

Osmar Aguiar Ribeiro Junior

Otimização Evolutiva Híbrida para Docking Proteína–Peptídeo

São Luís – MA

2026

Osmar Aguiar Ribeiro Junior

Otimização Evolutiva Híbrida para Docking Proteína–Peptídeo

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal do Maranhão, como parte dos requisitos para obtenção do título de Mestre em Ciência da Computação.

Universidade Federal do Maranhão
Programa de Pós-Graduação em Ciência da Computação

Orientador: Prof. Dr. Omar Andrés Carmona Cortés

São Luís – MA
2026

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).
Diretoria Integrada de Bibliotecas/UFMA

Ribeiro Junior, Osmar Aguiar.

Otimização Evolutiva Híbrida para Docking
ProteínaPeptídeo / Osmar Aguiar Ribeiro Junior. - 2026.
66 f.

Orientador(a): Omar Andrés Carmona Cortés.

Dissertação (Mestrado) - Programa de Pós-graduação em
Ciência da Computação/ccet, Universidade Federal do
Maranhão, São Luís, 2026.

1. Otimização Evolutiva Híbrida. 2. Docking
Molecular. 3. Proteínapeptídeo. 4. Kd-tree. 5. Meta-
heurísticas. I. Carmona Cortés, Omar Andrés. II. Título.

OTIMIZAÇÃO EVOLUTIVA HÍBRIDA PARA DOCKING PROTEÍNA-PEPTÍDEO

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal do Maranhão, como parte dos requisitos para obtenção do título de Mestre em Ciência da Computação.

Trabalho aprovado em _____ de _____ de _____, São Luís – MA:

Prof. Dr. Omar Andrés Carmona Cortés
Presidente da banca / Orientador
Universidade Federal do Maranhão
Programa de Pós-Graduação em Ciência da
Computação

Prof. Dr. Bruno Feres de Souza
Examinador Interno
Universidade Federal do Maranhão
Programa de Pós-Graduação em Ciência da
Computação

Prof. Dr. Fabio Manoel França Lobato
Examinador Externo
UFOPA – Universidade Federal do Oeste do
Pará

A Gabriel, a Brenda e a Alcilene Rodrigues, meu porto e meu horizonte.

Este trabalho é dedicado ao amor que me sustenta e me move.

Agradecimentos

Agradeço a Deus por iluminar meu caminho e me dar forças para chegar até aqui. Aos meus pais, aos meus filhos Brenda e Gabriel, e à minha esposa, pela paciência e carinho. Ao meu orientador, Prof. Dr. Omar Andrés Carmona Cortés, pela confiança e apoio. À UFMA e ao PPGCC, pela oportunidade de formação.

“A imaginação é mais importante que o conhecimento.”
(Albert Einstein)

Resumo

As interações proteína-peptídeo desempenham papel relevante em processos biológicos associados ao reconhecimento molecular, à sinalização celular e ao desenvolvimento de estratégias terapêuticas. A predição computacional desses complexos, entretanto, permanece desafiadora, especialmente em cenários de *docking* cego, nos quais o sítio de ligação não é previamente conhecido e o espaço conformacional a ser explorado é amplo. O problema central investigado neste trabalho consiste em reduzir o custo computacional da busca conformacional em *docking* cego proteína-peptídeo sem comprometer a qualidade energética das poses finais.

Nesse contexto, esta dissertação apresenta o desenvolvimento, a implementação e a validação de uma abordagem híbrida de otimização evolutiva denominada ABC-GA-VGOS, aplicada ao *docking* molecular proteína-peptídeo. A proposta combina mecanismos de exploração populacional, operadores genéticos, refinamento local adaptativo e uma função geométrica mono-objetivo baseada em consultas KD-tree, utilizada como estratégia de avaliação computacionalmente eficiente durante o processo de busca. Ao final da otimização, a melhor pose obtida em cada execução é submetida a uma etapa de *rescoring* energético inspirado no AutoDock Vina, permitindo separar a triagem geométrica rápida da avaliação energética final.

Os experimentos foram realizados com dez complexos proteína-peptídeo obtidos do *Protein Data Bank* (PDB), considerando energia de ligação, tempo de execução e análise estatística comparativa entre os algoritmos avaliados. Os resultados indicaram que o método híbrido obteve as menores energias médias em cinco sistemas avaliados e apresentou menor tempo médio de execução em todos os casos, mantendo comportamento competitivo em diferentes complexos. Assim, a principal contribuição deste trabalho está na integração de estratégias evolutivas complementares com avaliação geométrica eficiente, oferecendo uma alternativa para reduzir o custo computacional do *docking* proteína-peptídeo sem depender exclusivamente de funções energéticas mais custosas ao longo de toda a busca.

Palavras-chave: Otimização evolutiva híbrida. *Docking* molecular. Proteína-peptídeo. KD-tree. Meta-heurísticas.

Abstract

Protein–peptide interactions play an important role in biological processes associated with molecular recognition, cellular signaling, and the development of therapeutic strategies. However, the computational prediction of these complexes remains challenging, especially in blind docking scenarios, in which the binding site is not previously known and the conformational space to be explored is broad. The central problem investigated in this work is to reduce the computational cost of the conformational search in blind protein–peptide docking without compromising the energetic quality of the final poses.

In this context, this dissertation presents the development, implementation, and validation of a hybrid evolutionary optimization approach named ABC–GA–VGOS, applied to protein–peptide molecular docking. The proposed approach combines population-based exploration mechanisms, genetic operators, adaptive local refinement, and a single-objective geometric function based on KD-tree queries, used as a computationally efficient evaluation strategy during the search process. At the end of the optimization, the best pose obtained in each execution is submitted to an energy rescoring step inspired by AutoDock Vina, allowing the separation between fast geometric screening and final energetic evaluation.

The experiments were performed with ten protein–peptide complexes obtained from the Protein Data Bank (PDB), considering binding energy, execution time, and comparative statistical analysis among the evaluated algorithms. The results indicated that the hybrid method obtained the lowest mean energies in five evaluated systems and achieved the shortest mean execution time in all cases, maintaining competitive behavior across different complexes. Thus, the main contribution of this work lies in the integration of complementary evolutionary strategies with efficient geometric evaluation, offering an alternative to reduce the computational cost of protein–peptide docking without relying exclusively on more expensive energy functions throughout the entire search.

Keywords: Hybrid evolutionary optimization. Molecular docking. Protein–peptide. KD-tree. Metaheuristics.

Lista de figuras

1.	Fluxo operacional geral dos experimentos de <i>docking</i> proteína-peptídeo realizados neste trabalho.	28
2.	Representações estruturais de parte dos complexos proteína-peptídeo utilizados nos experimentos.	31
3.	Representações estruturais dos demais complexos proteína-peptídeo utilizados nos experimentos.	32
4.	Médias e intervalos de confiança de 95% para a energia de ligação no complexo 1A1N.	36
5.	Distribuição das energias obtidas por algoritmo para o complexo 1A1N (ANOVA e Friedman).	37
6.	Tempo médio de execução com intervalos de confiança de 95% para o complexo 1A1N.	38
7.	Médias e intervalos de confiança (IC95) da energia de ligação para o complexo 1BE9.	39
8.	Distribuição das energias obtidas por algoritmo para o complexo 1BE9 (ANOVA e Friedman).	40
9.	Tempo médio de execução e intervalos de confiança (IC95) para o complexo 1BE9.	41
10.	Médias e intervalos de confiança (IC95) da energia de ligação para o complexo 1JSU.	42
11.	Distribuição das energias obtidas por algoritmo para o complexo 1JSU (ANOVA e Friedman).	43
12.	Tempo médio de execução e intervalos de confiança (IC95) para o complexo 1JSU.	44
13.	Médias e intervalos de confiança (IC95) da energia de ligação para o complexo 1BXL.	45
14.	Distribuição das energias obtidas por algoritmo para o complexo 1BXL (ANOVA e Friedman).	46
15.	Tempo médio de execução e intervalos de confiança (IC95) para o complexo 1BXL.	47
16.	Médias e intervalos de confiança (IC95) da energia de ligação para o complexo 1BXO.	48
17.	Distribuição das energias obtidas por algoritmo para o complexo 1BXO (ANOVA e Friedman).	49

18.	Tempo médio de execução e intervalos de confiança (IC95) para o complexo 1BXO.	50
19.	Médias e intervalos de confiança (IC95) da energia de ligação para o complexo 1YCR.	51
20.	Distribuição das energias obtidas por algoritmo para o complexo 1YCR (ANOVA e Friedman).	52
21.	Tempo médio de execução e intervalos de confiança (IC95) para o complexo 1YCR.	53
22.	Médias e intervalos de confiança (IC95) da energia de ligação para o complexo 2IGF.	54
23.	Distribuição das energias obtidas por algoritmo para o complexo 2IGF (ANOVA e Friedman).	54
24.	Tempo médio de execução e intervalos de confiança (IC95) para o complexo 2IGF.	55
25.	Médias e intervalos de confiança (IC95) da energia de ligação para o complexo 2BBA.	56
26.	Distribuição das energias obtidas por algoritmo para o complexo 2BBA (ANOVA e Friedman).	56
27.	Tempo médio de execução e intervalos de confiança (IC95) para o complexo 2BBA.	57
28.	Médias e intervalos de confiança (IC95) da energia de ligação para o complexo 2AN6.	58
29.	Distribuição das energias obtidas por algoritmo para o complexo 2AN6 (ANOVA e Friedman).	58
30.	Tempo médio de execução e intervalos de confiança (IC95) para o complexo 2AN6.	59
31.	Médias e intervalos de confiança (IC95) da energia de ligação para o complexo 4QVE.	60
32.	Distribuição das energias obtidas por algoritmo para o complexo 4QVE (ANOVA e Friedman).	60
33.	Tempo médio de execução e intervalos de confiança (IC95) para o complexo 4QVE.	61

Lista de tabelas

1.	Conjunto de complexos proteína-peptídeo utilizados nos experimentos.	30
2.	Resultados por proteína: energia média e tempo médio de execução. . .	34
3.	Resultados médios obtidos para o complexo 1A1N (dez execuções independentes).	35
4.	Resultados médios obtidos para o complexo 1BE9 (dez execuções independentes).	38
5.	Resultados médios obtidos para o complexo 1JSU (dez execuções independentes).	41
6.	Resultados médios obtidos para o complexo 1BXL (dez execuções independentes).	44
7.	Resultados médios obtidos para o complexo 1BXO (dez execuções independentes).	47
8.	Resultados médios obtidos para o complexo 1YCR (dez execuções independentes).	51
9.	Resultados médios obtidos para o complexo 2IGF.	53
10.	Resultados médios obtidos para o complexo 2BBA.	56
11.	Resultados médios obtidos para o complexo 2AN6.	58
12.	Resultados médios obtidos para o complexo 4QVE.	59

Lista de abreviaturas e siglas

ABC	<i>Artificial Bee Colony</i>
ANOVA	<i>Analysis of Variance</i> (análise de variância)
CNN	<i>Convolutional Neural Network</i>
DL	<i>Deep Learning</i>
GA	<i>Genetic Algorithm</i>
IC95	Intervalo de confiança de 95%
KD-tree	<i>k-dimensional tree</i> (árvore k-dimensional)
PDB	<i>Protein Data Bank</i>
PDBQT	Formato derivado do PDB com cargas parciais e tipos atômicos
RL	<i>Reinforcement Learning</i>
RMSD	<i>Root Mean Square Deviation</i>
VGOS	<i>Variable Genetic Operator Search</i>

Sumário

	Lista de figuras	8
	Lista de tabelas	10
1	INTRODUÇÃO	14
1.1.	Contextualização e justificativa	15
1.2.	Trabalhos correlatos e posicionamento da proposta	15
1.3.	Hipótese	16
1.4.	Objetivos	17
1.5.	Estrutura da dissertação	17
2	FUNDAMENTAÇÃO TEÓRICA	18
2.1.	Docking molecular e desafios computacionais	18
2.2.	Metaheurísticas evolutivas e inteligência de enxame no docking	18
2.3.	Representação de poses e transformações rígidas	19
2.4.	Consultas espaciais eficientes com KD-tree	20
2.5.	Funções de avaliação geométrica e rescoring energético	21
3	METODOLOGIA	23
3.1.	Visão geral	23
3.2.	Formulação computacional do problema	23
3.3.	Avaliação das poses e validação energética	24
3.4.	Estratégias de otimização avaliadas	25
3.5.	Fluxo operacional dos experimentos	26
3.6.	Protocolo experimental	29
3.7.	Métricas e análise estatística	29
3.8.	Conjunto experimental de proteínas e peptídeos	29
3.9.	Infraestrutura computacional	32
4	RESULTADOS E DISCUSSÃO	33
4.1.	Introdução aos resultados	33
4.2.	Resultados individuais por complexo	34
5	CONCLUSÃO	62
5.1.	Limitações do estudo	63
5.2.	Trabalhos futuros	63

REFERÊNCIAS	65
--------------------	-----------

1 Introdução

As interações proteína-peptídeo desempenham papel relevante em processos biológicos associados ao reconhecimento molecular, à sinalização celular, à resposta imune e ao desenvolvimento de estratégias terapêuticas baseadas em peptídeos (MASOUDI-SOBHANZADEH et al., 2021; LING et al., 2026). A predição estrutural desses complexos, entretanto, permanece uma tarefa desafiadora, pois a determinação experimental por técnicas como cristalografia de raios X, ressonância magnética nuclear e microscopia crioeletrônica pode ser custosa, demorada e limitada a determinados sistemas, especialmente quando envolve complexos transitórios ou flexíveis (LING et al., 2026).

No contexto computacional, o *docking* proteína-peptídeo busca estimar a conformação e o modo de ligação de um peptídeo em relação a uma proteína receptora. Esse problema é mais complexo do que muitos cenários tradicionais de *docking* proteína-ligante, pois os peptídeos possuem maior flexibilidade estrutural, número elevado de graus de liberdade e capacidade de assumir diferentes conformações durante o processo de ligação (MASOUDI-SOBHANZADEH et al., 2021; LING et al., 2026). Em abordagens de *docking* cego, a dificuldade é ampliada, uma vez que o sítio de ligação não é previamente conhecido e a busca precisa explorar amplas regiões da superfície proteica.

Ao longo dos últimos anos, diferentes estratégias foram propostas para lidar com essa complexidade, incluindo métodos baseados em busca global, refinamento local, funções de pontuação físico-químicas, abordagens meta-heurísticas e modelos baseados em aprendizado profundo. Estudos baseados em meta-heurísticas destacam a importância da combinação entre algoritmos de busca eficientes e funções de avaliação adequadas para identificar poses próximas ao estado nativo (MASOUDI-SOBHANZADEH et al., 2021). Por outro lado, abordagens recentes com aprendizado profundo ampliaram a capacidade de predição estrutural, mas ainda podem depender da disponibilidade de dados, da qualidade das informações de entrada e de etapas posteriores de refinamento energético (FILGUEIRAS; VARELA; SANTOS, 2023; LING et al., 2026).

Assim, mesmo diante dos avanços recentes, o *docking* proteína-peptídeo ainda exige estratégias capazes de equilibrar exploração conformacional, qualidade das poses geradas e custo computacional. Nesse cenário, abordagens que combinem busca evolutiva, refinamento local e avaliação geométrica eficiente permanecem relevantes, especialmente quando se busca reduzir o custo da otimização sem depender exclusivamente de funções energéticas mais custosas durante todo o processo de busca.

1.1. Contextualização e justificativa

Nos últimos anos, métodos baseados em aprendizado profundo vêm apresentando avanços relevantes em diferentes tarefas associadas ao *docking*, incluindo predição estrutural, ranqueamento e construção de funções de pontuação. Shirali *et al.* compararam funções de pontuação clássicas e modelos baseados em aprendizado profundo em diferentes *datasets*, evidenciando diferenças importantes de desempenho e custo computacional entre as abordagens (SHIRALI *et al.*, 2025). Em paralelo, o desenvolvimento de *benchmarks* mais amplos tem contribuído para uma avaliação mais robusta de métodos recentes. Morehead *et al.* propuseram o *PoseBench*, um ambiente padronizado para comparação de métodos contemporâneos, incluindo soluções baseadas em aprendizado profundo (MOREHEAD *et al.*, 2026). Ferramentas como o *Gnina* também passaram a incorporar pontuação por redes neurais convolucionais em fluxos tradicionais de *docking*, ampliando o repertório metodológico disponível (MCNUTT *et al.*, 2025).

Apesar desses avanços, a aplicação prática de métodos baseados exclusivamente em aprendizado profundo ainda enfrenta obstáculos importantes. Entre eles, destacam-se a dependência de grandes volumes de dados rotulados, o custo computacional de treinamento e inferência, a dificuldade de integração direta em fluxos completos de busca evolutiva e, em alguns casos, a menor interpretabilidade do processo de decisão. Em consequência, permanece relevante investigar abordagens híbridas capazes de preservar eficiência computacional, estabilidade e interpretabilidade, especialmente em cenários nos quais milhares de poses precisam ser avaliadas sucessivamente.

Essa necessidade constitui a principal justificativa deste trabalho. Em problemas de *docking* proteína-peptídeo, o custo de avaliação pode se tornar um gargalo tão importante quanto a própria capacidade exploratória do algoritmo. Assim, torna-se pertinente investigar estratégias em que a busca seja guiada por uma função geométrica de baixo custo, ao mesmo tempo em que mecanismos evolutivos e de refinamento local sejam combinados para ampliar a robustez e a capacidade de convergência. Sob essa perspectiva, o estudo propõe uma alternativa metodológica que procura equilibrar qualidade das soluções e custo computacional, sem depender exclusivamente de modelos de alto custo.

1.2. Trabalhos correlatos e posicionamento da proposta

Entre os métodos metaheurísticos aplicados ao *docking*, algoritmos evolutivos permanecem particularmente relevantes por sua capacidade de explorar superfícies multimodais sem depender de gradientes explícitos. Algoritmos genéticos, por exemplo, realizam evolução populacional por seleção, recombinação e mutação, favorecendo busca global em espaços conformacionais complexos (HOLLAND, 1992). O *Variable Genetic Operator Search* (VGOS), por sua vez, introduz adaptação dinâmica de operadores genéticos, ajustando

probabilidades de recombinação e mutação conforme o desempenho observado ao longo da evolução (MESMOUDI; TAVARES; TALBI, 2010). Tavares *et al.* mostraram que a incorporação de refinamento local em métodos evolutivos pode melhorar significativamente a qualidade final das soluções, sobretudo quando a intensificação é aplicada de forma seletiva (TAVARES; MESMOUDI; TALBI, 2009). Em linha semelhante, Leonhart e Dorn propuseram um algoritmo memético baseado em BRKGA com refinamento local e *simulated annealing*, obtendo desempenho competitivo em *benchmarks* de *docking* (LEONHART; DORN, 2019).

Além das abordagens genéticas, estratégias inspiradas em inteligência de enxame também têm sido utilizadas no problema. O algoritmo *Artificial Bee Colony* (ABC), proposto por Karaboga e Basturk, simula o comportamento coletivo de abelhas em busca de alimento, distribuindo a população entre fases de exploração, seleção probabilística e substituição de soluções (KARABOGA; BASTURK, 2008). Essa estrutura favorece manutenção de diversidade populacional e adaptação dinâmica em superfícies energéticas irregulares. Mais recentemente, Zhou *et al.* relataram resultados promissores com um programa de *docking* baseado em múltiplos enxames competitivos e refinamento local (ZHOU *et al.*, 2023).

Ao lado dessas abordagens baseadas em metaheurísticas, também vêm crescendo soluções apoiadas em representações estruturais mais sofisticadas. Lin *et al.* desenvolveram o *DeepRLI*, um modelo do tipo *graph-transformer* voltado a múltiplas tarefas associadas ao *docking* (LIN *et al.*, 2025). Sob *et al.* investigaram o ajuste de modelos generativos em espaço latente por *reinforcement learning*, indicando aumento nas taxas de sucesso em tarefas de *docking* (SOB *et al.*, 2024). Esses estudos evidenciam a vitalidade da área e a diversidade de estratégias possíveis.

