

BIOLOGIA PREDITIVA: A REVOLUÇÃO IMPULSIONADA PELA INTELIGÊNCIA ARTIFICIAL <https://doi.org/10.63330/aurumpub.015-010>**Abraham Guerra**

Universidade Estadual Paulista "Júlio de Mesquita Filho - Faculdade de Ciências Agrárias e Veterinárias
Via de Acesso Professor Paulo Donato Castelane Castellane S/N - Vila Industrial, Jaboticabal, Sao Paulo

Nilo Ricardo Corrêa de Mello Júnior

Universidade Estadual Paulista "Júlio de Mesquita Filho - Faculdade de Ciências Agrárias e Veterinárias
Via de Acesso Professor Paulo Donato Castelane Castellane S/N - Vila Industrial, Jaboticabal, Sao Paulo

Anderson Manares-Romero

Universidade Estadual Paulista "Júlio de Mesquita Filho - Faculdade de Ciências Agrárias e Veterinárias
Via de Acesso Professor Paulo Donato Castelane Castellane S/N - Vila Industrial, Jaboticabal, Sao Paulo

Luzia Micaele Alves Barbosa

Universidade do Estado da Bahia
Rua Castro Alves, 291, Juazeiro, Bahia

RESUMO

A convergência da inteligência artificial e da bioinformática iniciou uma revolução científica, transformando a nossa capacidade de interpretar a complexidade biológica. Este trabalho argumenta que a recente solução para o problema do enovelamento de proteínas, alcançada por ferramentas de IA como o AlphaFold, proporcionou um arsenal tecnológico sem precedentes para enfrentar crises de saúde globais como a resistência aos antimicrobianos (RAM). Apresenta-se a "bioprospeção digital" como um novo paradigma que, em vez de cultivar microrganismos, explora diretamente os dados genômicos e proteômicos. Este método permite agora investigar fronteiras biológicas antes inacessíveis, destacando o domínio das Arqueias como um alvo de especial interesse. Devido à sua bioquímica única e por ser um ramo da vida vastamente inexplorado, a recente análise computacional dos seus proteomas revelou um rico potencial para novas famílias de compostos antimicrobianos ("arqueasinas"), validando assim a sua importância como uma fonte crucial para a descoberta de futuros antibióticos. Assim, a transição de uma ciência descritiva para uma preditiva, impulsionada por esta sinergia, inaugura uma nova era no desenho de fármacos e na engenharia biológica.

Palavras-chave: Machine learning; Bioinformática; AlphaFold; Resistência aos antimicrobianos; Arqueias.



1 INTRODUÇÃO

Na confluência da biologia e da computação, está a emergir uma revolução científica de uma escala sem precedentes. Esta transformação baseia-se na convergência de duas linguagens fundamentais: o código genético, a sequência de nucleótidos que dita a forma e a função de todo o ser vivo, e o código computacional da inteligência artificial (IA), que nos concede a capacidade de processar e compreender a complexidade biológica a uma velocidade e profundidade antes inimagináveis. A fusão destes dois códigos não representa uma melhoria incremental, mas sim uma mudança de paradigma que está a redefinir os limites da descoberta. Permite-nos, pela primeira vez, decifrar a "gramática" oculta das proteínas, as macromoléculas que executam as instruções da vida (Jumper et al., 2021).

Esta nova era é definida pela nossa capacidade de enfrentar desafios que durante muito tempo foram considerados intratáveis. O primeiro deles é o "grande desafio" do enovelamento de proteínas, um enigma que tem ocupado a comunidade científica por mais de 50 anos (Anfinsen, 1973; Jumper et al., 2021a). Prever a intrincada estrutura tridimensional de uma proteína a partir da sua sequência linear de aminoácidos era uma tarefa de uma complexidade computacional astronómica, mas fundamental, uma vez que a estrutura determina a função. O segundo desafio é uma crise de saúde pública de carácter urgente e global: a resistência aos antimicrobianos (RAM). Declarada pela Organização Mundial da Saúde (OMS) como uma das dez principais ameaças à humanidade, a RAM ameaça reverter um século de avanços médicos, fazendo com que infeções comuns voltem a ser mortais devido à estagnação na descoberta de novos antibióticos (Murray et al., 2022a).

O argumento central deste trabalho é que a recente solução para o primeiro desafio, impulsionada pela IA através de ferramentas como o AlphaFold, nos proporcionou um arsenal conceptual e tecnológico completamente novo para enfrentar o segundo. Ao decifrar o código estrutural das proteínas, abrimos caminhos inovadores para descobrir e desenhar os fármacos do futuro, iniciando uma nova era em que o código da vida já não é apenas lido, mas começa a ser compreendido e, potencialmente, reescrito.

