **International Journal of Biology and Biomedicine**, v. 3, p. 16–21, 2018. Citado 6 vezes nas páginas 7, 40, 41, 56, 57 e 71.

UNGER, R.; MOULT, J. Genetic algorithms for protein folding simulations. **Journal of Molecular Biology**, v. 231, n. 1, p. 75–81, may 1993. Citado na página 34.

VELDHUIZEN, D. A. van. Evolutionary computation and convergence to a pareto front. In: **Late breaking papers at the genetic programming 1998 conference**. [s.n.], 1998. p. 221–228. Disponível em: <<https://api.semanticscholar.org/CorpusID:18585705>>. Citado 2 vezes nas páginas 23 e 24.

Wang, Boyuan. Protein folding: Recent advances and analysis in computing. **BIO Web Conf.**, v. 214, p. 01012, 2026. Disponível em: <<https://doi.org/10.1051/bioconf/202621401012>>. Citado na página 41.

ZHANG, Q.; LI, H. Moea/d: A multiobjective evolutionary algorithm based on decomposition. **IEEE Transactions on evolutionary computation**, IEEE, v. 11, n. 6, p. 712–731, 2007. Citado 3 vezes nas páginas 23, 25 e 38.

ZHANG, Y.; WU, L.; WANG, S. Solving two-dimensional hp model by firefly algorithm and simplified energy function. **Mathematical Problems in Engineering**, v. 2013, 01 2013. Citado na página 37.

ZITZLER, E.; LAUMANN, M.; THIELE, L. **SPEA2: Improving the Strength Pareto Evolutionary Algorithm**. [S.l.], 2001. Disponível em: <<https://doi.org/10.3929/ethz-a-004284029>>. Citado na página 25.