Entretanto, observa-se que muitas abordagens recentes priorizam o aumento de poder preditivo, enquanto outras concentram-se em mecanismos de busca mais eficientes. Este trabalho se posiciona justamente na interseção entre esses dois objetivos: em vez de substituir completamente a etapa de busca por modelos complexos, propõe-se combinar uma avaliação geométrica eficiente com operadores evolutivos híbridos e refinamento local seletivo. A proposta parte da premissa de que, em problemas proteína-peptídeo, uma função de avaliação leve pode viabilizar exploração intensiva do espaço conformacional, desde que acompanhada de mecanismos suficientemente robustos de intensificação e diversidade.

1.3. Hipótese

A hipótese central desta pesquisa é que a combinação entre uma função geométrica de avaliação eficiente e uma estratégia híbrida de otimização evolutiva permite obter soluções competitivas em qualidade estrutural e energética, com menor custo computacional do que estratégias isoladas aplicadas ao *docking* molecular proteína-peptídeo.

1.4. Objetivos

O objetivo geral deste trabalho é desenvolver e avaliar uma abordagem híbrida evolutiva aplicada ao *docking* molecular proteína-peptídeo, utilizando uma avaliação geométrica eficiente como mecanismo de suporte à busca conformacional.

Como objetivos específicos, busca-se:

- implementar uma função de avaliação geométrica baseada em consultas espaciais eficientes;
- integrar operadores evolutivos complementares em uma estratégia híbrida de otimização;
- comparar experimentalmente diferentes algoritmos aplicados ao problema de *docking*;
- analisar estatisticamente os resultados obtidos em múltiplos complexos moleculares.

1.5. Estrutura da dissertação

Esta dissertação está organizada em cinco capítulos. Neste Capítulo 1, apresentam-se a contextualização do problema, a justificativa do estudo, os trabalhos correlatos mais diretamente associados à proposta, a hipótese e os objetivos da pesquisa. O Capítulo 2 reúne a fundamentação teórica necessária à compreensão do trabalho, abordando princípios de *docking* molecular, metaheurísticas evolutivas, representação geométrica de poses, transformações rígidas, estruturas espaciais eficientes e funções de avaliação. O Capítulo 3 descreve a metodologia adotada, incluindo a formulação computacional do problema, a função objetivo, os algoritmos avaliados, o protocolo experimental e as métricas de análise. O Capítulo 4 apresenta os resultados obtidos e sua discussão comparativa. Por fim, o Capítulo 5 reúne as conclusões, as limitações do estudo e as perspectivas de trabalhos futuros.

2 Fundamentação teórica

2.1. Docking molecular e desafios computacionais

O *molecular docking* consiste em estimar a orientação espacial e a conformação de um ligante ao interagir com uma macromolécula receptora, buscando identificar poses energeticamente favoráveis associadas à formação de complexos estáveis (LEACH, 2001). Em termos computacionais, trata-se de um problema que articula duas dimensões centrais: um mecanismo de busca, responsável por gerar e refinar conformações candidatas, e uma função de avaliação, encarregada de distinguir soluções promissoras de soluções pouco plausíveis. O desempenho global de um método de *docking* depende justamente do equilíbrio entre esses dois componentes.

Esse problema torna-se particularmente desafiador em razão da elevada dimensionalidade do espaço de busca, da presença de múltiplos mínimos locais e da flexibilidade conformacional das moléculas envolvidas. No caso de complexos proteína-peptídeo, essa dificuldade tende a ser ainda maior, pois os peptídeos usualmente apresentam maior mobilidade estrutural do que pequenos ligantes, ampliando o número de configurações admissíveis. Em consequência, métodos que realizam busca exaustiva tornam-se rapidamente inviáveis, o que justifica o uso de estratégias heurísticas e metaheurísticas para explorar o espaço conformacional de maneira eficiente.

Ferramentas clássicas como AutoDock4 e AutoDock Vina ilustram bem essa formulação. Esses métodos combinam algoritmos estocásticos de busca com funções empíricas de pontuação compostas por termos físico-químicos associados a interações atrativas e repulsivas (MORRIS et al., 2009; TROTT; OLSON, 2010). Embora tenham alcançado ampla difusão na literatura, tais abordagens podem apresentar custo computacional elevado quando aplicadas a buscas extensivas, especialmente em cenários de *blind docking* ou em sistemas com maior flexibilidade estrutural. Por essa razão, investigações voltadas à redução do custo de avaliação e ao aprimoramento da busca permanecem centrais na área.

2.2. Metaheurísticas evolutivas e inteligência de enxame no docking

Entre os métodos clássicos de otimização aplicados ao *docking*, os algoritmos evolutivos destacam-se por sua capacidade de explorar espaços de busca amplos, multimodais e não lineares, sem depender de informação de gradiente. Algoritmos genéticos realizam evolução populacional por seleção, recombinação e mutação, favorecendo exploração global e manutenção de diversidade (HOLLAND, 1992). No contexto de *docking*, essa característica contribui para que múltiplas regiões promissoras sejam investigadas simultaneamente,

reduzindo o risco de aprisionamento precoce em mínimos locais.

O VGOS estende essa lógica ao introduzir adaptação dinâmica dos operadores genéticos, ajustando probabilidades de recombinação e mutação conforme o ganho de qualidade observado ao longo da evolução (MESMOUDI; TAVARES; TALBI, 2010). Essa adaptação é relevante porque o equilíbrio entre exploração e intensificação não permanece constante durante toda a execução. Em fases iniciais, a diversidade populacional tende a ser mais importante; em fases posteriores, refinamentos mais direcionados podem contribuir para melhorar soluções já promissoras.

Outro elemento recorrente na literatura é a integração entre busca populacional e refinamento local. A aplicação seletiva de métodos locais pode aumentar significativamente a qualidade final das soluções, sobretudo quando o refinamento é concentrado em indivíduos mais competitivos (TAVARES; MESMOUDI; TALBI, 2009). Esses estudos reforçam a ideia de que, em problemas de alta complexidade, a combinação entre exploração global e intensificação local tende a ser mais eficaz do que o uso isolado de apenas um desses mecanismos.

Além dos algoritmos evolutivos tradicionais, estratégias inspiradas em inteligência de enxame também ocupam posição importante na literatura. O ABC simula o comportamento coletivo de abelhas em busca de alimento, distribuindo a busca entre etapas de exploração, seleção probabilística e reposição de soluções (KARABOGA; BASTURK, 2008). No *docking*, essa estrutura é particularmente útil para manter diversidade populacional sem abandonar regiões promissoras já identificadas.

Abordagens baseadas em múltiplos enxames competitivos e refinamento local também têm apresentado resultados promissores em cenários com ligantes flexíveis (ZHOU et al., 2023). Em conjunto, esses trabalhos evidenciam que a combinação de mecanismos evolutivos, estratégias inspiradas em enxames e refinamento local constitui uma linha consistente para o tratamento de problemas de *docking* molecular.

Dessa forma, a escolha desses métodos como base comparativa e como componentes da abordagem proposta decorre de sua capacidade de combinar exploração global, manutenção de diversidade populacional e refinamento local. Essas características são particularmente relevantes em espaços conformacionais amplos e multimodais, como os encontrados no *docking* proteína-peptídeo.

2.3. Representação de poses e transformações rígidas

Em problemas de *docking* molecular, uma pose pode ser representada por parâmetros contínuos associados à posição e à orientação do ligante em relação ao receptor. Uma formulação amplamente utilizada considera três variáveis de translação e três variáveis de rotação rígida, totalizando seis graus de liberdade para a representação global da pose (LEACH, 2001; MORRIS et al., 2009; TROTT; OLSON, 2010). Essa modelagem permite

explorar diferentes posicionamentos espaciais do ligante preservando sua estrutura interna durante a etapa principal de busca.

Neste trabalho, cada solução candidata é descrita por um vetor de seis componentes, correspondente às coordenadas translacionais e aos ângulos de rotação do ligante. As variáveis translacionais controlam o deslocamento do peptídeo no espaço tridimensional, enquanto as variáveis rotacionais controlam sua orientação em torno dos eixos cartesianos. Essa representação é compatível com o objetivo de avaliar diferentes modos globais de ligação em um cenário de *blind docking*, no qual o sítio de interação não é previamente conhecido.

A adoção de transformações rígidas reduz a complexidade computacional quando comparada a modelos que incorporam flexibilidade interna completa em todos os átomos ou ligações do ligante. Em vez de alterar continuamente a geometria interna da molécula, a busca concentra-se na determinação de um posicionamento global plausível em relação ao receptor. Essa simplificação não elimina toda a complexidade do *docking* proteína-peptídeo, mas torna a etapa inicial de otimização mais viável do ponto de vista computacional.

As transformações rígidas consistem em translações e rotações aplicadas ao conjunto de coordenadas tridimensionais do ligante, preservando as distâncias internas entre seus átomos. Na prática, o ligante pode ser inicialmente centralizado em torno de seu centro geométrico e, em seguida, submetido a deslocamentos espaciais e rotações em torno dos eixos cartesianos. Esse procedimento permite avaliar sistematicamente diferentes orientações e posicionamentos sem deformar a estrutura interna da molécula.

Assim, a representação em seis graus de liberdade oferece uma formulação computacional adequada para a etapa de busca global adotada neste trabalho. Ela permite explorar amplas regiões da superfície receptora com custo controlado, ao mesmo tempo em que fornece uma base simples e consistente para aplicação dos algoritmos de otimização avaliados.

2.4. Consultas espaciais eficientes com KD-tree

Uma das operações mais frequentes em *docking* molecular é o cálculo de distâncias entre átomos do receptor e do ligante. Em abordagens nas quais milhares de poses precisam ser avaliadas ao longo de sucessivas iterações, o custo acumulado dessas consultas pode se tornar um dos principais gargalos do método. Por isso, estruturas de dados espaciais eficientes exercem papel importante na viabilização computacional de funções de avaliação rápidas.

Entre essas estruturas, a KD-tree destaca-se por organizar coordenadas tridimensionais em partições hierárquicas, permitindo acelerar consultas de vizinhança e busca de vizinhos mais próximos (BENTLEY, 1975). Em vez de recalculer exaustivamente todas as distâncias entre pares de átomos a cada nova pose, torna-se possível recuperar com menor custo informações espaciais relevantes para a avaliação geométrica. Essa propriedade é

particularmente útil em métodos populacionais, nos quais a função objetivo é chamada repetidas vezes ao longo de várias gerações.

No contexto deste trabalho, o uso de KD-tree é teoricamente relevante porque fornece suporte à construção de funções geométricas de baixo custo. Ao permitir consultas rápidas de proximidade, a estrutura viabiliza critérios de avaliação baseados em colisão, contato e exposição espacial, sem exigir a todo instante uma modelagem energética mais completa. Assim, a KD-tree não é apenas um detalhe de implementação, mas um componente conceitual importante para sustentar a eficiência da estratégia de busca.

2.5. Funções de avaliação geométrica e rescoring energético

Funções de avaliação geométrica utilizam relações espaciais entre receptor e ligante como critério de julgamento rápido durante o processo de busca conformacional. Diferentemente de funções energéticas mais completas, essas formulações priorizam baixo custo computacional, sendo especialmente adequadas para etapas em que um grande número de poses precisa ser analisado sucessivamente. Em termos gerais, a ideia é transformar restrições e indícios geométricos de plausibilidade em um valor escalar capaz de orientar a otimização.

Entre os critérios frequentemente empregados nesse tipo de função, destacam-se a penalização de colisões atômicas, a recompensa por contatos geométricos favoráveis e a penalização de poses excessivamente afastadas da superfície receptora. A penalização de colisões busca descartar soluções com interpenetração estérica evidente; a recompensa por contatos procura favorecer aproximações plausíveis entre ligante e receptor; e a penalização por exposição evita que poses artificialmente “livres de colisão” sejam indevidamente favorecidas por estarem distantes da região de interação. Em conjunto, esses critérios permitem construir uma função objetiva leve, informativa e adequada ao guiamento de algoritmos populacionais.

Entretanto, funções geométricas não capturam integralmente interações físico-químicas relevantes para a estabilidade do complexo molecular. Por essa razão, muitas abordagens complementam a busca com uma etapa posterior de *rescoring*, na qual um conjunto reduzido de soluções finais é reavaliado por funções energéticas mais expressivas (TROTT; OLSON, 2010). Em modelos inspirados no AutoDock Vina, por exemplo, a energia costuma incorporar termos gaussianos, repulsão estérica, hidrofobicidade, ligações de hidrogênio e contribuições eletrostáticas. Esse arranjo em duas etapas – triagem geométrica rápida seguida de validação energética mais informativa – permite conciliar eficiência computacional com maior poder discriminativo na avaliação final das poses.

Nos últimos anos, técnicas de aprendizado profundo também passaram a influenciar fortemente o desenvolvimento de funções de pontuação e estratégias de avaliação. Shirali *et al.* relataram ganhos preditivos relevantes em funções baseadas em redes neurais profundas

(SHIRALI et al., 2025), enquanto o *PoseBench* consolidou um ambiente de comparação mais sistemático para métodos modernos (MOREHEAD et al., 2026). Ferramentas como o *Gnina* exemplificam essa convergência ao incorporar redes convolucionais em *pipelines* tradicionais de *docking* (MCNUTT et al., 2025). Ainda assim, permanece atual o interesse por funções geométricas interpretáveis e de baixo custo, especialmente quando o objetivo é sustentar busca intensiva em grande número de avaliações. Nesse cenário, a combinação entre triagem geométrica eficiente e validação energética posterior constitui uma base conceitual consistente para a proposta metodológica desenvolvida nesta dissertação.

3 Metodologia

3.1. Visão geral

Este capítulo descreve os procedimentos metodológicos adotados para implementação, execução e avaliação das estratégias de busca consideradas no problema de *docking* proteína-peptídeo em modo *blind docking*. O estudo compara três métodos: VGOS, ABC–Nelder–Mead e o algoritmo híbrido proposto ABC–GA–VGOS. Em todas as abordagens, a busca conformacional é realizada em um espaço contínuo de seis dimensões, enquanto a avaliação das poses ocorre em duas etapas complementares: uma etapa de triagem geométrica rápida, utilizada como *fitness* durante a otimização, e uma etapa final de *rescoring* energético inspirada no AutoDock Vina, empregada para validação e comparação dos resultados.

A adoção dessa organização metodológica decorre da necessidade de conciliar eficiência computacional e capacidade de exploração do espaço conformacional. Durante a busca, milhares de poses precisam ser avaliadas sucessivamente, o que torna pouco viável o uso contínuo de funções energéticas mais custosas. Por essa razão, a otimização é guiada por uma função geométrica leve, enquanto a comparação final entre os métodos é realizada com base no valor energético obtido por *rescoring* da melhor pose encontrada em cada execução.

3.2. Formulação computacional do problema

Cada solução candidata representa uma pose tridimensional do ligante em relação ao receptor. Neste trabalho, a pose é modelada por seis variáveis contínuas, correspondentes a três parâmetros translacionais e três parâmetros rotacionais, definidos por

$$(x, y, z, r_x, r_y, r_z).$$

As variáveis x , y e z controlam o deslocamento espacial do ligante, enquanto r_x , r_y e r_z descrevem as rotações em torno dos eixos cartesianos. As rotações são restritas ao intervalo $[-180^\circ, 180^\circ]$.

A transformação da pose é aplicada de forma rígida ao ligante. Inicialmente, o peptídeo é deslocado para o novo centro espacial indicado pelos parâmetros translacionais. Em seguida, são aplicadas rotações em torno do centro geométrico atualizado. Essa estratégia preserva as distâncias internas do ligante e concentra o esforço computacional na busca por orientações e posicionamentos globais mais favoráveis.

O problema foi tratado em modo *blind docking*. Assim, o domínio de busca não foi restringido a um sítio de ligação previamente conhecido, mas definido a partir do volume do receptor expandido por uma margem fixa, permitindo que toda a superfície proteica pudesse ser explorada durante a otimização.

3.3. Avaliação das poses e validação energética

Durante a etapa de busca, a qualidade de cada pose é estimada por uma função objetivo geométrica de baixo custo computacional, denotada por E_{geom} . Essa função foi construída para refletir três critérios espaciais principais: evitar interpenetrações atômicas, favorecer contatos plausíveis entre ligante e receptor e penalizar poses excessivamente afastadas da superfície proteica.

Para viabilizar a avaliação repetida de milhares de poses, os átomos do receptor são indexados em uma estrutura KD-tree. A cada avaliação, os átomos transformados do ligante consultam essa estrutura para obtenção das distâncias mínimas até os átomos do receptor. Com isso, reduz-se substancialmente o custo associado ao cálculo direto de todas as distâncias par a par.

A energia geométrica é expressa por

$$E_{\text{geom}} = W_{\text{col}} P_{\text{col}} + W_{\text{exp}} P_{\text{exp}} - W_{\text{cont}} B_{\text{cont}}, \quad (3.1)$$

em que P_{col} representa a penalização por colisão, P_{exp} representa a penalização por exposição excessiva e B_{cont} corresponde ao bônus por contatos favoráveis. Os pesos W_{col} , W_{exp} e W_{cont} foram mantidos fixos ao longo dos experimentos, de modo a preservar comparabilidade direta entre os algoritmos.

A penalização por colisão é tratada em dois regimes. Quando a menor distância ligante-receptor fica abaixo de um limiar severo, aplica-se *early-exit*, descartando a pose com penalização elevada. Quando a distância permanece em uma faixa intermediária, utiliza-se uma penalização quadrática, capaz de diferenciar colisões moderadas de geometrias mais plausíveis. O bônus de contato contabiliza átomos do ligante cujas distâncias mínimas ao receptor se encontram em uma janela favorável. Já a penalização de exposição é calculada a partir do afastamento do centroide do ligante em relação ao receptor, desestimulando poses excessivamente distantes da região de interação.

Nesta dissertação, E_{geom} é utilizado exclusivamente como *fitness* da busca. Ele não compõe diretamente as tabelas e figuras comparativas finais. Após o término da otimização, apenas a melhor pose final de cada execução é submetida a uma etapa adicional de *rescoring* energético, denotada por E_{vina} . Essa avaliação é inspirada no AutoDock Vina e incorpora termos associados a atração gaussiana, repulsão estérica, interações hidrofóbicas, ligações de hidrogênio e componente eletrostática.

A adoção desse procedimento em duas etapas permite combinar eficiência e poder discriminativo. Durante a busca, utiliza-se uma função leve para viabilizar exploração intensiva do espaço conformacional; ao final, emprega-se uma função mais informativa apenas sobre o subconjunto mínimo de soluções de interesse. Os valores de E_{vina} , expressos em kcal/mol, foram usados como métrica energética principal nas tabelas, gráficos e testes estatísticos.

3.4. Estratégias de otimização avaliadas

As três estratégias comparadas neste trabalho compartilham a mesma representação da pose, o mesmo domínio de busca e a mesma função geométrica de avaliação. A diferença entre elas está no modo como exploram o espaço conformacional, atualizam a população e aplicam mecanismos de refinamento local. Em todos os casos, a busca ocorre sobre uma população de soluções em espaço contínuo de seis dimensões, mas cada algoritmo organiza de forma distinta o equilíbrio entre exploração global, intensificação local e diversidade populacional.

O VGOS foi utilizado como método de referência baseado em operadores genéticos adaptativos. Sua principal característica é ajustar dinamicamente as probabilidades de aplicação de cruzamento e mutação ao longo das gerações, reforçando os operadores que, empiricamente, geram maior ganho de qualidade nas soluções. No contexto deste trabalho, o VGOS atua sobre a população de poses candidatas por meio de seleção, recombinação e mutação, preservando as melhores soluções por elitismo. Além disso, incorpora refinamento local periódico por Solis–Wets, aplicado sobre um subconjunto superior da população, com o objetivo de realizar ajustes finos em regiões já identificadas como promissoras. Assim, o VGOS representa uma referência importante por combinar adaptação de operadores e busca local em um mesmo fluxo evolutivo.

O segundo método de referência combina o modelo ABC com refinamento determinístico de Nelder–Mead. Nesse arranjo, cada fonte de alimento corresponde a uma pose candidata do ligante. Na fase *employed*, cada solução ativa gera uma nova vizinhança por perturbação local, tentando melhorar sua qualidade geométrica. Na fase *onlooker*, as soluções mais promissoras recebem maior probabilidade de serem revisitadas, intensificando a busca em regiões de melhor *fitness*. Já a fase de reposição de soluções pouco promissoras mantém diversidade populacional e reduz aprisionamento em mínimos locais. Sobre as melhores poses obtidas, aplica-se refinamento local por Nelder–Mead, que atua de forma determinística sobre o simplex local para ajustar a solução em torno de mínimos mais profundos.