2 BIOINFORMÁTICA: A LINGUAGEM DA BIOLOGIA NA ERA DIGITAL

2.1 DEFINIÇÃO E FUNDAMENTOS

A bioinformática é a disciplina científica que se situa na interseção da biologia, da informática, da matemática e da estatística (Carolina Cabral da Silva; Cidinaria Silva Alves, 2024; Uesaka et al., 2022). O seu propósito fundamental é desenvolver métodos e ferramentas para armazenar, analisar e, de forma crucial, interpretar a vasta quantidade de dados biológicos que as tecnologias modernas geram (Carolina Cabral da Silva; Cidinaria Silva Alves, 2024; Diniz; Canduri, 2017; Uesaka et al., 2022). Em essência, a bioinformática atua como a "parte linguística da genética". Assim como um linguista estuda os padrões da linguagem para compreender o seu significado e estrutura, um bioinformático analisa os padrões nas



sequências de ADN e proteínas para identificar genes, determinar as suas funções, estabelecer relações evolutivas e prever a sua conformação tridimensional. Esta disciplina é a ponte indispensável que liga os dados biológicos brutos, como os milhares de milhões de bases de um genoma sequenciado, ao conhecimento biológico aplicável e à compreensão funcional (Uesaka et al., 2022).

A necessidade da bioinformática surgiu de uma realidade inescapável: os avanços nas tecnologias, como as tecnologias de sequenciação de nova geração (NGS), produziram millones de dados biológicos (Rabbani; Tekin; Mahdih, 2014). Tal volume de informação excedia a capacidade dos métodos de análise tradicionais. Sem um arcabouço computacional para organizar, consultar e analisar esses dados, o Projeto Genoma Humano, por exemplo, teria resultado numa coleção ininteligível de letras em vez de um mapa funcional da nossa espécie. A bioinformática forneceu esse enquadramento, combinando bases de dados biológicas massivas com algoritmos sofisticados e modelos estatísticos (Clark; Lillard, 2024; Sayers et al., 2021). Portanto, o nascimento da "biologia de *big data*" não foi um produto exclusivo das tecnologias de sequenciação, mas o resultado da simbiose entre a geração de dados em grande escala e a capacidade da bioinformática para transformar esses dados em conhecimento estruturado.

2.2 O IMPACTO TRANSFORMADOR NA CIÊNCIA E NA MEDICINA

O catalisador que demonstrou o poder da bioinformática à escala mundial foi a conclusão do Projeto Genoma Humano em 2004 (Hood; Rowen, 2013). Este marco não foi apenas um triunfo da biologia molecular, mas também uma proeza bioinformática que tornou manejável uma tarefa de análise de dados de enorme escala e complexidade. Desde então, as suas aplicações permeiam quase todos os campos das ciências da vida, gerando um impacto transformador.

Entre as suas aplicações mais destacadas encontram-se:

- **Desenvolvimento de fármacos:** As ferramentas bioinformáticas aceleram drasticamente a descoberta de novos medicamentos. Permitem a identificação de alvos terapêuticos a nível molecular e a realização de simulações computacionais para prever como um fármaco potencial interagirá com o seu alvo, reduzindo custos e tempo nas fases iniciais de investigação (Behzadi; Gajdács, 2022; Bian; Xie, 2018; Xia, 2017; Zhang et al., 2025).
- **Saúde pública e epidemiologia:** O papel da bioinformática foi crucial durante a pandemia de COVID-19. A análise computacional rápida da sequência genómica do vírus SARS-CoV-2 permitiu decifrar a estrutura molecular das suas proteínas, um passo essencial para o desenho e desenvolvimento de vacinas eficazes em tempo recorde (Abdelsattar et al., 2021; Al-Janabi, 2022; Ma et al., 2021; Sawyer; Free; Martin, 2021; Torrington, 2022). Além disso, utilizam-se modelos bioinformáticos para prever a evolução e propagação de epidemias.
- **Indústria, biotecnologia e agricultura:** No setor industrial, a bioinformática é crucial para a

"mineração de genomas", que consiste em explorar os genomas de microrganismos em busca de genes que produzam enzimas de interesse. Estes biocatalisadores são otimizados e utilizados para desenvolver processos industriais mais eficientes e ecológicos no fabrico de detergentes, biocombustíveis e alimentos (Beller et al., 2018; Costessi et al., 2018; Mitra et al., 2022; Van den Bogert et al., 2019).

De forma semelhante, no setor agrícola, a integração entre bioinformática e tecnologias não destrutivas tem aberto novas fronteiras. Métodos como a espectroscopia no infravermelho próximo (NIR) vêm se mostrando promissores para avaliação de qualidade de frutas, permitindo a análise simultânea de vários atributos com mínima preparação da amostra. Essa técnica já demonstrou eficácia em culturas como mangas (Da Silva Alves et al., 2025; De Freitas et al., 2022), abacaxi (Amuah et al., 2019), abacate (Subedi; Walsh, 2020), laranja (Borba et al., 2021), nectarina (Scalisi; O'Connell, 2021), banana (Sripaurya et al., 2021), melão (Kim et al., 2021), maçã (Pourdarbani et al., 2022), pera (Lu et al., 2022), tangerina (Huang et al., 2022) e kiwi (Basile; Marsico; Perniola, 2022). Ao combinar esses grandes volumes de dados espectrais com algoritmos bioinformáticos, torna-se possível não apenas predizer atributos de qualidade, mas também identificar padrões ligados à nutrição, ao amadurecimento e até a distúrbios fisiológicos, aproximando a agricultura de uma abordagem cada vez mais data-driven.