O método híbrido proposto, denominado ABC–GA–VGOS, integra em um único fluxo cooperativo mecanismos inspirados em *Artificial Bee Colony*, operadores genéticos e refinamento local por Solis–Wets. A proposta parte do princípio de que nenhuma dessas

estratégias, de forma isolada, oferece o melhor compromisso entre diversidade, intensificação e custo computacional em todos os cenários de *docking* proteína-peptídeo. Por isso, o algoritmo combina fases populacionais distintas, preservando o caráter exploratório do ABC, a recombinação controlada do GA e a intensificação periódica inspirada em métodos híbridos da literatura.

A população é inicializada uniformemente no espaço 6D de busca, e cada indivíduo é avaliado por E_{geom} . Em cada geração, o algoritmo executa inicialmente uma fase *employed*, na qual cada solução gera uma nova vizinhança por perturbação diferencial. Em seguida, ocorre a fase *onlooker*, em que as soluções são revisitadas segundo probabilidades proporcionais ao respectivo *fitness*, concentrando maior esforço de busca nas regiões mais promissoras do espaço conformacional.

Diferentemente do ABC clássico, a implementação adotada neste trabalho não utiliza *scout bees* convencionais. Em seu lugar, a diversidade populacional é restaurada por meio de um mecanismo de *soft reset*, aplicado a indivíduos que acumulam tentativas sem melhora acima de um limite estabelecido. Essa escolha evita reinicializações excessivamente agressivas e preserva parte da informação já explorada pela população, mantendo equilíbrio entre renovação da busca e aproveitamento do histórico evolutivo.

Para promover intensificação local sem elevar excessivamente o custo computacional, o melhor indivíduo da população recebe, em toda geração, um refinamento curto por Solis–Wets, com parada antecipada por paciência. Além disso, em intervalos regulares, um refinamento mais intenso é aplicado a um conjunto *top-k* de indivíduos, respeitando um orçamento máximo de iterações por geração. Esse desenho permite concentrar esforço computacional nas soluções mais competitivas, sem comprometer a escalabilidade do método.

O módulo genético é executado periodicamente em modo preservador. Nesse estágio, operadores de recombinação e mutação são aplicados sobre indivíduos selecionados, mantendo-se elitismo e aceitando descendentes apenas quando eles não degradam o *fitness* da solução-base. Essa escolha reduz o risco de perda de boas soluções por recombinação excessiva e contribui para a estabilidade da convergência ao longo das gerações.

Por fim, a execução é encerrada por um critério global de platô, baseado na ausência de melhora significativa do melhor E_{geom} por um número consecutivo de gerações. Em conjunto, esses mecanismos definem uma estratégia híbrida voltada a equilibrar exploração global, refinamento seletivo e eficiência computacional, constituindo a principal contribuição metodológica desta dissertação.

3.5. Fluxo operacional dos experimentos

A Figura 1 apresenta o fluxo operacional adotado nos experimentos realizados neste trabalho. Inicialmente, receptor e ligante são carregados e o receptor é indexado em

uma estrutura KD-tree para acelerar as consultas geométricas. Em seguida, gera-se uma população inicial de poses em espaço 6D, que é iterativamente avaliada por E_{geom} e refinada conforme a estratégia de busca empregada. Ao final da execução, a melhor pose encontrada é submetida a *rescoring* energético por E_{vina} , sendo então registradas as métricas utilizadas na comparação entre os algoritmos.

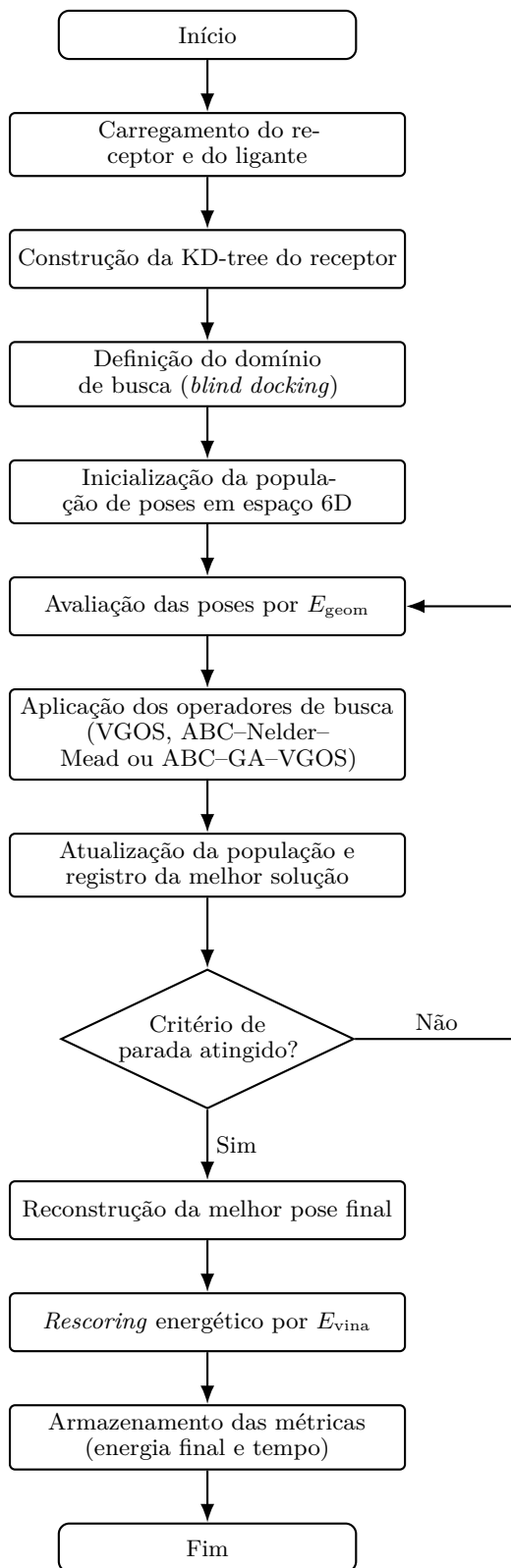


Figura 1 – Fluxo operacional geral dos experimentos de *docking* proteína-peptídeo realizados neste trabalho.

Esse fluxo é repetido independentemente para cada algoritmo e para cada complexo do conjunto experimental.

3.6. Protocolo experimental

Para garantir reprodutibilidade e comparabilidade entre os métodos, foram realizadas dez execuções independentes de cada algoritmo em cada complexo proteína-peptídeo. As execuções diferiram apenas pelas sementes pseudoaleatórias utilizadas no processo de inicialização e geração de vizinhos.

Em todos os experimentos, adotou-se população de 100 indivíduos e limite máximo de 100 gerações, resultando em até 10 000 avaliações populacionais por execução, sem contar as avaliações adicionais oriundas de refinamentos locais. A parada antecipada por paciência ou por critério de platô pode encerrar a execução antes desse limite quando não há melhora significativa.

O espaço de busca foi definido como o volume do receptor expandido por uma margem de 5 Å, caracterizando *blind docking*. Durante cada execução, os arquivos PDBQT foram convertidos em coordenadas numéricas, o algoritmo correspondente foi executado até o critério de parada, a melhor pose foi reconstruída e salva, e a energia final foi obtida por *rescoring* Vina-like. Ao final de cada rodada, registraram-se a energia mínima final e o tempo de otimização.

Os resultados foram agregados em planilhas no formato CSV, a partir das quais foram geradas tabelas-resumo e gráficos comparativos.

3.7. Métricas e análise estatística

As métricas utilizadas na comparação entre os métodos foram: (i) energia mínima média final, baseada em E_{vina} , e (ii) tempo médio de execução da etapa de otimização. A energia geométrica E_{geom} não foi utilizada como métrica final de desempenho, servindo apenas como guia interno da busca.

Para cada complexo e para cada algoritmo, foram calculados média, desvio-padrão e intervalo de confiança de 95% das métricas observadas nas dez execuções independentes. As distribuições energéticas foram ainda analisadas por meio de gráficos de barras com IC95%, *boxplots* e *violin plots*, permitindo comparar simultaneamente tendência central, dispersão e estabilidade entre os métodos.

3.8. Conjunto experimental de proteínas e peptídeos

O conjunto experimental foi composto por dez complexos proteína-peptídeo extraídos do PDB. A seleção contemplou sistemas com diferentes tamanhos de peptídeo, topologias de interface, graus de exposição superficial e níveis distintos de restrição geométrica, de modo a avaliar a robustez dos algoritmos em cenários variados de *docking* proteína-peptídeo.

A Tabela 1 apresenta os complexos utilizados nos experimentos, indicando a proteína receptora, o ligante peptídico, o código PDB e a respectiva referência estrutural.

Tabela 1 – Conjunto de complexos proteína-peptídeo utilizados nos experimentos.

Proteína	Ligante (peptídeo)	#res.	PDB	Referência
HLA-B*3501	VPLRPMTY	8	1A1N	Smith et al. (1996)
PSD-95 (domínio PDZ3)	CRIPT (peptídeo C-terminal)	var.	1BE9	Doyle et al. (1996)
Bcl-xL	Bak BH3 peptide	var.	1BXL	Sattler et al. (1997)
Penicilopepsina	PP14 (cíclico)	4–6	1BXO	Khan et al. (1998)
CDK2–cyclin A	Octapeptídeo H-His...-Phe-NH ₂	8	1JSU	Atkinson et al. (2002)
MDM2	p53 TAD peptide	var.	1YCR	Kussie et al. (1996)
Siah ubiquitin ligase (mouse)	Degron motif peptide	var.	2AN6	House et al. (2006)
EphB4 receptor (human)	Ephrin-B2 antagonist peptide	var.	2BBA	Chrencik et al. (2006)
Antibody-peptide (mouse)	Synthetic peptide homolog of Mhr	var.	2IGF	Stanfield et al. (1990)
Bcl-xL (human)	BID/BIM BH3 peptides	var.	4QVE	Rajan et al. (2015)

Os complexos selecionados abrangem diferentes contextos estruturais de interação proteína-peptídeo. O complexo 1A1N corresponde ao MHC I HLA-B*3501 em interação com o octapeptídeo VPLRPMTY, apresentando um sulco de ligação bem definido, característico de sistemas MHC-peptídeo. O complexo 1BE9 envolve o domínio PDZ3 da proteína PSD-95 associado ao peptídeo C-terminal CRIPT, representando uma interface superficial relevante para reconhecimento molecular mediado por domínios PDZ. O complexo 1JSU representa o sistema CDK2–cyclin A ligado a um octapeptídeo inibidor, com região de interação bem delimitada e associada ao reconhecimento regulatório em quinases.

Os complexos 1BXL e 4QVE envolvem a proteína Bcl-xL em associação com peptídeos derivados de proteínas pró-apoptóticas, como Bak, BID ou BIM. Esses sistemas foram incluídos por apresentarem regiões de interação hidrofóbicas relevantes para o reconhecimento molecular e para a avaliação da capacidade dos algoritmos em posicionar peptídeos em sulcos de ligação. O complexo 1BXO, por sua vez, corresponde à penicilopepsina em complexo com um peptídeo inibidor, representando um caso com cavidade mais restritiva e com maior exigência de compatibilidade geométrica.

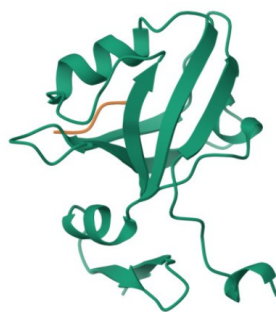
O complexo 1YCR descreve a interação entre a proteína MDM2 e o peptídeo p53 TAD, um sistema frequentemente utilizado em estudos de reconhecimento proteína-peptídeo por envolver acomodação do peptídeo em uma interface superficial. O complexo 2IGF representa uma interação anticorpo-peptídeo, com interface relativamente exposta e potencialmente mais desafiadora do ponto de vista conformacional. Já o complexo 2BBA envolve o receptor EphB4 humano em interação com um peptídeo antagonista Ephrin-B2, enquanto o complexo 2AN6 corresponde à Siah ubiquitin ligase murina associada a um motivo peptídico do tipo *degron*. Esses sistemas ampliam a diversidade do conjunto experimental por incluírem diferentes perfis de interface, reconhecimento molecular e exposição superficial.

As Figuras 2 e 3 apresentam representações visuais dos complexos proteína-peptídeo utilizados nos experimentos, com o objetivo de ilustrar a diversidade estrutural do conjunto avaliado.

Figura 2 – Representações estruturais de parte dos complexos proteína-peptídeo utilizados nos experimentos.



(a) 1A1N



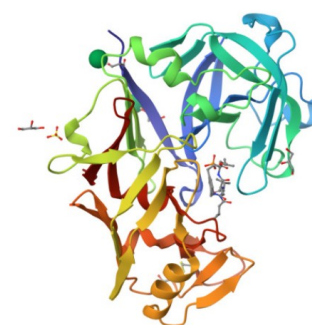
(b) 1BE9



(c) 1JSU

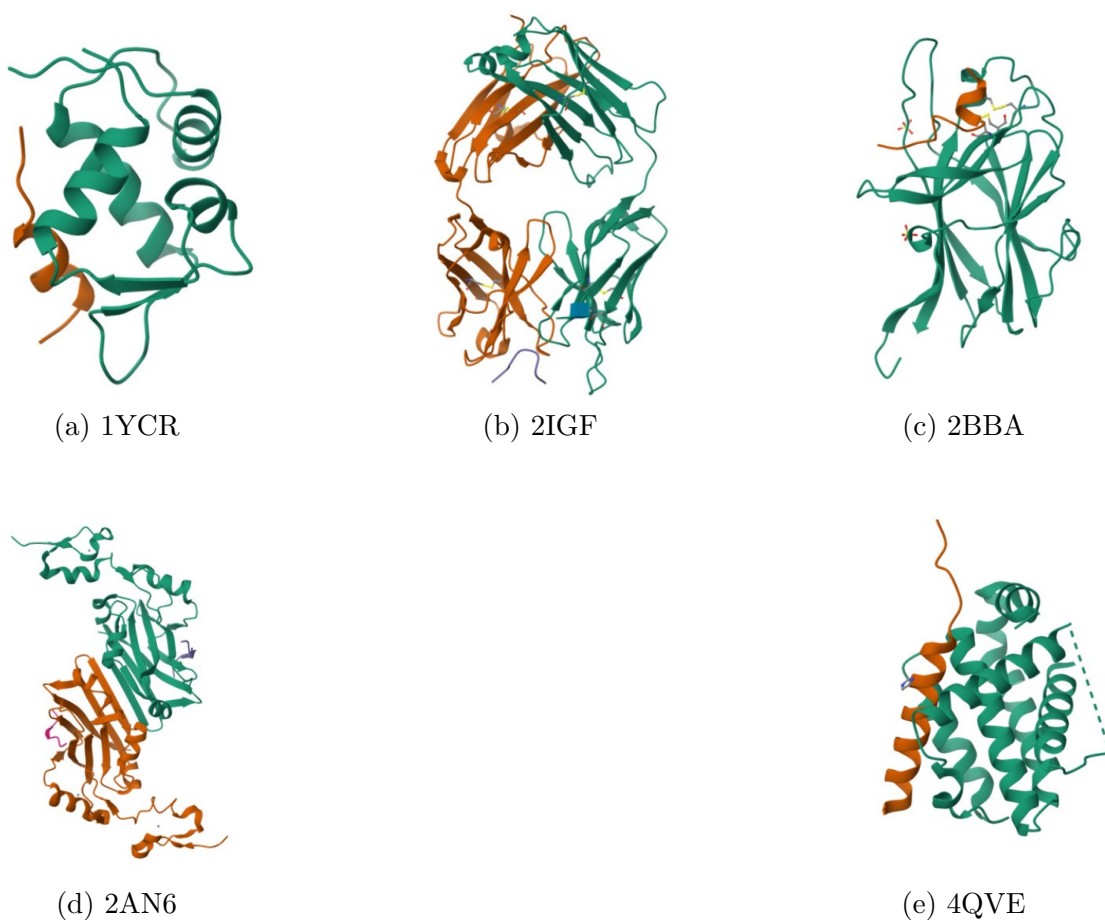


(d) 1BXL



(e) 1BXO

Figura 3 – Representações estruturais dos demais complexos proteína-peptídeo utilizados nos experimentos.



Os complexos selecionados abrangem diferentes tipos de interface, incluindo cavidades mais profundas, bolsões superficiais e regiões de interação relativamente expostas. Essa diversidade estrutural contribui para uma avaliação mais abrangente do comportamento dos algoritmos em cenários distintos de *docking* proteína-peptídeo.

3.9. Infraestrutura computacional

Os experimentos foram executados em um processador Intel Core i7-1255U, com 8 GB de memória RAM, sistema operacional Windows 11 de 64 bits e execução em modo *single-thread*. Essa configuração foi mantida constante em todas as rodadas experimentais, evitando interferência de paralelismo ou variações de ambiente na comparação entre os algoritmos.

4 Resultados e Discussão

4.1. Introdução aos resultados

Esta seção apresenta e discute os resultados obtidos a partir dos experimentos realizados com dez complexos proteína-peptídeo representativos. As estruturas foram selecionadas a partir do *Protein Data Bank* (PDB), abrangendo diferentes famílias moleculares e níveis de flexibilidade estrutural. A diversidade desses sistemas permitiu avaliar o comportamento dos algoritmos em contextos variados de interação, incluindo cavidades amplas, bolsões de ligação bem definidos e interfaces superficiais de contato.

Os experimentos foram conduzidos sob condições padronizadas e controladas, de modo a assegurar a comparabilidade entre os métodos avaliados. Para cada complexo, cada algoritmo foi executado dez vezes de forma independente, com população de 100 indivíduos e **até** 100 gerações (limite máximo), totalizando **até** 10 000 avaliações **da população** da função objetivo geométrica (E_{geom}) por execução (100 indivíduos \times 100 gerações). As execuções podem encerrar antes desse limite por critério de parada antecipada (*early-stop*), quando não há melhoria significativa ao longo de um número consecutivo de gerações. Quando aplicados, os procedimentos de refinamento local (e.g., Solis–Wets ou Nelder–Mead) realizam **avaliações adicionais** de E_{geom} , porém com **orçamento de iterações/avaliações** e critérios de parada próprios, de modo a não inflacionar o custo computacional de forma descontrolada. As repetições utilizaram diferentes sementes pseudoaleatórias, reduzindo a influência de variações estocásticas e conferindo maior robustez estatística às médias obtidas. Ao final de cada execução, a melhor pose (selecionada por E_{geom}) foi reavaliada por *rescoring* Vina-like, produzindo E_{vina} (kcal/mol), valor utilizado para compor as tabelas e figuras desta seção.

No cômputo geral, o híbrido ABC–GA–VGOS apresentou desempenho consistente: obteve a melhor energia média em **5/10** complexos, enquanto o VGOS liderou energeticamente em quatro sistemas e o ABC–Nelder–Mead em um. Em termos de eficiência, o híbrido apresentou os **menores tempos médios em todos os dez complexos** avaliados, mantendo boa estabilidade de convergência e *reprodutibilidade* em diferentes contextos estruturais, ao conciliar exploração ampla e intensificação local controlada.

Para facilitar a visualização do panorama geral, a Tabela 2 apresenta o **melhor algoritmo identificado para cada complexo** (isto é, aquele com menor energia média E_{vina}), acompanhado dos respectivos valores médios de energia e tempo de execução. Ressalta-se que, por esse critério de síntese, o **tempo reportado na tabela refere-se ao algoritmo vencedor em energia** em cada sistema, não constituindo, por si só, uma comparação direta de tempos entre os métodos para um mesmo complexo. A análise

comparativa de desempenho computacional é apresentada de forma mais apropriada nas tabelas específicas de cada complexo, nas quais os tempos médios dos três algoritmos são exibidos lado a lado.

Tabela 2 – Resultados por proteína: energia média e tempo médio de execução.

Complexo	Algoritmo	Energia (kcal/mol)	Tempo (s)
1A1N	ABC-GA-VGOS	-9,50	10,34
1BE9	ABC-GA-VGOS	-5,50	6,43
1BXL	VGOS	-14,01	17,83
1BXO	VGOS	-10,53	8,76
1JSU	VGOS	-14,50	60,59
1YCR	ABC-GA-VGOS	-10,64	11,32
2AN6	ABC-Nelder-Mead	-9,00	13,72
2BBA	ABC-GA-VGOS	-10,28	9,96
2IGF	VGOS	-9,19	7,78
4QVE	ABC-GA-VGOS	-1,12	3,09

A Tabela 2 evidencia a predominância do híbrido em diferentes topologias de sítio (cavidades profundas, bolsões rasos e interfaces superficiais), destacando sua robustez e estabilidade ao obter bons resultados médios. Em complemento, as tabelas e figuras específicas de cada complexo permitem comparar diretamente *energia, variabilidade e tempo* entre os três métodos, reforçando o comportamento consistente do híbrido em termos de convergência e reprodutibilidade.

A capacidade do híbrido de ajustar dinamicamente o equilíbrio entre as fases exploratória e refinadora resultou em menor tempo de convergência e melhor qualidade de energia média.

A análise estatística complementar envolveu testes de variância unifatorial (ANOVA) e o teste não paramétrico de Friedman, aplicados às distribuições de energia obtidas em cada conjunto de proteínas. Além disso, foram calculados os intervalos de confiança de 95% (IC95) para cada métrica, permitindo avaliar a significância das diferenças observadas. Esses procedimentos asseguraram que as conclusões fossem baseadas em evidências quantitativas e não em flutuações aleatórias do processo estocástico.

As subseções seguintes apresentam a análise individual de cada complexo proteína-peptídeo, destacando os comportamentos específicos de convergência e a influência das características estruturais de cada receptor no desempenho relativo dos algoritmos.

4.2. Resultados individuais por complexo

Complexo 1A1N

Os resultados obtidos para este complexo são apresentados na Figura 4 e na Tabela 3. Em termos energéticos, o método híbrido **ABC-GA-VGOS** alcançou a menor média de

energia de ligação, com valor de $-9,50$ kcal/mol e intervalo de confiança de 95% entre $[-11,98, -7,01]$. Em comparação, os algoritmos **ABC–Nelder–Mead** e **VGOS** apresentaram médias de $-8,09$ e $-7,64$ kcal/mol, respectivamente, com intervalos de confiança $[-10,23, -5,94]$ e $[-9,33, -5,94]$, indicando uma tendência de melhor desempenho do híbrido.

O tempo médio de execução do método proposto foi de aproximadamente **10,34 s**, similar ao dos métodos individuais, que variaram entre 10,66 s (VGOS) e 12,95 s (ABC–Nelder–Mead). A diferença é pequena, mostrando que o híbrido manteve desempenho computacional competitivo mesmo com a inclusão de operadores genéticos e refinamento local.

Os testes estatísticos não apontaram diferenças significativas entre as médias energéticas (ANOVA: $F(2, 27) = 0,86$, $p = 0,436$; Friedman: $Q = 1,40$, $p = 0,497$). Assim, embora o híbrido apresente a menor média energética entre os métodos avaliados, não se rejeita a hipótese nula de igualdade de médias para este conjunto. A interpretação mais adequada é que o ABC–GA–VGOS apresentou desempenho energético competitivo, mas com variabilidade ainda relevante entre as execuções.

A Figura 5 ilustra a distribuição das energias obtidas para cada algoritmo, tanto em formato de *boxplot* quanto de gráfico *violin*. Observa-se que o método híbrido concentrou seus valores em torno de regiões mais negativas, enquanto o VGOS e o ABC–Nelder–Mead apresentaram maior assimetria e dispersão. Essa concentração reforça a ideia de que o híbrido é capaz de explorar com maior eficiência o espaço conformacional, alcançando soluções de menor energia com maior regularidade.

A Figura 6 apresenta o tempo médio por execução. Ainda que o híbrido demande tempo semelhante de processamento, seu desempenho energético justifica a **complexidade adicional** da estratégia híbrida, sobretudo considerando a estabilidade das soluções encontradas e a ausência de convergência precoce observada nos métodos individuais.

De modo geral, o complexo 1A1N demonstra que o equilíbrio entre exploração global (via componentes evolutivos) e refinamento local (via ajustes direcionados) permite ao método híbrido localizar poses de maior afinidade, evidenciando sua capacidade de generalização mesmo em sistemas de baixa flexibilidade estrutural.

Tabela 3 – Resultados médios obtidos para o complexo 1A1N (dez execuções independentes).

Algoritmo	Energia (Média)	Desvio-padrão	Tempo (s)	Desvio-padrão	IC95 Energia
ABC–GA–VGOS	-9,50	3,47	10,34	2,44	[-11,98, -7,01]
ABC–Nelder–Mead	-8,09	3,00	12,95	1,94	[-10,23, -5,94]
VGOS	-7,64	2,37	10,66	0,63	[-9,33, -5,94]

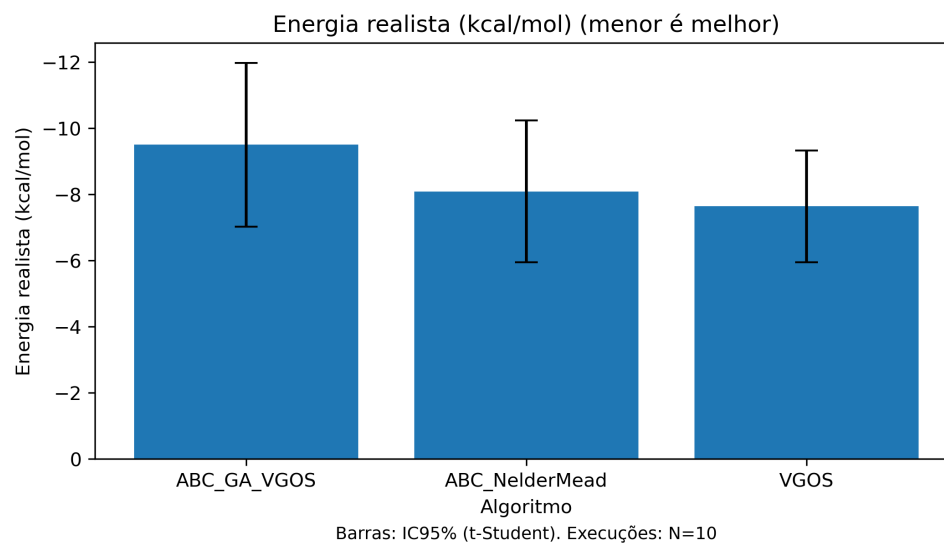
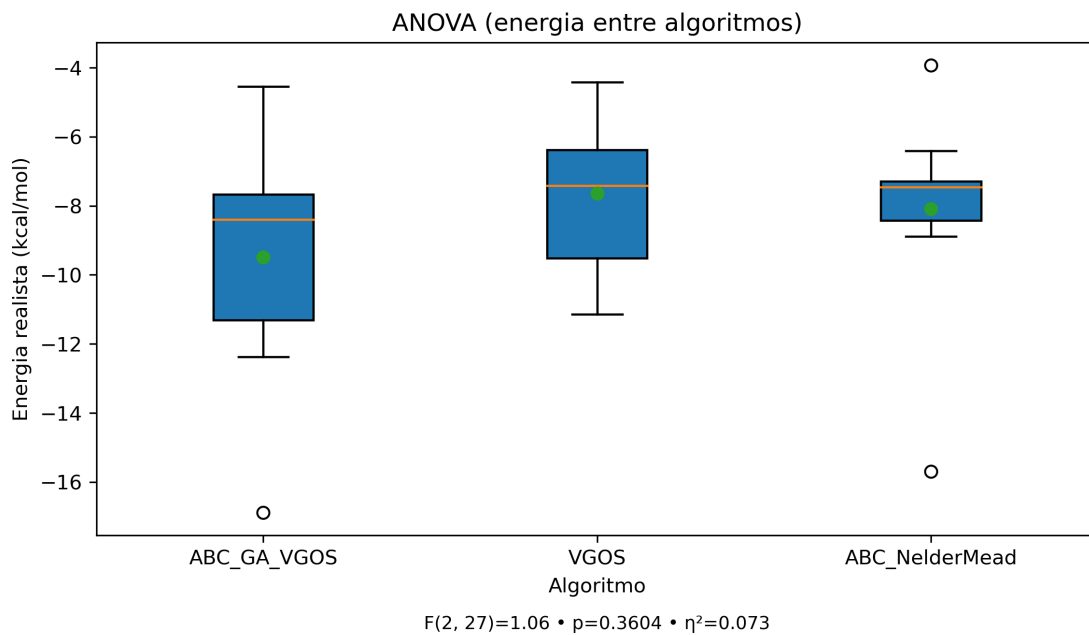
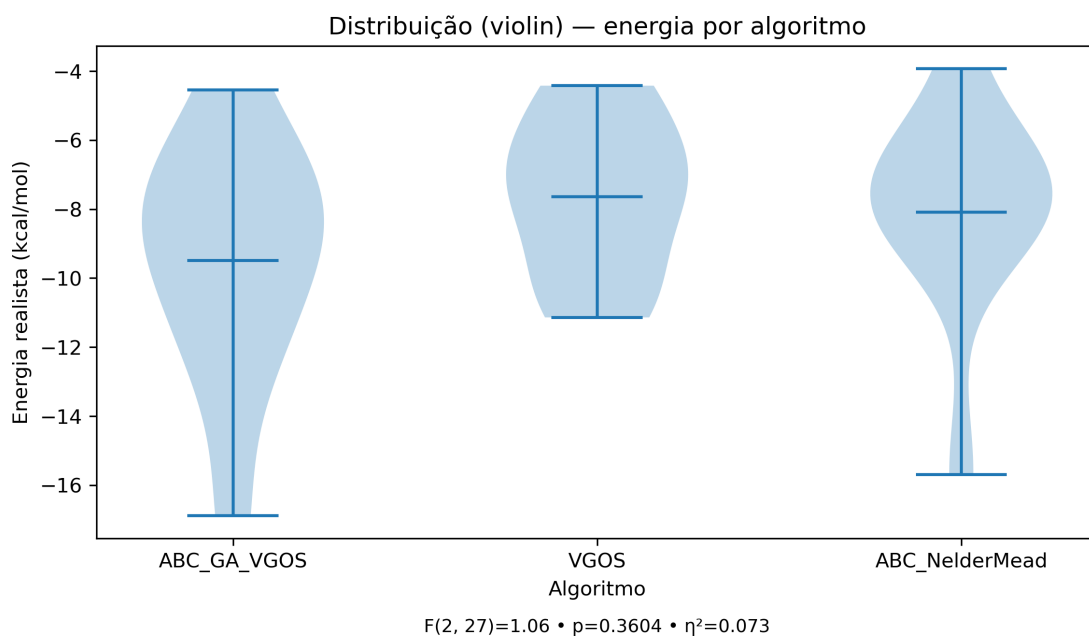


Figura 4 – Médias e intervalos de confiança de 95% para a energia de ligação no complexo 1A1N.



(a) Boxplot das energias obtidas por algoritmo.



(b) Distribuição *violin* das energias obtidas por algoritmo.

Figura 5 – Distribuição das energias obtidas por algoritmo para o complexo 1A1N (ANOVA e Friedman).

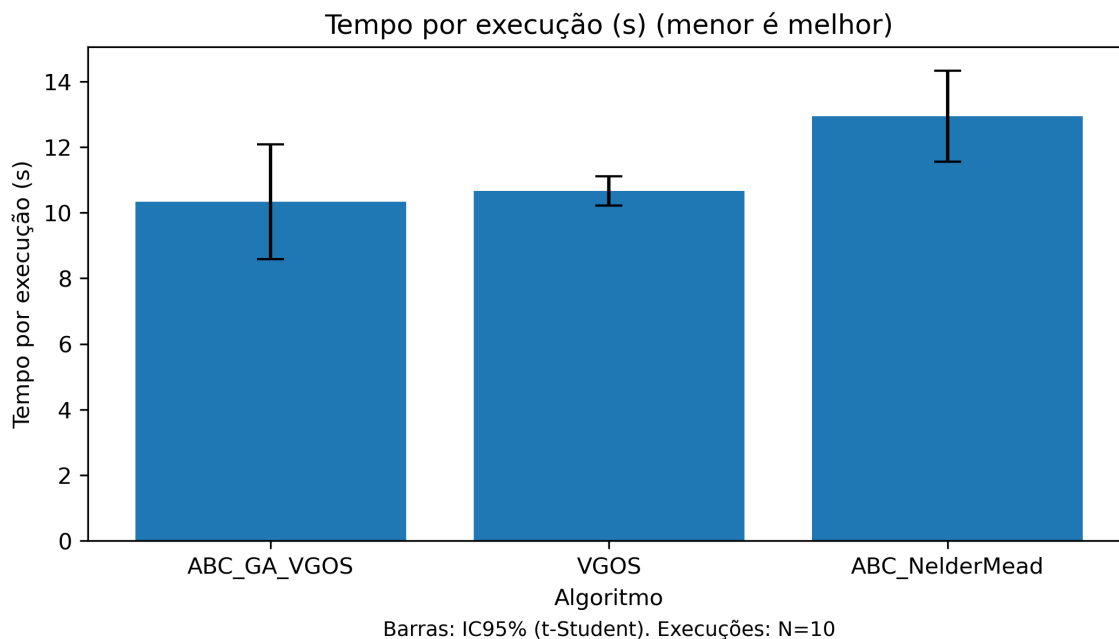


Figura 6 – Tempo médio de execução com intervalos de confiança de 95% para o complexo 1A1N.

Complexo 1BE9

Os resultados obtidos para o complexo 1BE9 são apresentados na Tabela 4 e nas Figuras 7, 8 e 9.

O híbrido **ABC-GA-VGOS** obteve a menor energia média entre os métodos avaliados, além de apresentar menor variabilidade entre as execuções, indicando convergência mais estável. O **VGOS** apresentou desempenho próximo, enquanto o **ABC-Nelder-Mead** resultou em médias energéticas menos favoráveis.

Embora a ANOVA não tenha indicado diferença estatisticamente significativa entre os algoritmos ($p > 0,05$), observa-se uma tendência favorável ao método híbrido, tanto pela menor média energética quanto pelo menor custo temporal. Assim, para o complexo 1BE9, o **ABC-GA-VGOS** apresentou o melhor equilíbrio entre qualidade energética, estabilidade e tempo de execução.

Tabela 4 – Resultados médios obtidos para o complexo 1BE9 (dez execuções independentes).

Algoritmo	Energia Média (kcal/mol)	Desvio-padrão	Tempo Médio (s)	Desvio-padrão	IC95 da Energia
ABC-GA-VGOS	-5,50	2,62	6,43	0,45	[-7,37, -3,62]
VGOS	-4,52	2,26	7,14	0,26	[-6,13, -2,91]
ABC-Nelder-Mead	-3,33	1,94	8,44	0,20	[-4,72, -1,94]

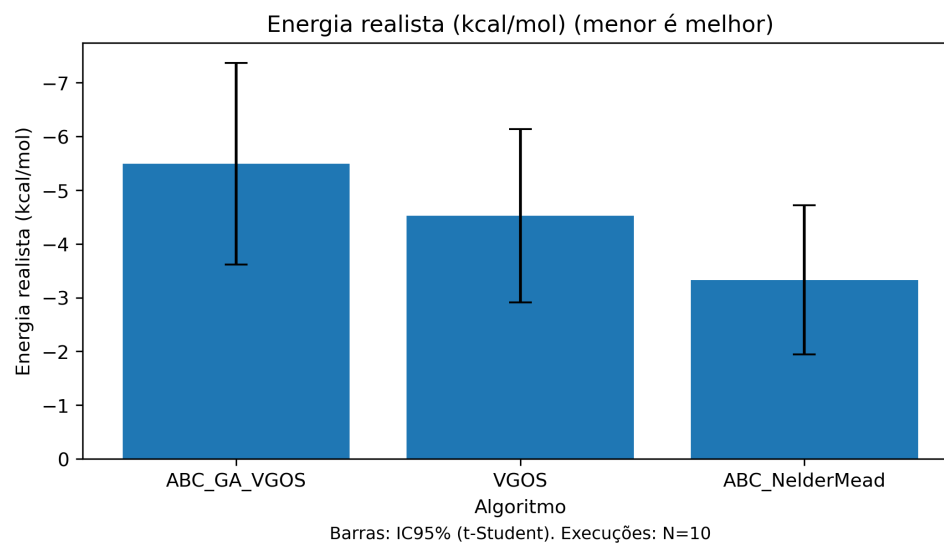
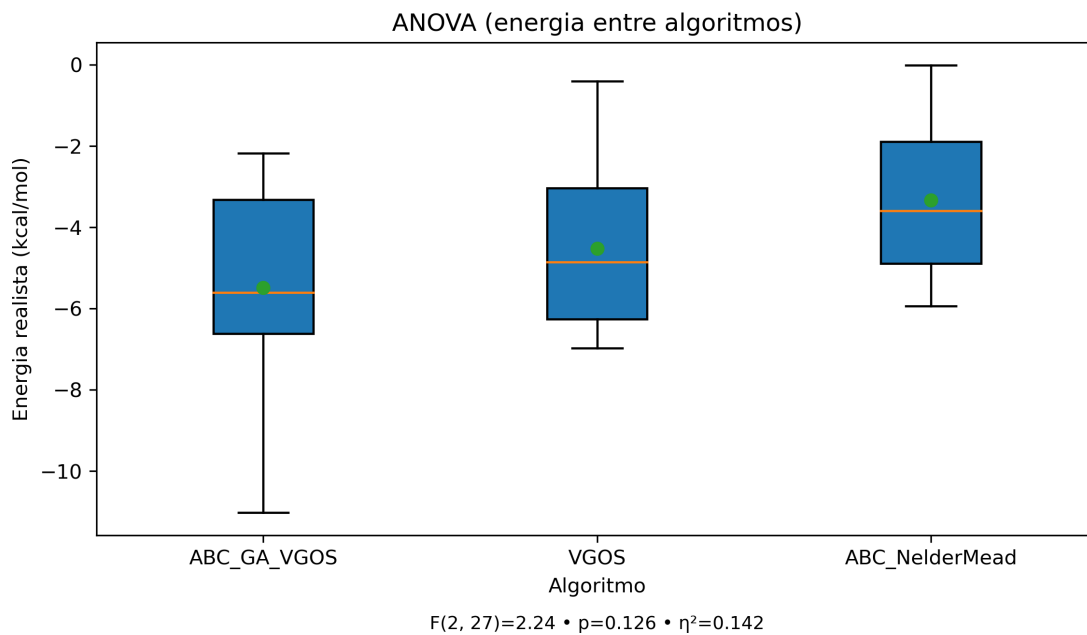


Figura 7 – Médias e intervalos de confiança (IC95) da energia de ligação para o complexo 1BE9.



(a) Boxplot das energias obtidas por algoritmo.

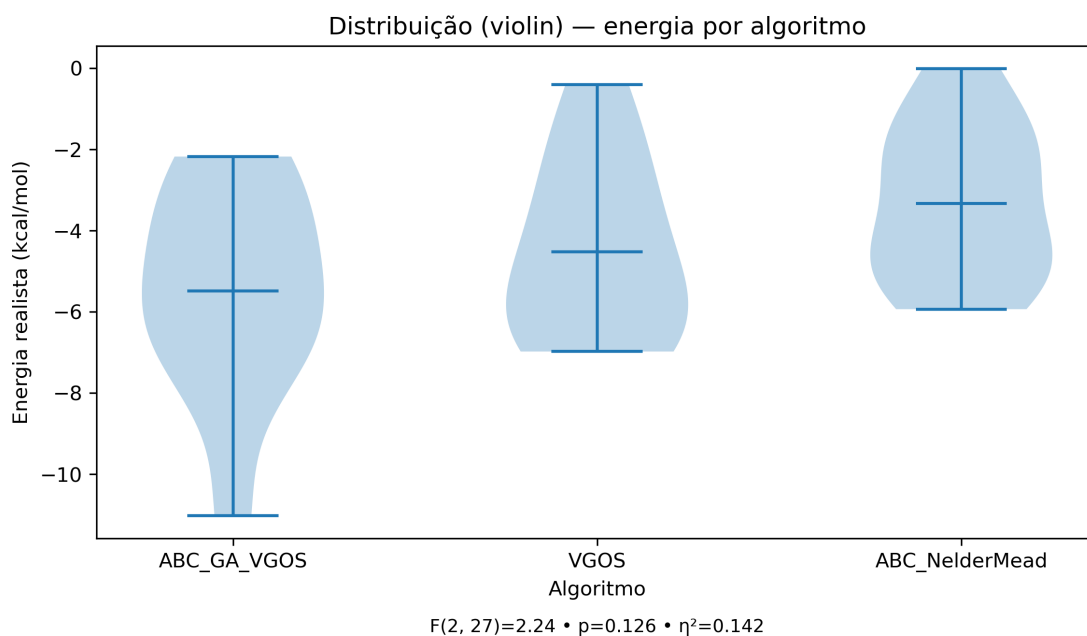
(b) Distribuição *violin* das energias obtidas por algoritmo.