3 INTELIGÊNCIA ARTIFICIAL: O MOTOR COMPUTACIONAL DA DESCOBERTA BIOLÓGICA

No contexto das ciências da vida, a inteligência artificial não é uma consciência abstrata, mas um conjunto de ferramentas computacionais avançadas desenhadas para analisar dados biológicos complexos, modelar processos naturais e, fundamentalmente, fazer previsões baseadas em padrões que são frequentemente indetectáveis para a análise humana. A IA permite aos investigadores decodificar vastos conjuntos de dados genômicos, simular comportamentos moleculares e modelar interações ecológicas a uma escala e velocidade sem precedentes (Hamid Jamialahmadi et al., 2024).

Para compreender a sua aplicação, é útil distinguir dois conceitos-chave:

Aprendizagem de máquina (Machine Learning): É um ramo da IA onde os algoritmos são treinados com grandes quantidades de dados para "aprender" a realizar tarefas específicas sem serem programados explicitamente para cada regra (Janiesch; Zschech; Heinrich, 2021). Por exemplo, um modelo de machine learning pode ser alimentado com milhares de sequências de mutações genéticas, algumas rotuladas como benignas e outras como patogênicas. Com o tempo, o algoritmo aprende a identificar os padrões distintivos de cada classe e pode prever a natureza de uma nova mutação que nunca viu antes (Diaz et al., 2023; Zhu; Ong; Huttley, 2020)

Redes neurais e aprendizagem profunda (Deep Learning): O deep learning é um subcampo do machine learning que utiliza arquiteturas chamadas redes neurais artificiais, que se inspiram vagamente na estrutura do cérebro humano ^{40,41}. Estas redes consistem em múltiplas camadas de "neurónios" interligados (daí o termo "profundo"), onde cada camada aprende a reconhecer características cada vez mais complexas e abstratas dos dados ⁴⁰. Esta estrutura hierárquica é particularmente eficaz para analisar os dados biológicos, que são massivos, de alta dimensionalidade e muitas vezes não estruturados, como as sequências genômicas ou as imagens microscópicas. Um diferencial importante do *deep learning* é a sua robustez em lidar com dados massivos e não estruturados. Alharbi et al. (2022) apontam que “*deep learning methods have shown unprecedented performance in genomics, especially when analyzing large-scale, high-dimensional data sets*” ⁴². Essa característica tem permitido avanços expressivos em bioinformática, como na análise de sequências genômicas, no reconhecimento de padrões em imagens biomédicas e na predição de estruturas de proteínas.

Além disso, as redes neurais profundas têm alcançado desempenhos comparáveis ou mesmo superiores ao de especialistas humanos. Como descrevem Goodfellow et al. (2016), “*deep learning has drastically improved the state of the art in speech recognition, visual object recognition, object detection, and many other domains*”, e essa revolução tecnológica se estende agora à biologia preditiva (Goodfellow; Bengio; Courville, [S.d.]).

A verdadeira revolução da IA na biologia não reside apenas na automatização ou no aumento da velocidade de análise, mas numa mudança fundamental de paradigma: de uma ciência predominantemente descritiva para uma ciência preditiva. A biologia tradicional e a bioinformática clássica centraram-se em descrever e catalogar o mundo natural: sequenciar um genoma, determinar uma estrutura proteica, anotar a função de um gene. A IA, e em particular o deep learning, sobressai na previsão. Não se limita a descrever o que existe, mas prevê o que poderia ser ou o que acontecerá. Exemplos como a previsão da estrutura tridimensional de uma proteína a partir da sua sequência linear, a previsão da patogenicidade de uma variante genética ou a previsão da eficácia de um composto farmacológico ilustram esta mudança (Hsu; Lu; Hsu, 2024; Jumper et al., 2021; Zhou; Astore; Skolnick, 2022). Este poder preditivo permite que as experiências de laboratório sejam guiadas por hipóteses geradas computacionalmente com uma alta probabilidade de sucesso, otimizando drasticamente o uso de tempo e recursos na investigação.

3.1 APLICAÇÕES FUNDAMENTAIS DA IA EM BIOLOGIA

As capacidades preditivas da IA já estão a ser aproveitadas numa vasta gama de aplicações biológicas e médicas:

- **Genómica:** Ferramentas como o DeepVariant da Google utilizam redes neurais profundas para analisar milhões de sequências genómicas e identificar variantes genéticas com uma

precisão superior aos métodos anteriores, o que é crucial para diagnosticar doenças genéticas (Poplin et al., 2018).

- **Descoberta de fármacos:** A IA está a transformar o longo e dispendioso processo de desenvolvimento de medicamentos. Os algoritmos podem analisar vastas bibliotecas químicas para identificar moléculas candidatas com potencial terapêutico, prever as suas propriedades farmacológicas e até otimizar a sua estrutura para melhorar a sua eficácia e segurança (Blanco-González et al., 2023; Rehman et al., 2025).
- **Biologia de sistemas:** A IA permite modelar as complexas redes de interação entre genes, proteínas e metabolitos que governam a célula (Dasgupta; De, 2023; Kannan et al., 2025; Kitano, 2012). Estes modelos ajudam os cientistas a compreender como pequenas perturbações, como uma mutação ou a ação de um fármaco, podem afetar o sistema biológico no seu todo.
- **Diagnóstico médico:** Os algoritmos de deep learning, especialmente os baseados em visão por computador, estão a demonstrar uma notável capacidade para analisar imagens médicas. Podem, por exemplo, identificar sinais precoces de melanoma em fotografias de lesões cutâneas, detetar anomalias em radiografias ou prever o resultado da coloração de Gram diretamente a partir de imagens de microscopia, automatizando e acelerando assim a identificação bacteriana (McMahon et al., 2025; Seven et al., 2025).