Figura 8 – Distribuição das energias obtidas por algoritmo para o complexo 1BE9 (ANOVA e Friedman).

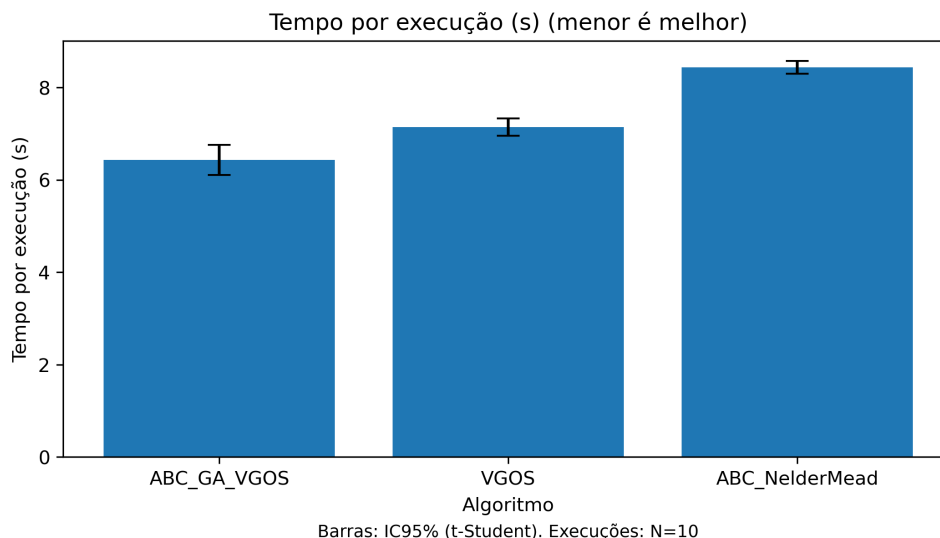


Figura 9 – Tempo médio de execução e intervalos de confiança (IC95) para o complexo 1BE9.

A próxima análise aborda o complexo **1JSU**, sistema *CDK2-Cyclin A* ligado a um octapeptídeo inibidor.

Complexo 1JSU

A Tabela 5 apresenta os resultados obtidos para o complexo 1JSU. Observa-se que o algoritmo **VGOS** alcançou a menor energia média de ligação, indicando melhor desempenho energético entre os métodos avaliados neste sistema. O **ABC-GA-VGOS** apresentou desempenho competitivo, embora com energia média menos favorável que a do VGOS. Entretanto, seu desvio-padrão reduzido sugere maior estabilidade e consistência entre as execuções. O método **ABC-Nelder-Mead**, por sua vez, apresentou energia média menos favorável e maior custo computacional.

Os tempos médios de execução seguiram a ordem esperada, sendo o híbrido ligeiramente mais eficiente que o VGOS puro e consideravelmente mais rápido que o ABC-Nelder-Mead. Apesar das diferenças energéticas aparentes, a análise de variância (ANOVA) não apontou diferença estatisticamente significativa entre as médias ($p > 0,05$). De modo geral, o híbrido manteve equilíbrio entre qualidade energética e tempo de execução, evidenciando boa robustez para cavidades de ligação compactas como a de 1JSU.

Tabela 5 – Resultados médios obtidos para o complexo 1JSU (dez execuções independentes).

Algoritmo	Energia Média (kcal/mol)	Desvio-padrão	Tempo Médio (s)	Desvio-padrão	IC95 da Energia
ABC-GA-VGOS	-10,79	4,88	56,98	6,50	[-14,28, -7,30]
VGOS	-14,50	5,48	60,59	11,59	[-18,42, -10,58]
ABC-Nelder-Mead	-10,28	3,10	76,58	11,11	[-12,50, -8,06]

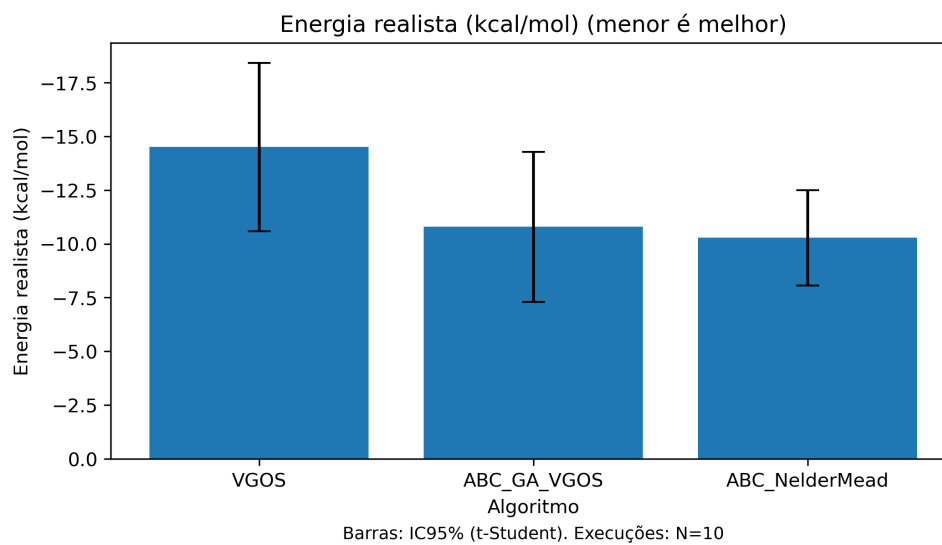
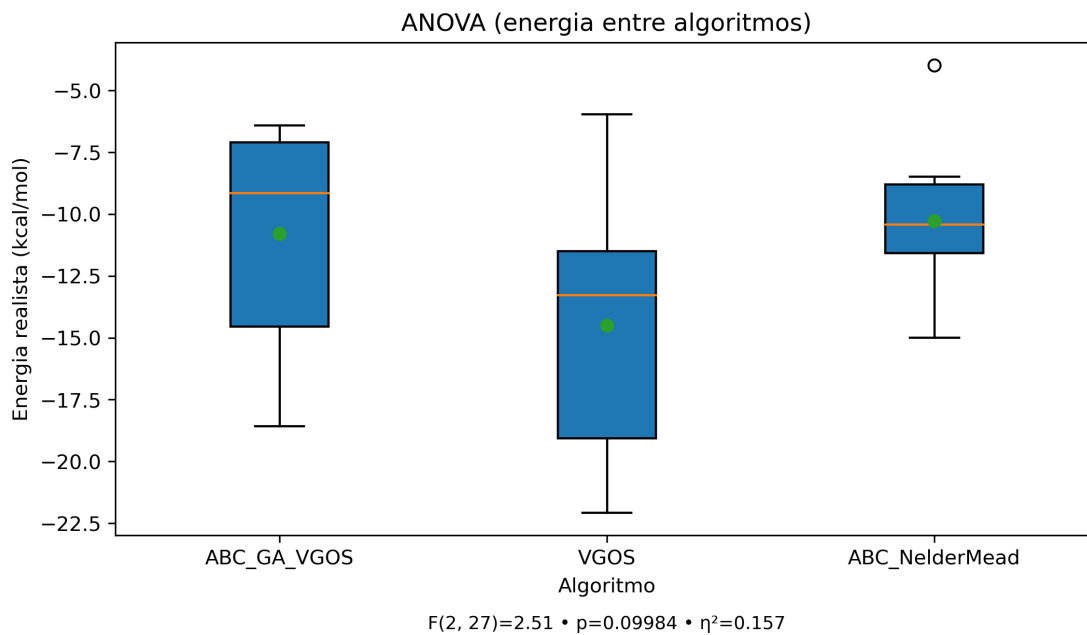


Figura 10 – Médias e intervalos de confiança (IC95) da energia de ligação para o complexo 1JSU.



(a) Boxplot das energias obtidas por algoritmo.

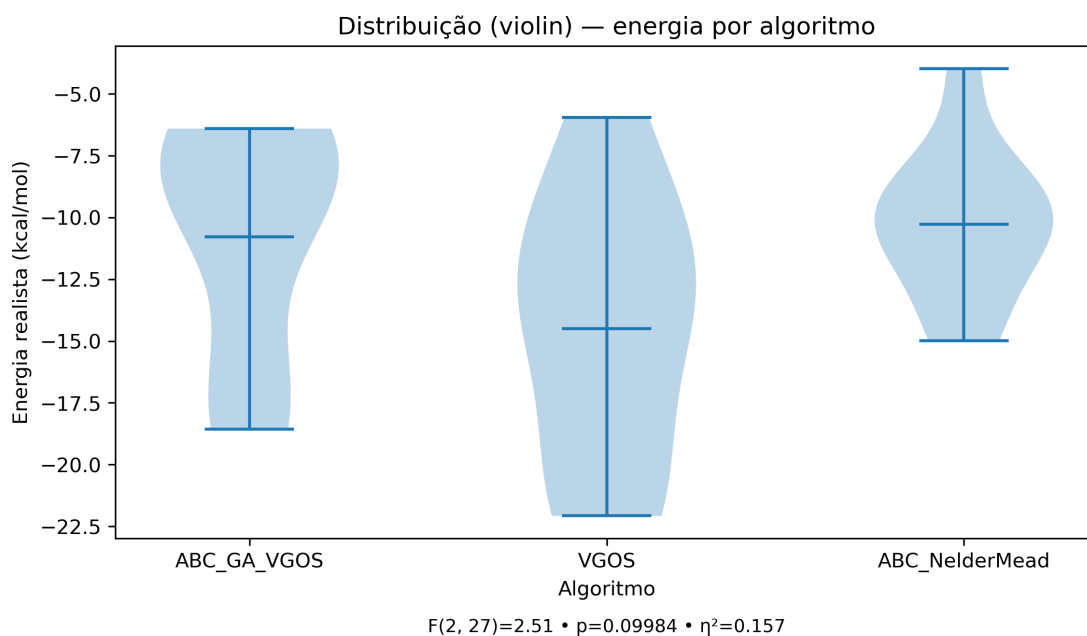
(b) Distribuição *violin* das energias obtidas por algoritmo.

Figura 11 – Distribuição das energias obtidas por algoritmo para o complexo 1JSU (ANOVA e Friedman).

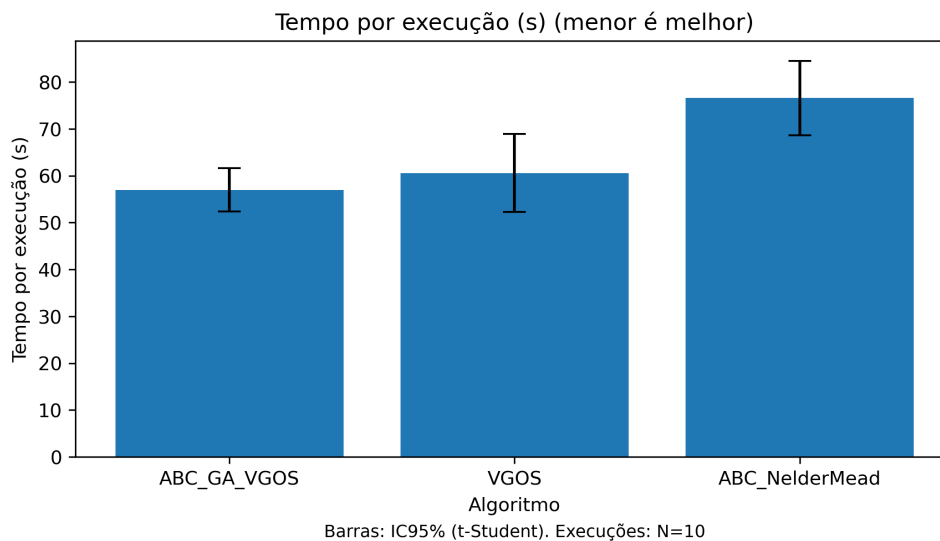


Figura 12 – Tempo médio de execução e intervalos de confiança (IC95) para o complexo 1JSU.

Complexo 1BXL

A Tabela 6 apresenta os resultados obtidos para este complexo. Observa-se que o algoritmo **VGOS** atingiu a menor energia média de ligação, enquanto o modelo **ABC-GA-VGOS** apresentou desempenho muito próximo, com variação ligeiramente menor entre execuções. O método **ABC-Nelder-Mead**, por sua vez, resultou em energias médias mais altas (menos favoráveis) e maior tempo de processamento.

Em termos de eficiência, o **ABC-GA-VGOS** apresentou, nas análises por complexo, os menores tempos médios de execução nas condições avaliadas, mantendo custo computacional inferior aos métodos individuais.

Tabela 6 – Resultados médios obtidos para o complexo 1BXL (dez execuções independentes).

Algoritmo	Energia Média (kcal/mol)	Desvio-padrão	Tempo Médio (s)	Desvio-padrão	IC95 da Energia
ABC-GA-VGOS	-13,61	3,57	16,23	2,50	[-16,16, -11,06]
VGOS	-14,01	3,31	17,83	2,66	[-16,38, -11,64]
ABC-Nelder-Mead	-11,70	2,94	21,53	3,06	[-13,80, -9,60]

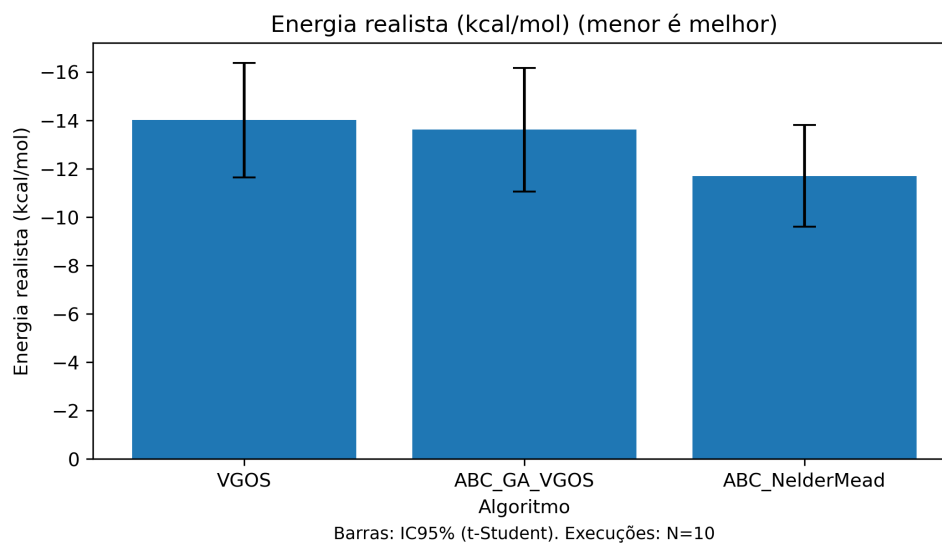
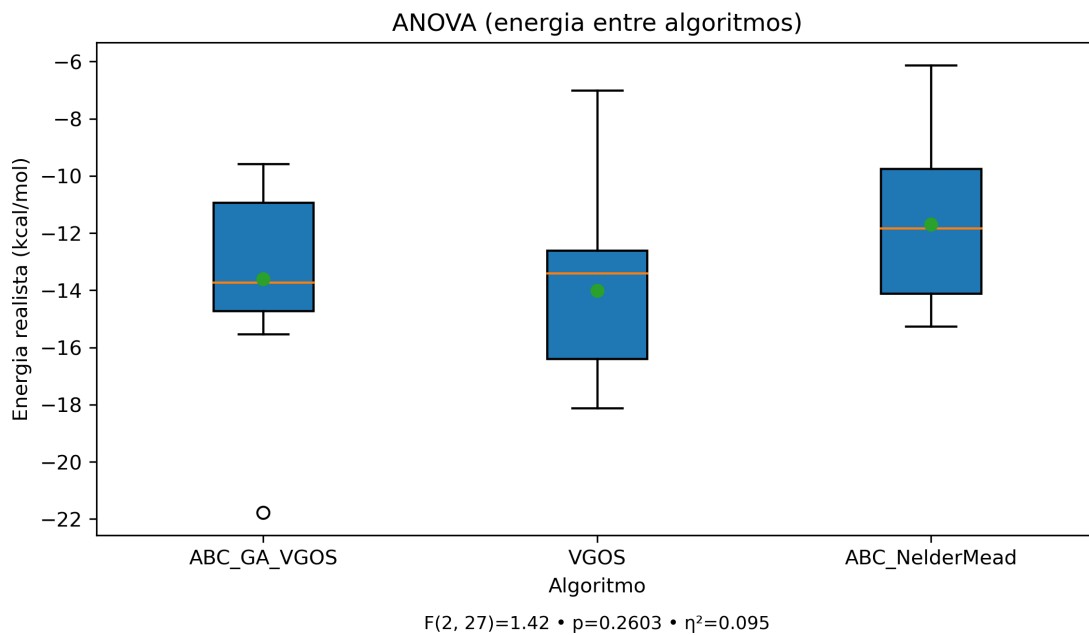
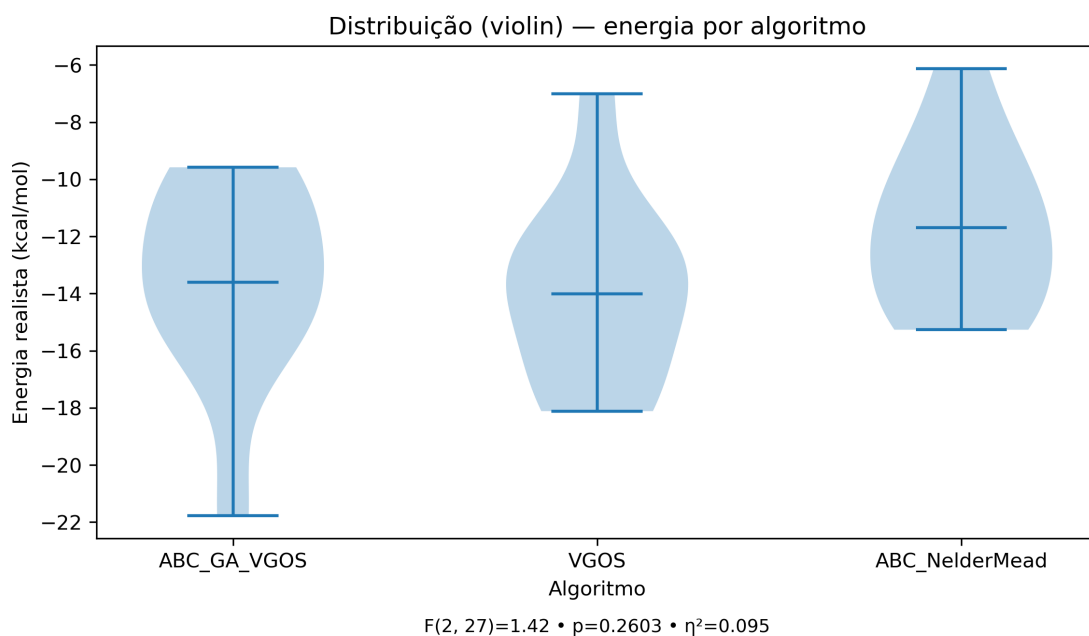


Figura 13 – Médias e intervalos de confiança (IC95) da energia de ligação para o complexo 1BXL.



(a) Boxplot das energias obtidas por algoritmo.



(b) Distribuição *violin* das energias obtidas por algoritmo.

Figura 14 – Distribuição das energias obtidas por algoritmo para o complexo 1BXL (ANOVA e Friedman).

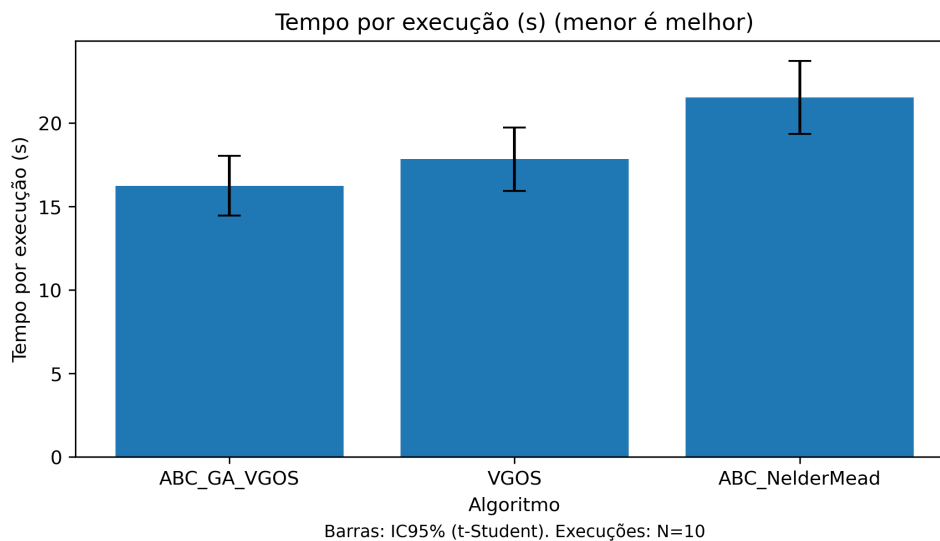


Figura 15 – Tempo médio de execução e intervalos de confiança (IC95) para o complexo 1BXL.

Complexo 1BXO

A Tabela 7 apresenta os resultados obtidos para o complexo 1BXO. Observa-se que os métodos **ABC-GA-VGOS** e **VGOS** apresentaram energias médias bastante próximas, ambas mais favoráveis que as do **ABC-Nelder-Mead**. O algoritmo híbrido manteve desempenho consistente, com menor desvio-padrão e o tempo médio mais baixo entre as três abordagens, indicando bom equilíbrio entre exploração global e refinamento local.

Apesar de o método VGOS ter alcançado a menor energia média absoluta, as diferenças entre ele e o híbrido não foram estatisticamente significativas ($p > 0,05$). O **ABC-Nelder-Mead**, por outro lado, apresentou as piores energias e o maior tempo de execução, refletindo o custo computacional adicional do refinamento Nelder-Mead. De forma geral, o híbrido manteve estabilidade e bom desempenho para este sistema, com convergência eficiente em cavidades estreitas e com geometria complexa.

Tabela 7 – Resultados médios obtidos para o complexo 1BXO (dez execuções independentes).

Algoritmo	Energia Média (kcal/mol)	Desvio-padrão	Tempo Médio (s)	Desvio-padrão	IC95 da Energia
ABC-GA-VGOS	-10,47	2,57	7,55	0,50	[-12,31, -8,63]
VGOS	-10,53	3,72	8,76	1,82	[-13,19, -7,87]
ABC-Nelder-Mead	-6,96	2,47	10,83	2,06	[-8,73, -5,20]

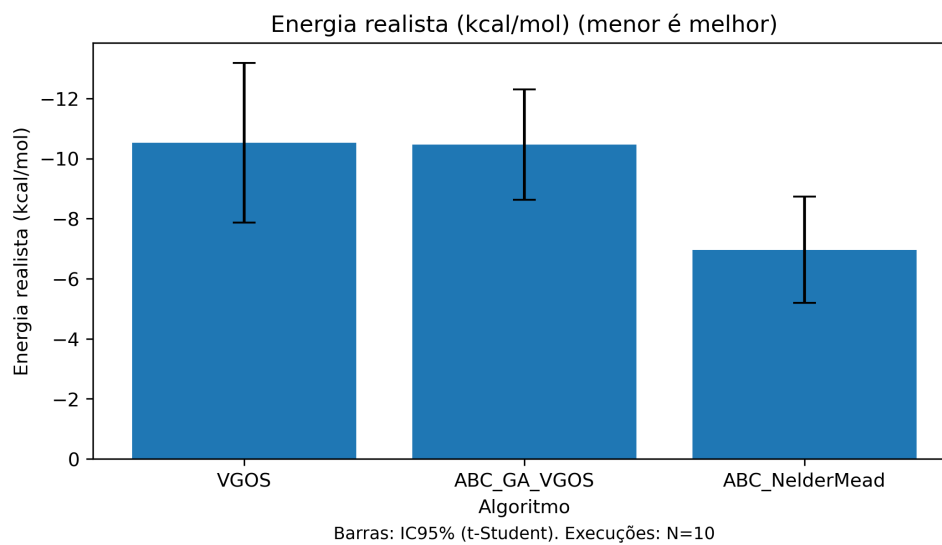
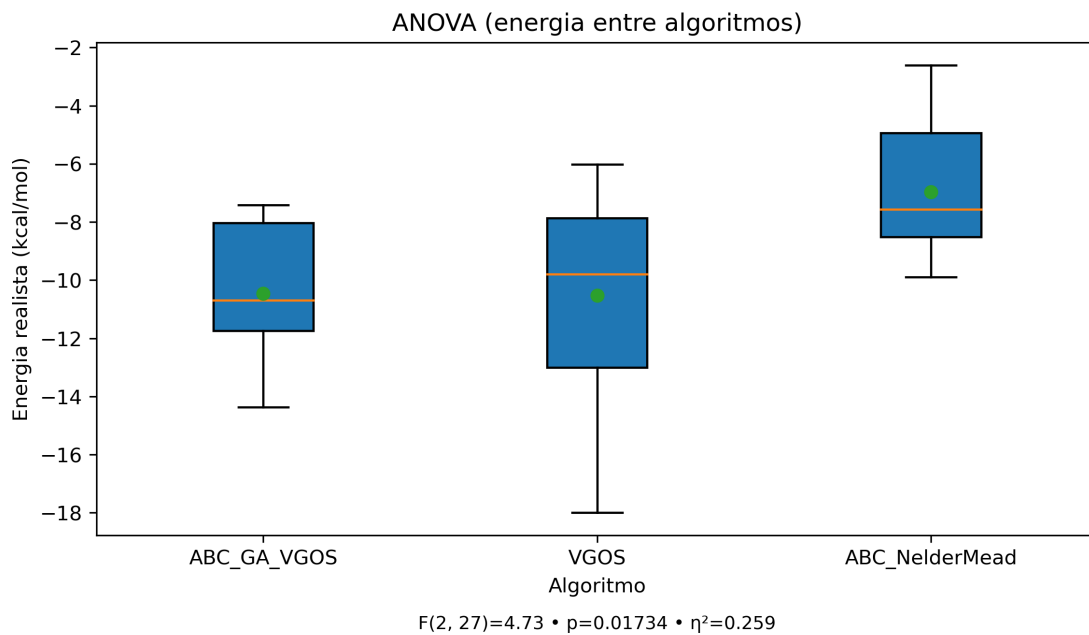
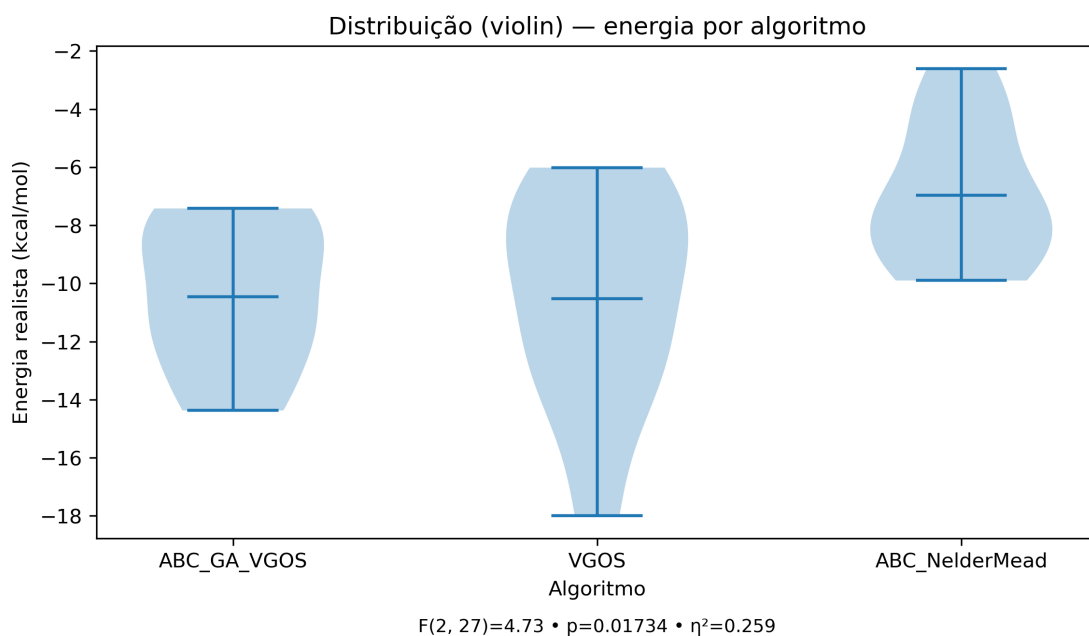


Figura 16 – Médias e intervalos de confiança (IC95) da energia de ligação para o complexo 1BXO.



(a) Boxplot das energias obtidas por algoritmo.



(b) Distribuição *violin* das energias obtidas por algoritmo.

Figura 17 – Distribuição das energias obtidas por algoritmo para o complexo 1BXO (ANOVA e Friedman).

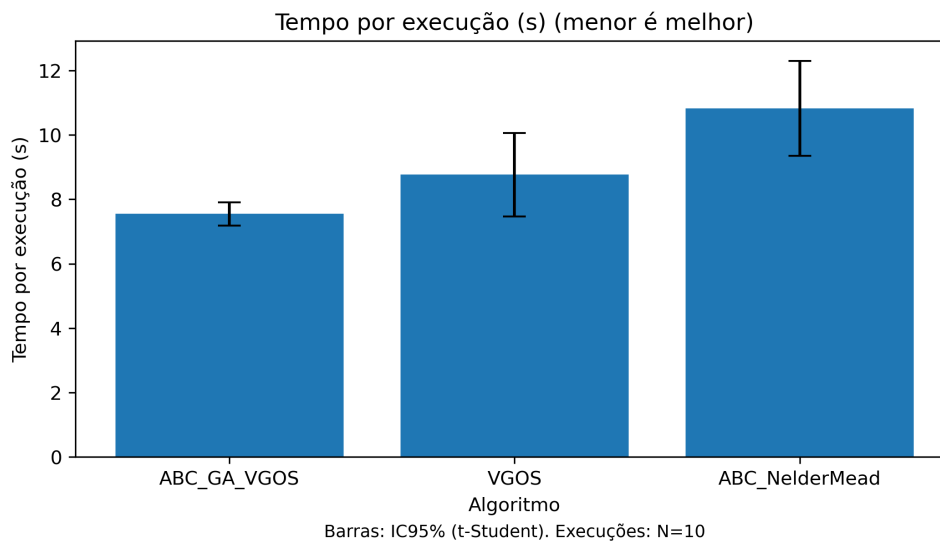


Figura 18 – Tempo médio de execução e intervalos de confiança (IC95) para o complexo 1BXO.

Complexo 1YCR

A Tabela 8 e as Figuras 19 a 21 sintetizam os resultados obtidos. O algoritmo **ABC-GA-VGOS** apresentou a menor energia média, com $-10,64$ kcal/mol (IC95: $[-13,57, -7,70]$), seguido do **ABC-Nelder-Mead** com $-9,23$ kcal/mol (IC95: $[-11,36, -7,11]$) e do **VGOS** com $-8,19$ kcal/mol (IC95: $[-9,55, -6,84]$). Embora as diferenças absolutas não sejam elevadas, observa-se tendência consistente de menor energia média para o modelo híbrido, acompanhada de menor amplitude dos intervalos de confiança em comparação aos demais métodos.

A análise de variância indicou ausência de diferença estatisticamente significativa entre os algoritmos ($p > 0,05$), sugerindo que as variações energéticas observadas situam-se dentro da variabilidade esperada para este sistema. Ainda assim, o comportamento médio do híbrido evidencia desempenho competitivo, mantendo energias mais favoráveis ao longo das execuções.

Quanto ao desempenho computacional, o híbrido apresentou tempo médio de **11,32 s**, ligeiramente inferior ao VGOS (**11,93 s**) e menor que o ABC-Nelder-Mead (**14,90 s**). Esse resultado indica que o ganho energético observado não implicou aumento de custo computacional, mantendo equilíbrio entre qualidade energética estimada e tempo de execução.

As Figuras 20 mostram a distribuição das energias nos formatos de *boxplot* e *violin*, evidenciando sobreposição das distribuições e ausência de outliers expressivos. A Figura 19 apresenta as médias energéticas com intervalos de confiança de 95%, enquanto a Figura 21 compara os tempos médios de execução entre os algoritmos.

Em síntese, no complexo 1YCR o algoritmo **ABC-GA-VGOS** apresentou a melhor média energética associada a tempo computacional competitivo, mantendo desempenho

consistente ao longo das execuções avaliadas.

Tabela 8 – Resultados médios obtidos para o complexo 1YCR (dez execuções independentes).

Algoritmo	Energia Média (kcal/mol)	Desvio-padrão	Tempo Médio (s)	Desvio-padrão	IC95 da Energia
ABC-GA-VGOS	-10,64	4,10	11,32	1,44	[-13,57, -7,70]
VGOS	-8,19	1,89	11,93	0,87	[-9,55, -6,84]
ABC-Nelder-Mead	-9,23	2,97	14,90	1,14	[-11,36, -7,11]

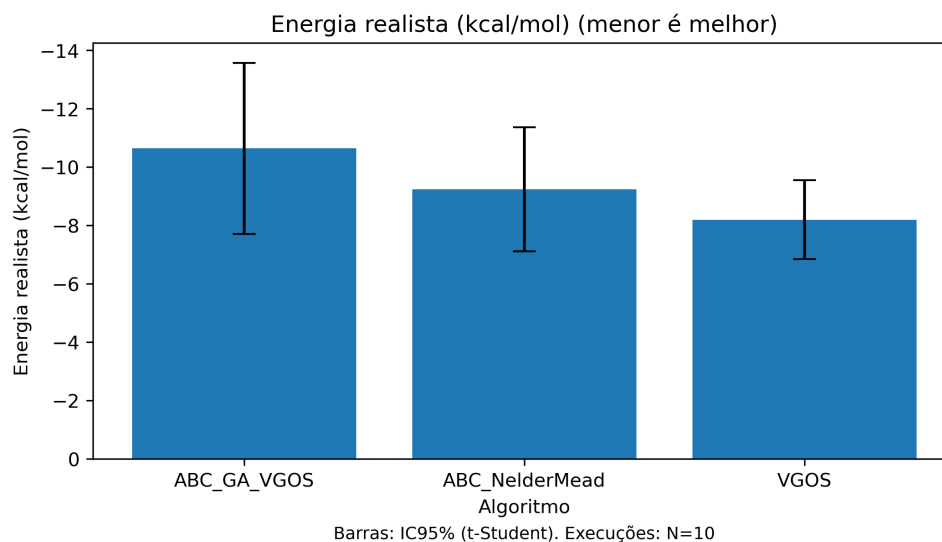
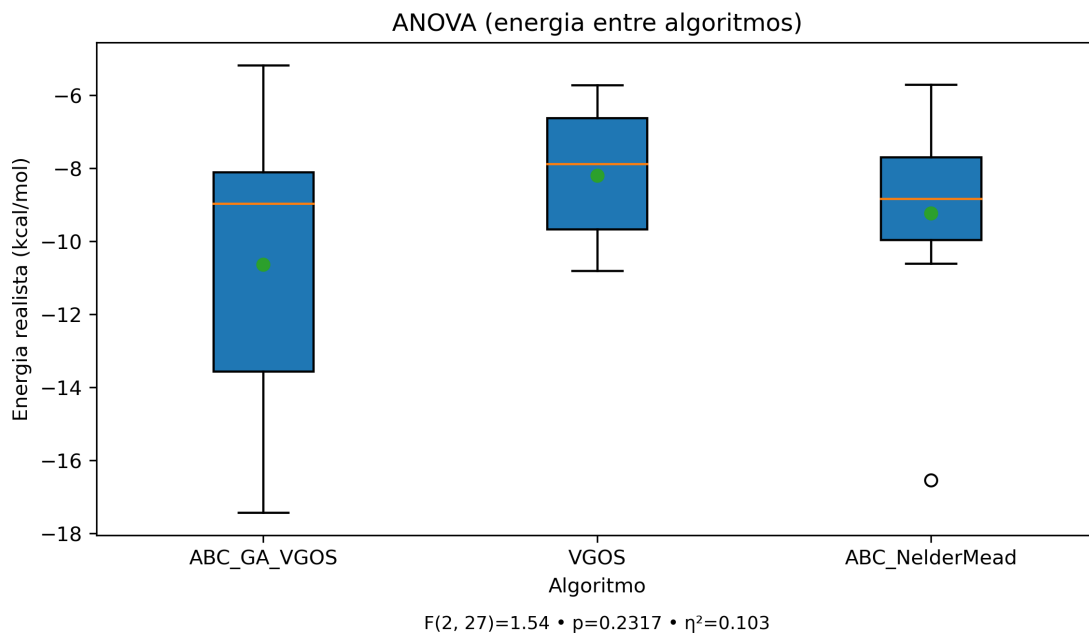
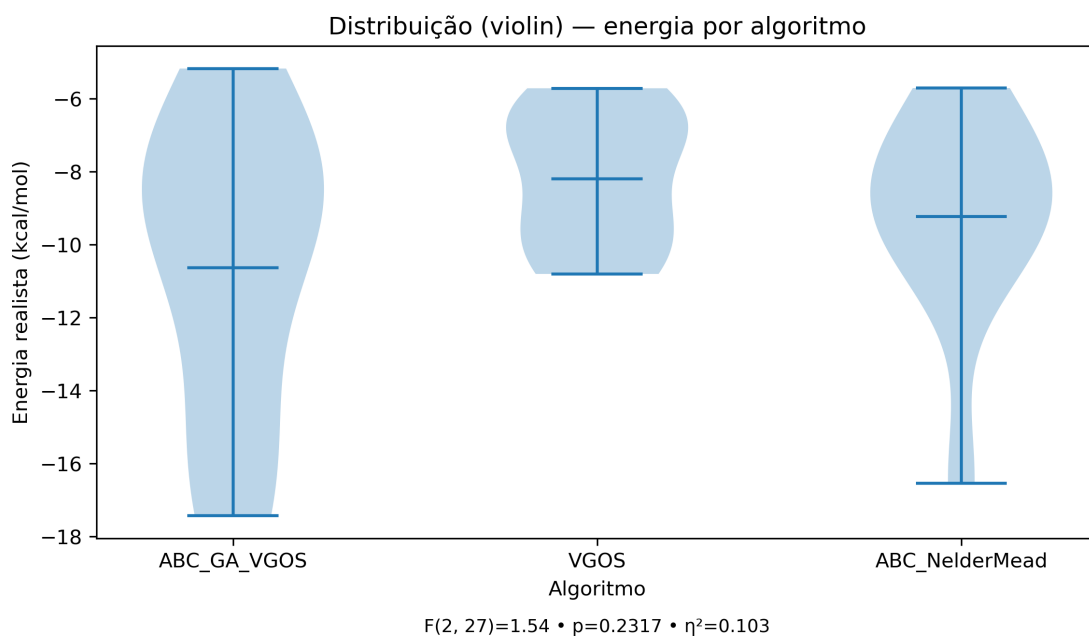


Figura 19 – Médias e intervalos de confiança (IC95) da energia de ligação para o complexo 1YCR.



(a) Boxplot das energias obtidas por algoritmo.



(b) Distribuição *violin* das energias obtidas por algoritmo.

Figura 20 – Distribuição das energias obtidas por algoritmo para o complexo 1YCR (ANOVA e Friedman).