3.2 A SOLUÇÃO PARA O ENIGMA DO ENOVELAMENTO DE PROTEÍNAS

Em 1973, o bioquímico Christian Anfinsen postulou a sua "hipótese termodinâmica", que estabelece que toda a informação necessária para que uma cadeia de aminoácidos se enovele na sua estrutura tridimensional única e funcional está contida na sua própria sequência (Anfinsen, 1973). Este princípio estabeleceu as bases de um dos maiores desafios da biologia computacional: prever a estrutura tridimensional (3D) de uma proteína conhecendo apenas a sua sequência. O problema é de uma complexidade avassaladora; uma proteína pode ter um número de conformações possíveis que supera o número de átomos no universo (Hou et al., 2024).

Resolver este enigma era de uma importância capital. A função biológica de uma proteína — seja catalisar uma reação, transportar uma molécula ou transmitir um sinal — depende intimamente da sua forma tridimensional. Portanto, conhecer a estrutura é um pré-requisito para compreender os mecanismos das doenças a nível molecular, desenhar fármacos que se acoplem a alvos específicos e engenhar novas enzimas para aplicações industriais. Durante décadas, o "padrão-ouro" para determinar estas estruturas foram técnicas experimentais como a cristalografia de raios X e a ressonância magnética nuclear (RMN) (Rahimi et al., 2022; Smyth; Martin, 2000). No entanto, estes métodos são lentos, extremamente dispendiosos e não aplicáveis a todas as proteínas (Bertoline et al., 2023). Isto gerou uma enorme lacuna de conhecimento:



enquanto as bases de dados continham centenas de milhões de sequências de proteínas, apenas se tinha resolvido experimentalmente a estrutura de uma pequena fração, cerca de 200.000 (Bertoline et al., 2023).

3.3 ALPHAFOLD 2: A REVOLUÇÃO NA BIOLOGIA ESTRUTURAL

Em 2020, durante a 14^a edição da competição bienal de avaliação crítica de técnicas para a previsão da estrutura de proteínas (CASP14), o sistema de IA AlphaFold 2, desenvolvido pela DeepMind da Google, produziu resultados que foram descritos como transformacionais e de alta precisão (Jumper et al., 2021). Pela primeira vez, um método computacional era capaz de prever a estrutura das proteínas com uma precisão comparável à dos métodos experimentais, resolvendo em grande medida o problema de 50 anos.

O funcionamento do AlphaFold 2 representa um marco na aplicação do deep learning a problemas científicos. Ao contrário de modelos anteriores, é um sistema *end-to-end* que integra conhecimento físico e biológico diretamente no desenho da sua arquitetura de rede neuronal profunda (Jumper et al., 2021). O processo pode ser resumido da seguinte forma:

1. O sistema toma como entrada a sequência de aminoácidos da proteína alvo.
2. Procura em enormes bases de dados públicas sequências geneticamente relacionadas para construir um alinhamento de sequências múltiplas (MSA). O MSA revela que posições de aminoácidos tenderam a mutar juntas ao longo da evolução, um forte sinal de que estes resíduos, embora distantes na sequência linear, provavelmente estão em contacto físico na estrutura 3D enovelada (Jumper et al., 2021).
3. Uma rede neuronal baseada num "mecanismo de atenção" (um conceito retirado do processamento de linguagem natural) analisa as complexas relações dentro do MSA para inferir as distâncias e orientações entre pares de aminoácidos.
4. Esta informação é utilizada para construir um "grafo espacial" que representa a estrutura da proteína. O sistema refina iterativamente este grafo, passando a informação de um lado para o outro entre a representação 1D da sequência, uma representação 2D das distâncias e a estrutura 3D final, até convergir numa previsão de alta confiança.

O impacto deste avanço foi monumental. O artigo que descreve o método, tornou-se um dos trabalhos científicos mais influentes da história recente, acumulando mais de 40000 citações até setembro de 2025 (Jumper et al., 2021). No entanto, a decisão estratégica da DeepMind de tornar público o código-fonte e, em colaboração com o Instituto Europeu de Bioinformática (EMBL-EBI), criar a AlphaFold Protein Structure Database, foi o que verdadeiramente catalisou a revolução (Varadi et al., 2022). Esta base de dados de acesso livre democratizou o acesso a mais de 300000 previsões estruturais. Esta estratégia de ciência aberta funcionou como um ciclo de retroalimentação massiva: permitiu a milhões de investigadores de todo o mundo utilizar, validar e construir sobre a tecnologia. Isto não só acelerou descobertas em áreas



tão diversas como o desenvolvimento de vacinas contra a malária, a engenharia de enzimas para degradar plásticos e a luta contra a resistência aos antibióticos (Behling et al., 2023; Ko et al., 2022; Liu et al., 2024; Yang et al., 2023), mas também consolidou o estatuto do AlphaFold como o padrão-ouro de uma forma muito mais rápida e sólida do que uma abordagem proprietária poderia ter alcançado.