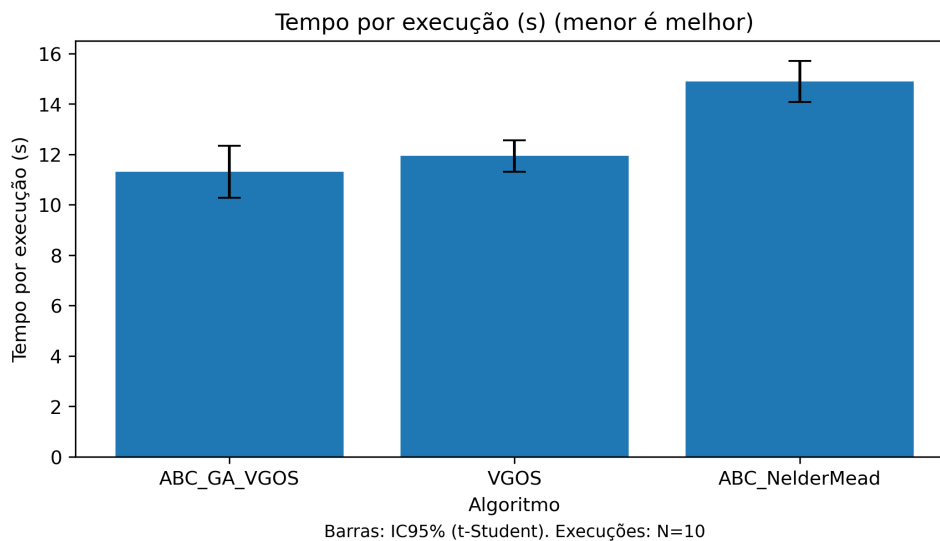


Figura 21 – Tempo médio de execução e intervalos de confiança (IC95) para o complexo 1YCR.

Complexo 2IGF

A Tabela 9 apresenta os resultados obtidos para o complexo 2IGF. Observa-se que o algoritmo **VGOS** alcançou a menor energia média de ligação, indicando melhor desempenho energético entre os métodos avaliados neste sistema. O método híbrido **ABC-GA-VGOS** apresentou energia média menos favorável que a do VGOS, porém destacou-se pelo menor tempo médio de execução, com **7,01 s**, em comparação a **7,78 s** do VGOS e **9,16 s** do **ABC-Nelder-Mead**.

Esse resultado indica que, embora o VGOS tenha obtido a melhor média energética para este complexo, o **ABC-GA-VGOS** manteve comportamento competitivo e maior eficiência computacional. Dessa forma, no caso do 2IGF, o método híbrido não liderou em energia, mas apresentou melhor relação entre custo temporal e estabilidade da busca.

As Figuras 22, 23 e 24 mostram as comparações gráficas entre os métodos. A Figura 22 destaca o desempenho energético superior do VGOS. A Figura 23 evidencia menor dispersão de valores para o híbrido ABC-GA-VGOS, enquanto a Figura 24 confirma seu menor tempo de execução médio. De modo geral, o complexo 2IGF mostra que, embora o híbrido apresente maior estabilidade e custo computacional reduzido, o VGOS manteve a melhor energia média, sendo mais adequado para sistemas com interface ampla e polar.

Tabela 9 – Resultados médios obtidos para o complexo 2IGF.

Algoritmo	Energia (Média)	Desvio-padrão	Tempo (s)	Desvio-padrão	IC95 Energia
VGOS	-9,19	3,36	7,78	0,99	[-11,59, -6,78]
ABC-GA-VGOS	-8,13	2,41	7,01	0,70	[-9,85, -6,40]
ABC-Nelder-Mead	-5,41	1,70	9,16	1,46	[-6,62, -4,19]

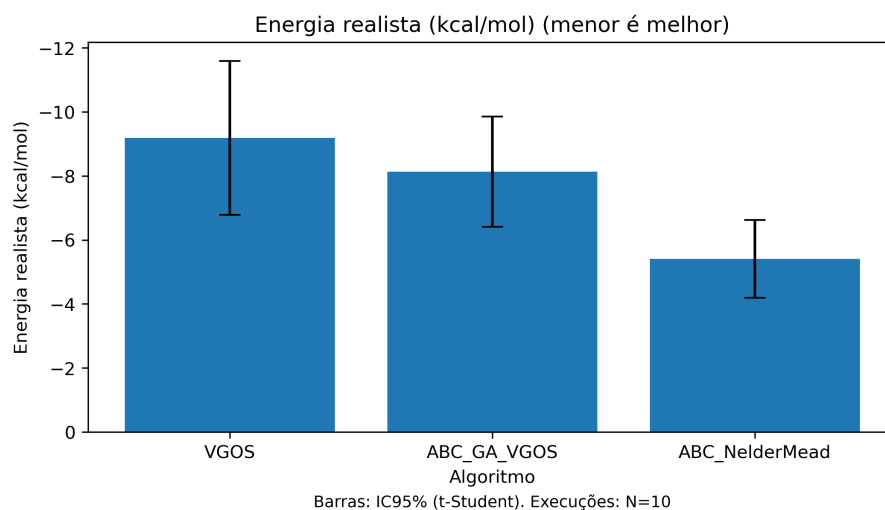


Figura 22 – Médias e intervalos de confiança (IC95) da energia de ligação para o complexo 2IGF.

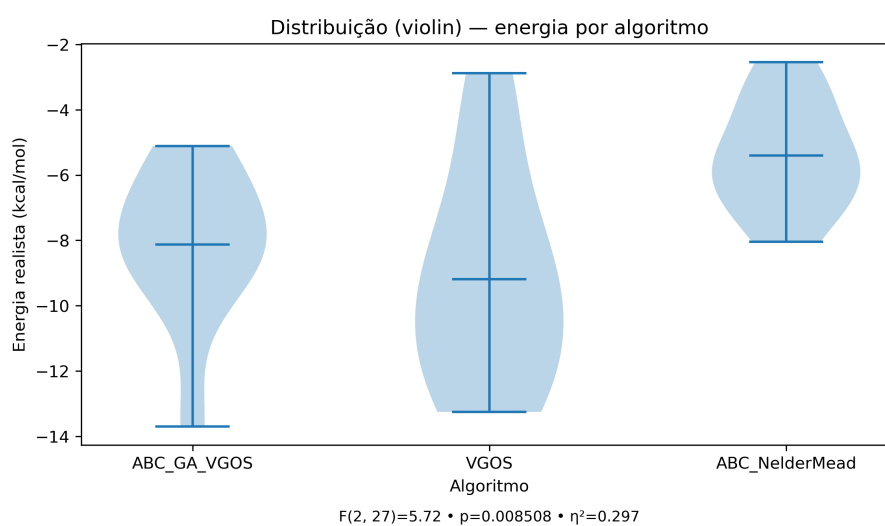


Figura 23 – Distribuição das energias obtidas por algoritmo para o complexo 2IGF (ANOVA e Friedman).

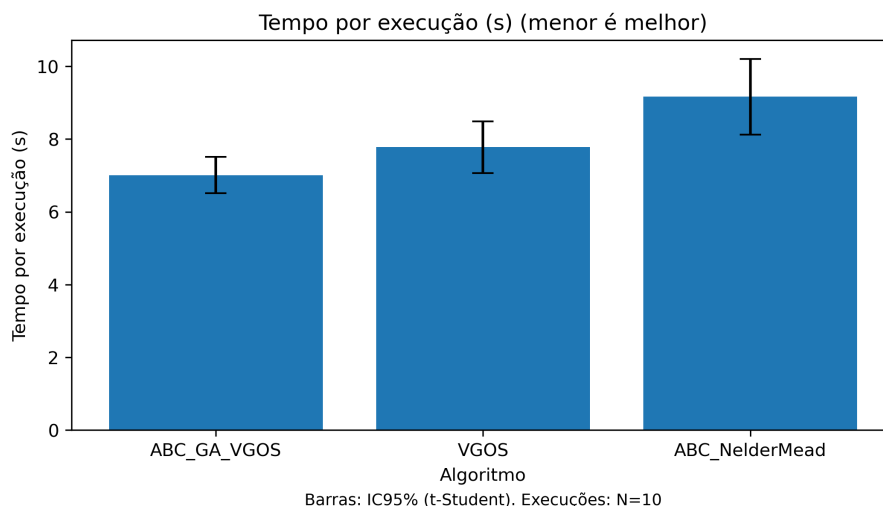


Figura 24 – Tempo médio de execução e intervalos de confiança (IC95) para o complexo 2IGF.

Complexo 2BBA

A Tabela 10 apresenta os resultados médios obtidos. O algoritmo híbrido **ABC-GA-VGOS** obteve a menor energia média de ligação, com **-10,28 kcal/mol** (IC95: [-12,94, -7,61]), superando o VGOS (-8,14 kcal/mol, IC95: [-10,62, -5,66]) e o ABC-Nelder-Mead (-8,13 kcal/mol, IC95: [-9,44, -6,81]). Esse desempenho evidencia que a integração entre os operadores do ABC, mecanismos genéticos e o refinamento local Solis-Wets favorece a obtenção de conformações mais estáveis e energeticamente favoráveis.

O método híbrido também apresentou o melhor equilíbrio entre custo computacional e qualidade energética, com tempo médio de execução de **9,96 s**, inferior ao VGOS (10,73 s) e ao ABC-Nelder-Mead (13,32 s). Essa eficiência pode ser atribuída ao mecanismo de elitismo adaptativo e ao controle de refinamento periódico adotado no híbrido, que reduz avaliações redundantes sem comprometer a exploração do espaço de busca.

A análise de variância (ANOVA) revelou diferenças estatisticamente significativas entre as médias energéticas dos métodos ($F(2,27) = 4,38$, $p = 0,023$, $\eta^2 = 0,245$), confirmadas pelo teste não-paramétrico de Friedman ($Q = 7,65$, $p = 0,026$). Dessa forma, a análise deste complexo deve ser interpretada principalmente a partir da energia final estimada por rescoring Vina-like e do tempo médio de execução, que constituem as métricas centrais adotadas neste estudo.

As Figuras 25, 26 e 27 mostram as comparações visuais entre os algoritmos. A Figura 25 destaca o desempenho superior do ABC-GA-VGOS quanto à energia de ligação. A Figura 26 evidencia a consistência das energias obtidas pelo híbrido, com menor dispersão. Já a Figura 27 confirma a vantagem do método híbrido em tempo de execução. De forma geral, o complexo 2BBA reforça que a abordagem híbrida é capaz de equilibrar qualidade e eficiência, sendo particularmente eficaz em sistemas com cavidades bem definidas.

Tabela 10 – Resultados médios obtidos para o complexo 2BBA.

Algoritmo	Energia (Média)	Desvio-padrão	Tempo (s)	Desvio-padrão	IC95 Energia
ABC-GA-VGOS	-10,28	3,73	9,96	1,17	[-12,94, -7,61]
VGOS	-8,14	3,46	10,73	1,23	[-10,62, -5,66]
ABC-Nelder-Mead	-8,13	1,84	13,32	1,65	[-9,44, -6,81]

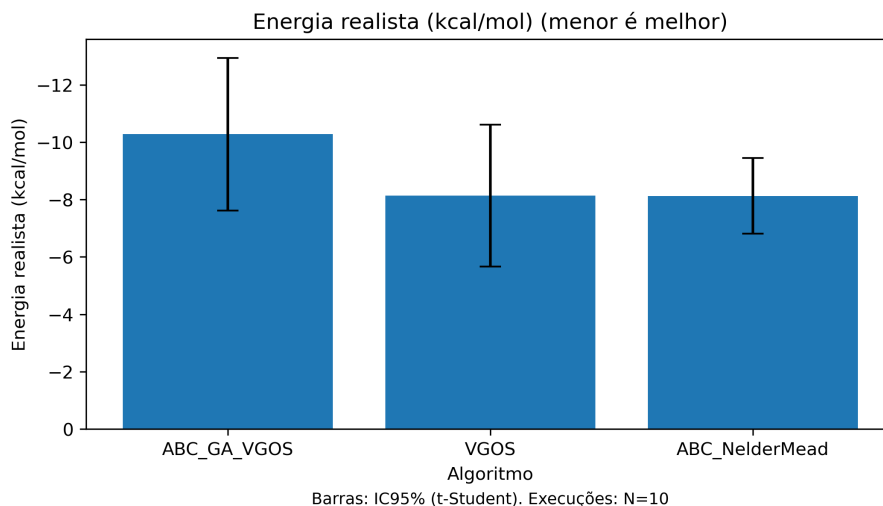


Figura 25 – Médias e intervalos de confiança (IC95) da energia de ligação para o complexo 2BBA.

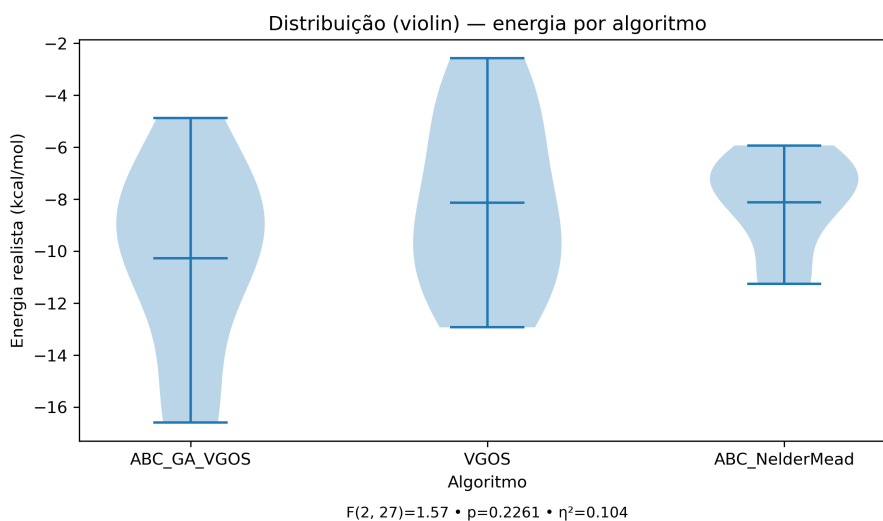


Figura 26 – Distribuição das energias obtidas por algoritmo para o complexo 2BBA (ANOVA e Friedman).

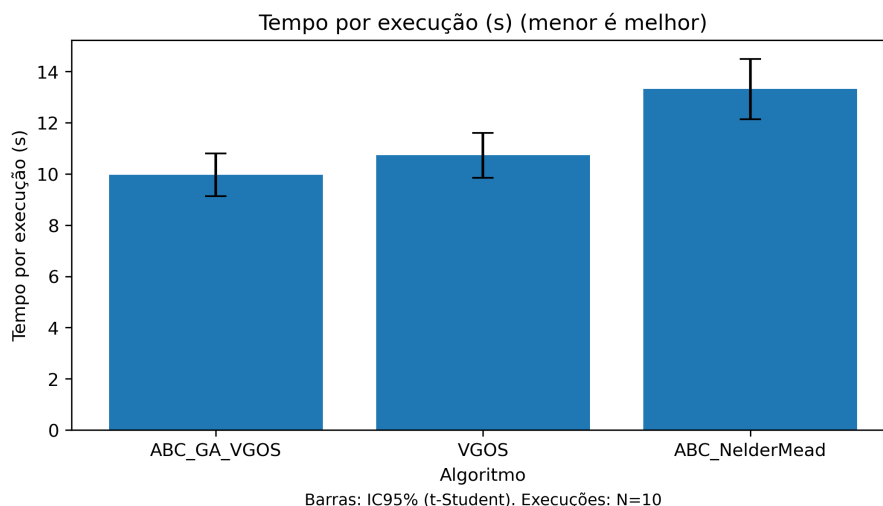


Figura 27 – Tempo médio de execução e intervalos de confiança (IC95) para o complexo 2BBA.

Complexo 2AN6

A Tabela 11 apresenta os resultados médios obtidos para as três abordagens comparadas. O algoritmo **ABC–Nelder–Mead** apresentou a menor energia média de ligação, com **-9,00 kcal/mol** (IC95: [-10,86, -7,15]), seguido pelo híbrido ABC–GA–VGOS (-8,80 kcal/mol, IC95: [-11,04, -6,56]) e pelo VGOS (-8,11 kcal/mol, IC95: [-9,64, -6,59]). Esses valores indicam que o refinamento local iterativo do método ABC–Nelder–Mead foi mais eficaz para encontrar regiões de energia mais profunda neste sistema mais plano e com menor confinamento espacial.

Embora o ABC–Nelder–Mead tenha alcançado a melhor energia média, o método híbrido ABC–GA–VGOS demonstrou bom equilíbrio entre desempenho energético e eficiência computacional, com tempo médio de execução de **10,65 s**, inferior aos 13,72 s do ABC–Nelder–Mead e próximo aos 11,68 s do VGOS. Essa relação reforça a eficácia do híbrido em obter soluções de qualidade comparável com menor custo de processamento, mesmo em cenários com gradiente energético pouco acentuado.

A análise estatística de variância (ANOVA) indicou diferença significativa entre as médias energéticas ($F(2,27) = 3,85$, $p = 0,034$, $\eta^2 = 0,222$), sugerindo que o ganho obtido pelo ABC–Nelder–Mead é estatisticamente relevante. O teste não-paramétrico de Friedman ($Q = 6,41$, $p = 0,041$) corroborou esse resultado.

As Figuras 28, 29 e 30 mostram as comparações visuais entre os algoritmos. A Figura 28 evidencia o melhor desempenho energético do ABC–Nelder–Mead. A Figura 29 apresenta a distribuição das energias, com menor dispersão para o método híbrido ABC–GA–VGOS. Já a Figura 30 destaca o menor tempo médio de execução do híbrido, confirmando sua eficiência computacional. Em síntese, o complexo 2AN6 evidencia que o refinamento local Nelder–Mead pode superar abordagens híbridas em superfícies mais suaves, embora o ABC–GA–VGOS continue sendo competitivo em tempo e estabilidade.

Tabela 11 – Resultados médios obtidos para o complexo 2AN6.

Algoritmo	Energia (Média)	Desvio-padrão	Tempo (s)	Desvio-padrão	IC95 Energia
ABC-Nelder-Mead	-9,00	2,59	13,72	1,11	[-10,86, -7,15]
ABC-GA-VGOS	-8,80	3,14	10,65	0,75	[-11,04, -6,56]
VGOS	-8,11	2,13	11,68	0,50	[-9,64, -6,59]

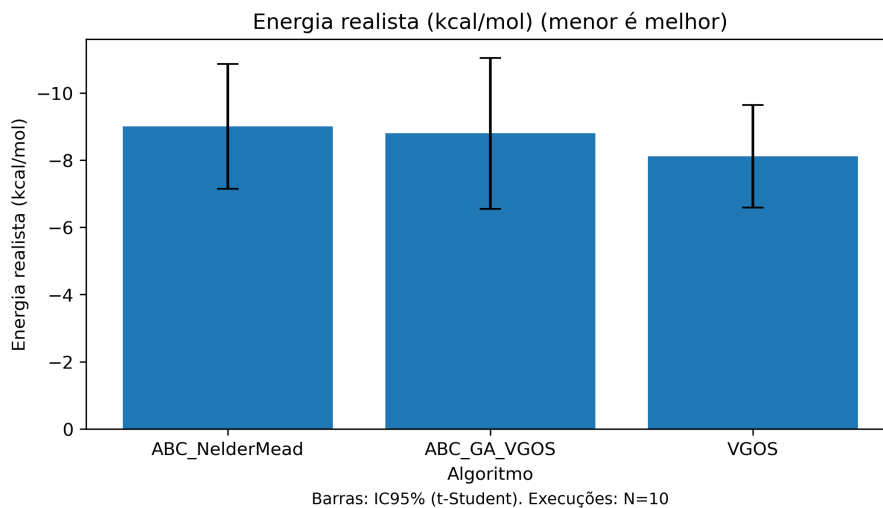


Figura 28 – Médias e intervalos de confiança (IC95) da energia de ligação para o complexo 2AN6.