3.4 ALPHAFOLD 3: A ARQUITETURA DA INTERAÇÃO MOLECULAR

Se o AlphaFold 2 resolveu o problema da estrutura das proteínas como atores individuais, o AlphaFold 3, apresentado em 2024 por Abramson et al., representa o passo lógico e conceptual seguinte: entender as proteínas como componentes de um ecossistema molecular interativo (Abramson et al., 2024). A função biológica raramente é o resultado de uma única molécula; as proteínas interagem constantemente com outras biomoléculas para realizar as suas tarefas. A evolução do AlphaFold 2 para o AlphaFold 3 não é apenas uma melhoria técnica, mas um reflexo computacional de um amadurecimento no pensamento biológico, passando de uma abordagem centrada na "peça" para uma centrada na "máquina molecular" completa.

O avanço na precisão é notável: o modelo mostra pelo menos uma melhoria de 50% na previsão de interações entre proteínas e outros tipos de moléculas em comparação com os métodos existentes (Abramson et al., 2024). Isto é alcançado através de uma arquitetura de rede neuronal completamente redesenhada, baseada num "modelo de difusão". Esta abordagem começa com uma "nuvem" de átomos desordenada e, através de um processo iterativo de refinamento, converge na estrutura tridimensional mais provável do complexo molecular como um todo (Abramson et al., 2024).

3.5 ESTUDO DE CASO: A CAÇA POR ANTIBIÓTICOS NO "TERCEIRO DOMÍNIO" DA VIDA

A resistência aos antimicrobianos (RAM) é uma pandemia silenciosa que avança a um ritmo alarmante. Segundo dados de 2019, as infeções resistentes a medicamentos foram a causa direta de pelo menos 1,27 milhões de mortes em todo o mundo e estiveram associadas a quase 5 milhões de mortes adicionais (Murray et al., 2022b). As projeções são ainda mais sombrias: um relatório apoiado pelas Nações Unidas adverte que, se não forem tomadas medidas drásticas, o número de mortes anuais por RAM poderá ascender a 10 milhões até 2050 (Naddaf, 2024). O impacto económico é igualmente devastador, com custos que incluem estadias hospitalares mais longas, a necessidade de medicamentos mais caros e uma possível redução de 3,4 biliões de dólares (Ahmed et al., 2024; Naddaf, 2024).

Face à crescente crise de resistência antimicrobiana, a necessidade de descobrir novas classes de antibióticos é desesperada. A busca tradicional, centrada em bactérias do solo e fungos, tem apresentado rendimentos decrescentes, o que impulsionou os investigadores a explorar fronteiras biológicas mais insólitas. Uma das mais promissoras é o domínio das Arqueias (Archaea), o "terceiro domínio" da vida,



formado por organismos unicelulares com características genéticas e bioquímicas que os distinguem tanto das bactérias como dos eucariotas como a sua parede celular e a sua membrana celular única (Baker et al., 2020; Spang; Offre, 2019).

As arqueias são conhecidas por prosperarem em alguns dos ambientes mais extremos da Terra, desde fontes termais a ferver até águas hipersalinas e fontes vulcânicas submarinas (Baker et al., 2020; Chaban; Ng; Jarrell, 2011). A sua capacidade de sobreviver nestas condições sugere uma bioquímica única. No entanto, a sua relevância não se limita a estes nichos, uma vez que também fazem parte de microbiomas complexos como o do intestino humano, o que as torna uma fonte potencial e diretamente relevante de novos compostos bioativos (Guerra, 2024). Apesar deste potencial, muitas arqueias são extremamente difíceis de cultivar, uma barreira que tem limitado o seu estudo.

Para superar esta barreira, emerge uma mudança de paradigma: a "bioprospeção digital. Em vez de depender dos limites do cultivo microbiano, esta abordagem explora diretamente a informação genómica e proteómica armazenada em bases de dados. A inteligência artificial atua como um "prospetor" virtual que analisa estes enormes repositórios em busca de sequências com potencial antimicrobiano. Esta metodologia não só acelera radicalmente a descoberta, como também revela o potencial de organismos previamente inacessíveis.

Um estudo recente demonstra o poder desta abordagem. Utilizando uma plataforma de aprendizagem automática (APEX 1.1), foi realizado uma busca sistemática nos proteomas de 233 espécies de arqueias (Torres; Wan; De La Fuente-Nunez, 2025). Os resultados foram extraordinários: o modelo de IA identificou mais de 12.000 sequências de péptidos com alto potencial antimicrobiano, aos quais deram o nome de "arqueasinas". Para validar as previsões, a equipa sintetizou 80 destas arqueasinas e, em testes *in vitro*, uns espantosos 93% mostraram atividade contra bactéria farmacorresistente. Os testes avançaram para modelos animais (*in vivo*), onde o candidato mais promissor, a arqueasina-73, demonstrou uma eficácia no tratamento de infeções comparável à da polimixina B, um potente antibiótico de último recurso (Torres; Wan; De La Fuente-Nunez, 2025). Este trabalho pioneiro não só valida as arqueias como uma nova e rica fonte de antibióticos, mas também estabelece o poder da IA para desbloquear o potencial oculto no código genético dos cantos mais inexplorados da vida.

4 O FUTURO DA BIOLOGIA É COMPUTACIONAL

A sinergia entre a inteligência artificial e a bioinformática consolidou uma mudança de paradigma fundamental, transformando a biologia de uma ciência descritiva para uma eminentemente preditiva. O sucesso do AlphaFold na resolução do problema do enovelamento de proteínas não representa um ponto final, mas sim o catalisador para uma nova era de investigação aplicada, cujo potencial é demonstrado no seu uso para a prospeção de novos antibióticos em linhagens biológicas pouco exploradas para combater a



crise da resistência aos antimicrobianos. A capacidade de prever com alta precisão a estrutura e as interações de complexos moleculares permite agora guiar a experimentação laboratorial com hipóteses geradas computacionalmente, otimizando drasticamente os ciclos de descoberta. Este poder preditivo estabelece as bases para futuras aplicações, como o desenho de tratamentos personalizados baseados na dinâmica estrutural de proteínas e o desenvolvimento da biologia sintética para criar biomoléculas com funções à medida. A convergência do código biológico e do computacional dotou a ciência das ferramentas para interpretar e, potencialmente, reescrever a lógica molecular da vida, estabelecendo que o futuro da investigação biológica é, inequivocamente, computacional.



REFERÊNCIAS

- ABDELSATTAR, Abdallah S. *et al.* The Role of Molecular Modeling and Bioinformatics in Treating a Pandemic Disease: The Case of COVID-19. **The Open COVID Journal**, v. 1, n. 1, p. 216–234, 23 dez. 2021.
- ABRAMSON, Josh *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. **Nature**, v. 630, n. 8016, p. 493–500, 13 jun. 2024.
- AHMED, Sirwan Khalid *et al.* Antimicrobial resistance: Impacts, challenges, and future prospects. **Journal of Medicine, Surgery, and Public Health**, v. 2, p. 100081, 1 abr. 2024.
- AL-JANABI, Aisha. Has DeepMind's AlphaFold Solved the Protein Folding Problem? **BioTechniques**, v. 72, n. 3, p. 73–76, 4 mar. 2022.
- AMUAH, Charles L. Y. *et al.* Feasibility Study of the Use of Handheld NIR Spectrometer for Simultaneous Authentication and Quantification of Quality Parameters in Intact Pineapple Fruits. **Journal of Spectroscopy**, v. 2019, n. 1, p. 5975461, 1 jan. 2019.
- ANFINSEN, Christian B. Principles that Govern the Folding of Protein Chains. **Science**, v. 181, n. 4096, p. 223–230, 20 jul. 1973a.
- BAKER, Brett J. *et al.* Diversity, ecology and evolution of Archaea. **Nature Microbiology**, v. 5, n. 7, p. 887–900, 1 jul. 2020.
- BASILE, Teodora; MARSICO, Antonio Domenico; PERNIOLA, Rocco. Use of Artificial Neural Networks and NIR Spectroscopy for Non-Destructive Grape Texture Prediction. **Foods**, v. 11, n. 3, p. 281, 1 fev. 2022.
- BEHLING, Anna H. *et al.* Addressing antibiotic resistance: computational answers to a biological problem? **Current Opinion in Microbiology**, v. 74, p. 102305, 1 ago. 2023.
- BEHZADI, Payam; GAJDÁCS, Márió. Worldwide Protein Data Bank (wwPDB): A virtual treasure for research in biotechnology. **European Journal of Microbiology and Immunology**, v. 11, n. 4, p. 77–86, 3 fev. 2022.
- BELLER, Harry R. *et al.* Discovery of enzymes for toluene synthesis from anoxic microbial communities. **Nature Chemical Biology**, v. 14, n. 5, p. 451–457, 19 maio 2018.
- BERTOLINE, Leticia M. F. *et al.* Before and after AlphaFold2: An overview of protein structure prediction. **Frontiers in Bioinformatics**, v. 3, p. 1120370, 28 fev. 2023.
- BIAN, Yuemin; XIE, Xiang-Qun. Computational Fragment-Based Drug Design: Current Trends, Strategies, and Applications. **The AAPS Journal**, v. 20, n. 3, p. 59, 9 maio 2018.
- BLANCO-GONZÁLEZ, Alexandre *et al.* The Role of AI in Drug Discovery: Challenges, Opportunities, and Strategies. **Pharmaceuticals**, v. 16, n. 6, p. 891, 1 jun. 2023.
- BORBA, Karla Rodrigues *et al.* Non-invasive quantification of vitamin C, citric acid, and sugar in 'Valência' oranges using infrared spectroscopies. **Journal of Food Science and Technology**, v. 58, n. 2, p. 731–738, 1 fev. 2021.



CAROLINA CABRAL DA SILVA, Ruana; CIDINARIA SILVA ALVES, Maria. Os avanços e desafios da bioinformática aplicada à saúde: uma revisão. **Diversitas Journal**, v. 9, n. 3, 9 ago. 2024.

CHABAN, Bonnie; NG, Sandy Y. M.; JARRELL, Ken F. Archaeal habitats — from the extreme to the ordinary. <https://doi.org/10.1139/w05-147>, v. 52, n. 2, p. 73–116, fev. 2011.

CLARK, Alexis J.; LILLARD, James W. A Comprehensive Review of Bioinformatics Tools for Genomic Biomarker Discovery Driving Precision Oncology. **Genes**, v. 15, n. 8, p. 1036, 6 ago. 2024.