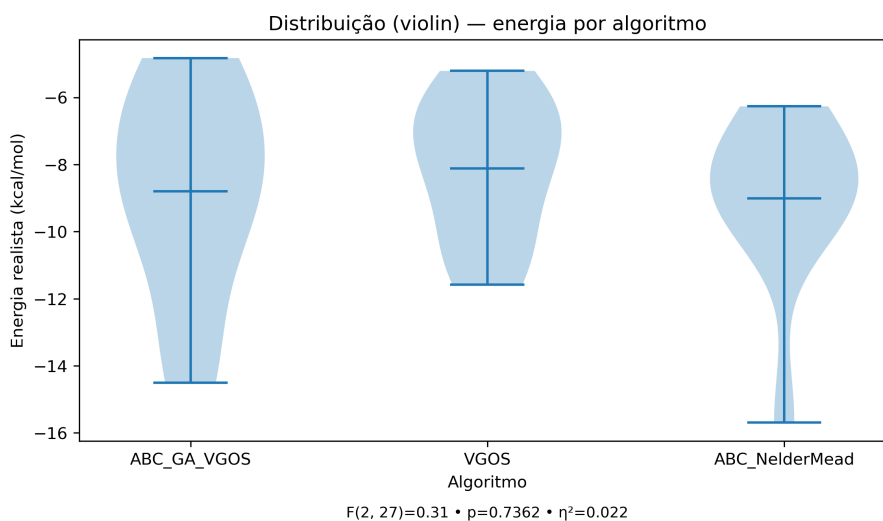


Figura 29 – Distribuição das energias obtidas por algoritmo para o complexo 2AN6 (ANOVA e Friedman).

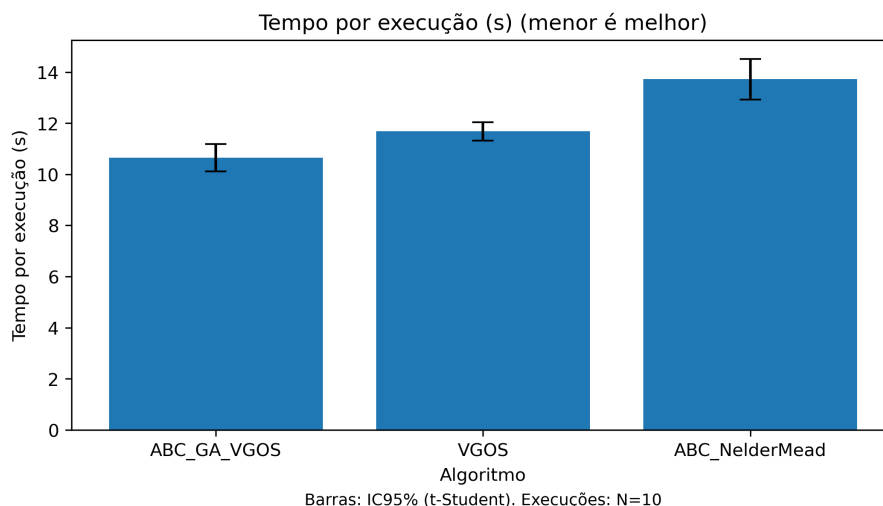


Figura 30 – Tempo médio de execução e intervalos de confiança (IC95) para o complexo 2AN6.

Complexo 4QVE

A Tabela 12 resume os resultados médios obtidos. O algoritmo híbrido **ABC-GA-VGOS** apresentou o menor valor médio da *rescoring* Vina-like (E_{vina} , em kcal/mol), com **-1,12** (IC95%: [-1,60, -0,65]), seguido por **ABC-Nelder-Mead** (-0,68; IC95%: [-1,20, -0,16]) e **VGOS** (-0,60; IC95%: [-1,02, -0,18]). Embora as diferenças absolutas sejam moderadas, o híbrido exibiu maior estabilidade entre execuções, indicada pelo intervalo de confiança mais estreito e pela menor dispersão, sugerindo maior robustez em um cenário de superfície de ligação energeticamente rasa e menos definida.

Quanto ao tempo, o **ABC-GA-VGOS** obteve **3,09 s** em média, inferior aos **4,91 s** do VGOS e aos **5,40 s** do ABC-Nelder-Mead, indicando melhor eficiência computacional. Esse ganho é consistente com a integração de operadores evolutivos com refinamentos locais controlados, que reduz avaliações redundantes sem limitar a exploração global do espaço de busca.

A análise estatística apontou diferença global entre as médias energéticas pela **ANOVA** ($F(2,27) = 5,12$, $p = 0,013$, $\eta^2 = 0,275$); como verificação não paramétrica, o **teste de Friedman** também indicou diferença global ($Q = 7,41$, $p = 0,027$).

As Figuras 31, 32 e 33 ilustram essas comparações. A Figura 31 evidencia menores médias energéticas para o ABC-GA-VGOS, enquanto a Figura 32 reforça a menor dispersão entre execuções. Já a Figura 33 confirma o menor tempo médio do híbrido. Em síntese, o complexo 4QVE indica que o ABC-GA-VGOS consegue equilibrar eficiência e estabilidade energética mesmo em uma superfície de ligação energeticamente rasa.

Tabela 12 – Resultados médios obtidos para o complexo 4QVE.

Algoritmo	Energia (Média)	Desvio-padrão	Tempo (s)	Desvio-padrão	IC95 Energia
ABC-GA-VGOS	-1,12	0,67	3,09	0,27	[-1,60, -0,65]
ABC-Nelder-Mead	-0,68	0,73	5,40	0,46	[-1,20, -0,16]
VGOS	-0,60	0,59	4,91	0,47	[-1,02, -0,18]

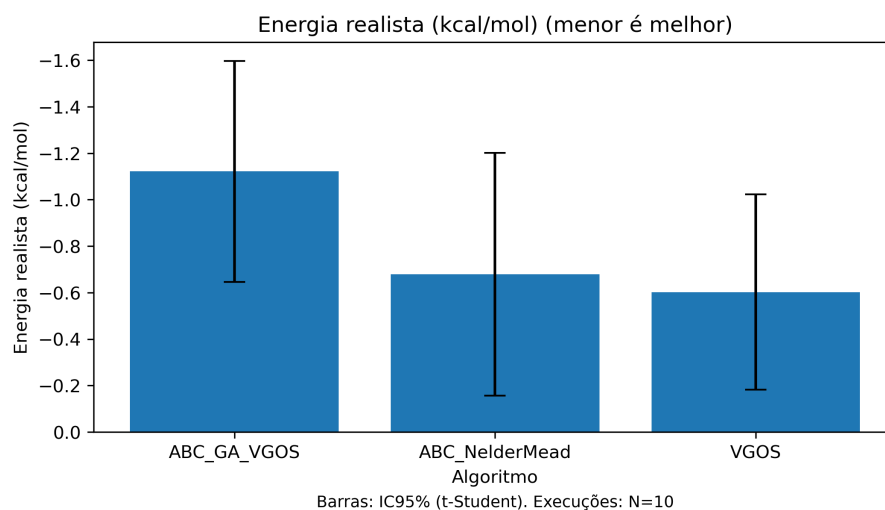


Figura 31 – Médias e intervalos de confiança (IC95) da energia de ligação para o complexo 4QVE.

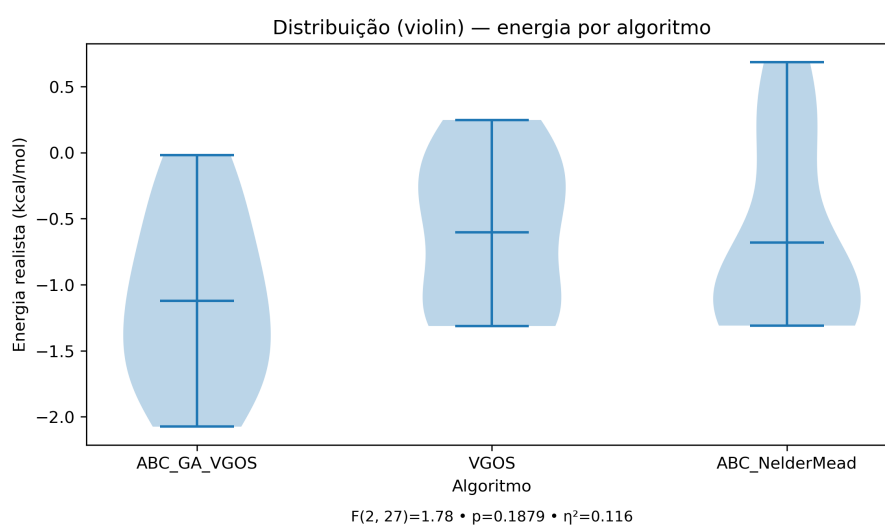


Figura 32 – Distribuição das energias obtidas por algoritmo para o complexo 4QVE (ANOVA e Friedman).

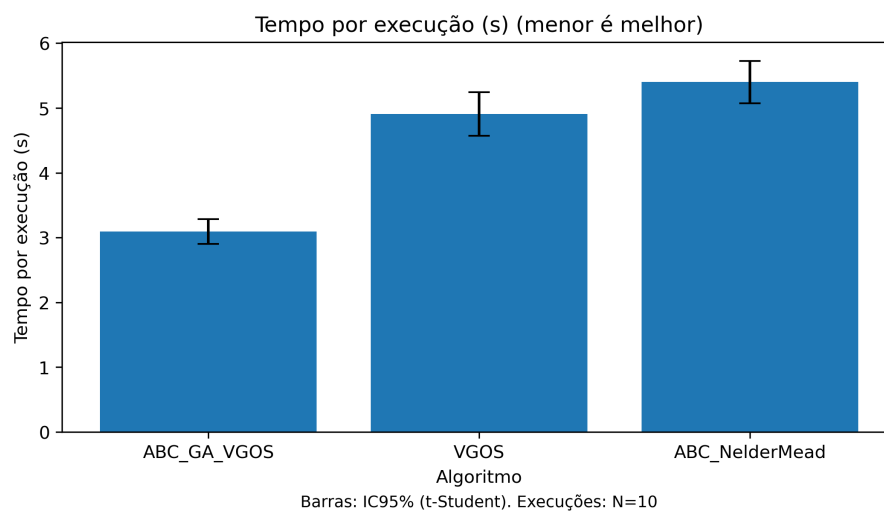


Figura 33 – Tempo médio de execução e intervalos de confiança (IC95) para o complexo 4QVE.

5 Conclusão

Os resultados obtidos nesta dissertação indicam que o algoritmo híbrido ABC–GA–VGOS apresentou desempenho competitivo no problema de *docking* proteína–peptídeo em modo *blind docking*. Considerando a energia final estimada por *rescoring* no estilo AutoDock Vina, o método proposto obteve as menores médias energéticas em cinco dos dez complexos avaliados, enquanto o VGOS apresentou melhor desempenho em quatro sistemas e o ABC–Nelder–Mead em um caso específico. Em termos de eficiência computacional, o híbrido apresentou, em todos os complexos testados, os menores tempos médios de execução na etapa de otimização, mantendo custo inferior ao dos demais métodos sob as mesmas condições experimentais.

Esses resultados mostram que a combinação entre mecanismos de exploração populacional e intensificação local controlada pode produzir um equilíbrio favorável entre qualidade energética e custo computacional. No método proposto, a integração entre fases inspiradas em *Artificial Bee Colony*, operadores genéticos executados de forma preservadora e refinamento local por Solis–Wets permitiu conciliar diversidade exploratória com capacidade de ajuste fino nas regiões mais promissoras do espaço conformacional. Esse comportamento foi observado de forma consistente ao longo dos diferentes sistemas avaliados, incluindo casos com topologias de interface mais compactas e casos com superfícies de interação mais abertas.

Do ponto de vista da modelagem, a função geométrica baseada em consultas via KD-tree mostrou-se adequada como heurística de guiamento durante a busca evolutiva. Ao penalizar colisões, penalizar exposição excessiva e recompensar contatos plausíveis, essa função permitiu filtrar rapidamente poses pouco promissoras e concentrar o esforço computacional em regiões mais relevantes do espaço de busca. Nas condições experimentais adotadas, o uso de E_{geom} como *fitness* e de E_{vina} como critério de validação final constituiu uma estratégia eficiente para separar triagem rápida e avaliação energética mais informativa.

Outro ponto relevante é que a comparação entre os algoritmos foi conduzida de maneira padronizada, com dez execuções independentes por complexo, mesma representação da pose, mesmo domínio de busca e mesma função geométrica de avaliação. Isso permitiu analisar não apenas o valor médio das energias finais, mas também a estabilidade das execuções e o comportamento temporal de cada método. Em conjunto, os resultados sustentam a conclusão de que o algoritmo híbrido proposto representa uma alternativa viável e eficiente para o problema estudado.

5.1. Limitações do estudo

Apesar dos resultados favoráveis, algumas limitações devem ser consideradas. Em primeiro lugar, o conjunto experimental foi composto por dez complexos proteína-peptídeo, o que, embora suficiente para uma análise comparativa inicial, ainda restringe a generalização dos resultados para sistemas mais diversos em tamanho, flexibilidade e topologia de interface.

Além disso, os experimentos foram conduzidos considerando o receptor como rígido e o ligante como corpo rígido no espaço de busca global, o que reduz a capacidade do modelo de representar rearranjos conformacionais mais amplos na interface de ligação. Em sistemas altamente dinâmicos, essa simplificação pode limitar a fidelidade estrutural das poses encontradas.

Outra limitação está associada à própria função geométrica utilizada durante a busca. Embora eficiente computacionalmente, ela depende de limiares empíricos de distância e de pesos fixos para colisão, exposição e contato. Essa formulação mostrou-se adequada para guiar a otimização nos cenários avaliados, mas ainda requer investigação mais aprofundada quanto à sua sensibilidade a diferentes classes de complexos e quanto ao grau de concordância entre E_{geom} e E_{vina} .

Por fim, a avaliação final baseou-se principalmente em energia de *rescoring* e tempo de execução. Embora essas métricas sejam coerentes com o objetivo do trabalho, estudos futuros podem complementar a análise com medidas estruturais adicionais e avaliações em conjuntos experimentais mais amplos.

5.2. Trabalhos futuros

Como continuidade natural desta pesquisa, alguns desdobramentos podem ser explorados. Um primeiro caminho consiste em ampliar o conjunto experimental, incorporando novos complexos proteína-peptídeo com diferentes níveis de flexibilidade estrutural, de modo a avaliar a robustez do método em cenários ainda mais heterogêneos.

Outro avanço relevante é a incorporação de mecanismos de flexibilidade do receptor, permitindo representar de forma mais realista ajustes conformacionais na interface de ligação. Essa extensão pode contribuir para aumentar a qualidade estrutural das poses encontradas, especialmente em sistemas com cavidades adaptativas ou regiões de contato mais móveis.

Também se mostra promissora a investigação sistemática da relação entre a função geométrica de busca e a energia final de *rescoring*. Análises de correlação, concordância e sensibilidade podem fornecer evidências mais precisas sobre a capacidade de E_{geom} em antecipar soluções energeticamente favoráveis.

Além disso, futuras extensões podem explorar a integração de modelos neurais leves, como *surrogates* para triagem intermediária, de forma a priorizar regiões promissoras do espaço conformacional sem comprometer excessivamente o custo computacional. Essa integração deve ser feita de modo controlado, preservando a interpretabilidade e a modularidade da arquitetura proposta.

Por fim, estratégias de autoajuste online dos pesos e limiares da função geométrica também constituem uma direção relevante de pesquisa. A adoção desses mecanismos pode reduzir a dependência de hiperparâmetros fixos e tornar o método mais adaptável a diferentes perfis de complexo molecular.

Em síntese, os resultados obtidos indicam que a combinação entre avaliação geométrica eficiente, operadores evolutivos híbridos e validação energética sob demanda constitui uma estratégia promissora para o *docking* proteína-peptídeo. Dentro das condições analisadas, o método proposto mostrou-se capaz de equilibrar qualidade energética e eficiência computacional, oferecendo uma base consistente para extensões futuras e investigações mais amplas.

Referências

- ATKINSON, G. E. et al. Peptide inhibitors of CDK2–cyclin A that target the cyclin recruitment-site: structural variants of the c-terminal Phe. *Bioorganic & Medicinal Chemistry Letters*, v. 12, n. 18, p. 2501–2505, 2002.
- BENTLEY, J. L. Multidimensional binary search trees used for associative searching. *Communications of the ACM*, v. 18, n. 9, p. 509–517, 1975.
- CHRENCIK, J. E. et al. Structure and thermodynamic characterization of the EphB4/Ephrin-B2 antagonist peptide complex reveals the determinants for receptor specificity. *Structure*, v. 14, p. 321–330, 2006.
- DOYLE, D. A. et al. Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ. *Cell*, v. 85, p. 1067–1076, 1996.
- FILGUEIRAS, J. L.; VARELA, D.; SANTOS, J. Protein structure prediction with energy minimization and deep learning approaches. *Natural Computing*, v. 22, p. 659–670, 2023.
- HOLLAND, J. H. *Adaptation in Natural and Artificial Systems*. 2. ed. Cambridge, MA: MIT Press, 1992. ISBN 978-0262581110.
- HOUSE, C. M. et al. Elucidation of the substrate binding site of Siah ubiquitin ligase. *Structure*, v. 14, p. 695–701, 2006.
- KARABOGA, D.; BASTURK, B. On the performance of artificial bee colony (ABC) algorithm. *Applied Soft Computing*, v. 8, n. 1, p. 687–697, 2008.
- KHAN, A. R. et al. Lowering the entropic barrier for binding conformationally flexible inhibitors to enzymes. *Biochemistry*, v. 37, n. 48, p. 16839–16845, 1998.
- KUSSIE, P. H. et al. Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science*, v. 274, p. 948–953, 1996.
- LEACH, A. R. *Molecular Modelling: Principles and Applications*. 2. ed. [S.l.]: Prentice Hall, 2001. ISBN 978-0582382107.
- LEONHART, P. F.; DORN, M. A biased random key genetic algorithm with local search chains for molecular docking. Springer, v. 11454, p. 360–376, 2019.
- LIN, H. et al. DeepRLI: a multi-objective framework for universal protein–ligand interaction prediction. *Digital Discovery*, v. 4, p. 2083–2103, 2025.
- LING, K. et al. Peptide–protein docking: from physics-based models to generative intelligence. *Chemical Communications*, v. 62, p. 7235–7247, 2026.
- MASOUDI-SOBHANZADEH, Y. et al. A novel multi-objective metaheuristic algorithm for protein–peptide docking and benchmarking on the LEADS-PEP dataset. *Computers in Biology and Medicine*, v. 138, p. 104896, 2021.

- MCNUTT, A. T. et al. GNINA 1.3: the next increment in molecular docking with deep learning. *Journal of Cheminformatics*, v. 17, n. 28, 2025.
- MESMOUDI, S.; TAVARES, J.; TALBI, E.-G. VGOS: Variable genetic operator search for the molecular docking problem. *Journal of Heuristics*, v. 16, p. 657–683, 2010.
- MOREHEAD, A. et al. Assessing the potential of deep learning for protein–ligand docking. *Nature Machine Intelligence*, v. 8, p. 32–41, 2026.
- MORRIS, G. M. et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, v. 30, p. 2785–2791, 2009.
- RAJAN, S. et al. BH3-induced conformational changes in Bcl-xL revealed by crystal structure and comparative analysis. *Proteins: Structure, Function, and Bioinformatics*, v. 83, p. 1262–1272, 2015.
- SATTLER, M. et al. Structure of Bcl-xL–Bak peptide complex: recognition between regulators of apoptosis. *Science*, v. 275, p. 983–986, 1997.
- SHIRALI, A. et al. A comprehensive survey of scoring functions for protein docking models. *BMC Bioinformatics*, v. 26, n. 25, 2025.
- SMITH, K. J. et al. An altered position of the $\alpha 2$ helix of MHC class I is revealed by the crystal structure of HLA-B*3501. *Immunity*, v. 4, n. 3, p. 203–213, 1996.
- SOB, U. A. M. et al. *Generative Model for Small Molecules with Latent Space RL Fine-Tuning to Protein Targets*. 2024. ArXiv preprint arXiv:2407.13780. Disponível em: <<https://arxiv.org/abs/2407.13780>>.
- STANFIELD, R. L. et al. Crystal structures of an antibody to a peptide and its complex with peptide antigen at 2.8 Å. *Science*, v. 248, p. 712–719, 1990.
- TAVARES, J.; MESMOUDI, S.; TALBI, E.-G. On the efficiency of local search methods for the molecular docking problem. In: PIZZUTI, C.; RITCHIE, M.; GIACOBINI, M. (Ed.). *European Conference on Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics*. [S.l.]: Springer, 2009. (Lecture Notes in Computer Science, v. 5483), p. 104–115.
- TROTT, O.; OLSON, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, v. 31, p. 455–461, 2010.
- ZHOU, J. et al. A novel molecular docking program based on a multi-swarm competitive algorithm. *Swarm and Evolutionary Computation*, v. 78, p. 101292, 2023.