COSTESSI, Adalberto *et al.* Novel sequencing technologies to support industrial biotechnology. **FEMS Microbiology Letters**, v. 365, n. 16, 1 ago. 2018.

DA SILVA ALVES, Jasciane *et al.* Non-Destructive Detection of Current Internal Disorders and Prediction of Future Appearance in Mango Fruit Using Portable Vis-NIR Spectroscopy. **Horticulturae** **2025**, Vol. 11, Page 759, v. 11, n. 7, p. 759, 1 jul. 2025.

DASGUPTA, Abhijit; DE, Rajat K. Artificial intelligence in systems biology. **Handbook of Statistics**, v. 49, p. 153–201, 1 jan. 2023.

DE FREITAS, Sergio Tonetto *et al.* Mango dry matter content at harvest to achieve high consumer quality of different cultivars in different growing seasons. **Postharvest Biology and Technology**, v. 189, p. 111917, 1 jul. 2022.

DIAZ, Daniel J. *et al.* Using machine learning to predict the effects and consequences of mutations in proteins. **Current Opinion in Structural Biology**, v. 78, p. 102518, fev. 2023.

DINIZ, W. J. S.; CANDURI, F. REVIEW-ARTICLE Bioinformatics: an overview and its applications. **Genetics and Molecular Research**, v. 16, n. 1, 2017.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. Deep Learning. *[S.d.]*.

GUERRA, Abraham. Human associated Archaea: a neglected microbiome worth investigating. **World Journal of Microbiology and Biotechnology**, v. 40, n. 2, p. 1–13, 1 fev. 2024.

HAMID JAMIALAHMADI *et al.* Artificial Intelligence and Bioinformatics: A Journey from Traditional Techniques to Smart Approaches. **Gastroenterology and Hepatology from Bed to Bench**, v. 17, n. 3, 2024.

HOOD, Leroy; ROWEN, Lee. The human genome project: big science transforms biology and medicine. **Genome Medicine**, v. 5, n. 9, p. 79, 2013.

HOU, Minghua *et al.* Protein Multiple Conformation Prediction Using Multi-Objective Evolution Algorithm. **Interdisciplinary Sciences - Computational Life Sciences**, v. 16, n. 3, p. 519–531, 1 set. 2024.

HSU, Jason C.; LU, Christine Y.; HSU, Min Huei. Editorial: Artificial intelligence in infectious diseases: pathogenesis and therapy. **Frontiers in Medicine**, v. 11, p. 1414056, 14 maio 2024.

HUANG, Chujun *et al.* Fusion models for detection of soluble solids content in mandarin by Vis/NIR transmission spectroscopy combined external factors. **Infrared Physics & Technology**, v. 124, p. 104233, 1 ago. 2022.



JANIESCH, Christian; ZSCHECH, Patrick; HEINRICH, Kai. Machine learning and deep learning. **Electronic Markets**, v. 31, n. 3, p. 685–695, 8 set. 2021.

JUMPER, John *et al.* Highly accurate protein structure prediction with AlphaFold. **Nature**, v. 596, n. 7873, p. 583–589, 26 ago. 2021.

KANNAN, Meera *et al.* Leveraging public AI tools to explore systems biology resources in mathematical modeling. **npj Systems Biology and Applications**, v. 11, n. 1, p. 1–8, 1 dez. 2025.

KIM, Sang Yeon *et al.* Neural Network Based Prediction of Soluble Solids Concentration in Oriental Melon Using VIS/NIR Spectroscopy. **Applied Engineering in Agriculture**, v. 37, n. 4, p. 653–663, 2021.

KITANO, Hiroaki. Systems Biology Powered by Artificial Intelligence. p. 1–1, 2012.

KO, Kuang Ting *et al.* Structure of the malaria vaccine candidate Pfs48/45 and its recognition by transmission blocking antibodies. **Nature Communications**, v. 13, n. 1, p. 1–11, 1 dez. 2022.

LIU, Wei *et al.* Structure-guided discovery and rational design of a new poly(ethylene terephthalate) hydrolase from AlphaFold protein structure database. **Journal of Hazardous Materials**, v. 480, p. 136389, 5 dez. 2024.

LU, Zhaohui *et al.* Nondestructive Testing of Pear Based on Fourier Near-Infrared Spectroscopy. **Foods** **2022, Vol. 11, Page 1076**, v. 11, n. 8, p. 1076, 8 abr. 2022.

MA, Lifei *et al.* Comprehensive analyses of bioinformatics applications in the fight against COVID-19 pandemic. **Computational Biology and Chemistry**, v. 95, p. 107599, dez. 2021.

MCMAHON, Jack *et al.* A novel framework for the automated characterization of Gram-stained blood culture slides using a large-scale vision transformer. **Journal of Clinical Microbiology**, v. 63, n. 3, 12 mar. 2025.

MITRA, Debasis *et al.* Evolution of bioinformatics and its impact on modern bio-science in the twenty-first century: Special attention to pharmacology, plant science and drug discovery. **Computational Toxicology**, v. 24, p. 100248, nov. 2022.

MURRAY, Christopher J. L. *et al.* Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. **The Lancet**, v. 399, n. 10325, p. 629–655, fev. 2022a.

MURRAY, Christopher J. L. *et al.* Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. **Lancet (London, England)**, v. 399, n. 10325, p. 629–655, 12 fev. 2022b.

NADDAF, Miryam. 40 million deaths by 2050: toll of drug-resistant infections to rise by 70. **Nature**, v. 633, n. 8031, p. 747–748, 1 set. 2024.

POPLIN, Ryan *et al.* A universal snp and small-indel variant caller using deep neural networks. **Nature Biotechnology**, v. 36, n. 10, p. 983, 1 nov. 2018.

POURDARBANI, Raziieh *et al.* Using metaheuristic algorithms to improve the estimation of acidity in Fuji apples using NIR spectroscopy. **Ain Shams Engineering Journal**, v. 13, n. 6, p. 101776, 1 nov. 2022.



RABBANI, Bahareh; TEKIN, Mustafa; MAHDIEH, Nejat. The promise of whole-exome sequencing in medical genetics. **Journal of Human Genetics**, v. 59, n. 1, p. 5–15, 7 jan. 2014.

RAHIMI, Mehdi *et al.* A toolset for the solid-state NMR-based 3D structure calculation of proteins. **Journal of Magnetic Resonance**, v. 339, p. 107214, 1 jun. 2022.

REHMAN, Ashfaq Ur *et al.* Role of artificial intelligence in revolutionizing drug discovery. **Fundamental Research**, v. 5, n. 3, p. 1273–1287, 1 maio 2025.

SAWYER, Abigail; FREE, Tristan; MARTIN, Joseph. Metagenomics: Preventing Future Pandemics. **BioTechniques**, v. 70, n. 1, p. 1–4, 15 jan. 2021.

SAYERS, Eric W. *et al.* GenBank. **Nucleic Acids Research**, v. 49, n. D1, p. D92–D96, 8 jan. 2021.

SCALISI, Alessio; O'CONNELL, Mark Glenn. Application of visible/NIR spectroscopy for the estimation of soluble solids, dry matter and flesh firmness in stone fruits. **Journal of the Science of Food and Agriculture**, v. 101, n. 5, p. 2100–2107, 30 mar. 2021.

SEVEN, İsmet *et al.* Predicting hepatocellular carcinoma survival with artificial intelligence. **Scientific Reports**, v. 15, n. 1, p. 1–14, 1 dez. 2025.

SMYTH, M. S.; MARTIN, J. H. J. x ray crystallography. **Molecular pathology : MP**, v. 53, n. 1, p. 8–14, 2000.

SPANG, Anja; OFFRE, Pierre. Towards a systematic understanding of differences between archaeal and bacterial diversity. **Environmental microbiology reports**, v. 11, n. 1, p. 9–12, 1 fev. 2019.

SRIPAURYA, Tanachart *et al.* Gros Michel banana soluble solids content evaluation and maturity classification using a developed portable 6 channel NIR device measurement. **Measurement**, v. 173, p. 108615, 1 mar. 2021.

SUBEDI, Phul P.; WALSH, Kerry B. Assessment of avocado fruit dry matter content using portable near infrared spectroscopy: Method and instrumentation optimisation. **Postharvest Biology and Technology**, v. 161, p. 111078, 1 mar. 2020.

TORRES, Marcelo D. T.; WAN, Fangping; DE LA FUENTE-NUNEZ, Cesar. Deep learning reveals antibiotics in the archaeal proteome. **Nature Microbiology**, v. 10, n. 9, p. 2153–2167, 1 set. 2025.

TORRINGTON, Ebony. Bioinformaticians: the Hidden Heroes of the COVID-19 Pandemic. **BioTechniques**, v. 72, n. 5, p. 171–174, 5 maio 2022.

UESAKA, Kazuma *et al.* Bioinformatics in bioscience and bioengineering: Recent advances, applications, and perspectives. **Journal of Bioscience and Bioengineering**, v. 134, n. 5, p. 363–373, nov. 2022.

VAN DEN BOGERT, Bartholomeus *et al.* On the Role of Bioinformatics and Data Science in Industrial Microbiome Applications. **Frontiers in Genetics**, v. 10, 9 ago. 2019.

VARADI, Mihaly *et al.* AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. **Nucleic Acids Research**, v. 50, n. D1, p. D439–D444, 7 jan. 2022.



XIA, Xuhua. Bioinformatics and Drug Discovery. **Current Topics in Medicinal Chemistry**, v. 17, n. 15, p. 1709–1726, 26 abr. 2017.

YANG, Zhenyu *et al.* AlphaFold2 and its applications in the fields of biology and medicine. **Signal Transduction and Targeted Therapy**, v. 8, n. 1, p. 1–14, 1 dez. 2023.

ZHANG, Shujun *et al.* The role and application of bioinformatics techniques and tools in drug discovery. **Frontiers in Pharmacology**, v. 16, 13 fev. 2025.

ZHOU, Hongyi; ASTORE, Courtney; SKOLNICK, Jeffrey. PHEVIR: an artificial intelligence algorithm that predicts the molecular role of pathogens in complex human diseases. **Scientific Reports**, v. 12, n. 1, p. 1–14, 1 dez. 2022.

ZHU, Yicheng; ONG, Cheng Soon; HUTTLEY, Gavin A. Machine Learning Techniques for Classifying the Mutagenic Origins of Point Mutations. **Genetics**, v. 215, n. 1, p. 25–40, 1 maio 2020.