ISBN: 978-65-83849-10-6

Aurum EDITORA

CHAPTER 10

PREDICTIVE BIOLOGY: THE REVOLUTION DRIVEN BY ARTIFICIAL INTELLIGENCE

https://doi.org/10.63330/aurumpub.015-010

Abraham Guerra[1], Nilo Ricardo Corrêa de Mello Júnior[2], Anderson Manares-Romero[3] and Luzia Micaele Alves Barbosa[4]

ABSTRACT

The convergence of artificial intelligence and bioinformatics has initiated a scientific revolution, transforming our ability to interpret biological complexity. This paper argues that the recent solution to the protein folding problem, achieved by AI tools such as AlphaFold, has provided an unprecedented technological arsenal to address global health crises such as antimicrobial resistance (AMR). The concept of "digital bioprospecting" is introduced as a new paradigm that, instead of cultivating microorganisms, directly explores genomic and proteomic data. This method now enables the investigation of previously inaccessible biological frontiers, highlighting the domain of Archaea as a target of particular interest. Due to their unique biochemistry and status as a largely unexplored branch of life, recent computational analyses of their proteomes have revealed a rich potential for new families of antimicrobial compounds ("archaeasins"), thereby validating their importance as a crucial source for the discovery of future antibiotics. Thus, the transition from a descriptive to a predictive science, driven by this synergy, inaugurates a new era in drug design and biological engineering.

Keywords: Machine learning; Bioinformatics; AlphaFold; Antimicrobial resistance; Archaea.

[1] São Paulo State University "Júlio de Mesquita Filho" – School of Agricultural and Veterinary Sciences
Via de Acesso Professor Paulo Donato Castelane Castellane S/N – Vila Industrial, Jaboticabal, São Paulo
[2] São Paulo State University "Júlio de Mesquita Filho" – School of Agricultural and Veterinary Sciences
Via de Acesso Professor Paulo Donato Castelane Castellane S/N – Vila Industrial, Jaboticabal, São Paulo
[3] São Paulo State University "Júlio de Mesquita Filho" – School of Agricultural and Veterinary Sciences
Via de Acesso Professor Paulo Donato Castelane Castellane S/N – Vila Industrial, Jaboticabal, São Paulo
[4] State University of Bahia
Rua Castro Alves, 291, Juazeiro, Bahia

INTRODUCTION

At the confluence of biology and computing, a scientific revolution of unprecedented scale is emerging. This transformation is grounded in the convergence of two fundamental languages: the genetic code—the sequence of nucleotides that dictates the form and function of every living being—and the computational code of artificial intelligence (AI), which grants us the ability to process and comprehend biological complexity at a speed and depth previously unimaginable. The fusion of these two codes does not represent an incremental improvement but rather a paradigm shift that is redefining the boundaries of discovery. For the first time, it enables us to decipher the hidden "grammar" of proteins, the macromolecules that execute life's instructions (Jumper et al., 2021).

This new era is defined by our capacity to confront challenges that, for a long time, were considered intractable. The first of these is the "grand challenge" of protein folding, an enigma that has occupied the scientific community for more than 50 years (Anfinsen, 1973; Jumper et al., 2021a). Predicting the intricate three-dimensional structure of a protein from its linear amino acid sequence was a task of astronomical computational complexity, yet fundamental, since structure determines function. The second challenge is an urgent global public health crisis: antimicrobial resistance (AMR). Declared by the World Health Organization (WHO) as one of the top ten threats to humanity, AMR threatens to reverse a century of medical progress, rendering common infections once again lethal due to stagnation in the discovery of new antibiotics (Murray et al., 2022a).

The central argument of this work is that the recent solution to the first challenge, driven by AI through tools such as AlphaFold, has provided us with an entirely new conceptual and technological arsenal to tackle the second. By deciphering the structural code of proteins, we open innovative pathways for discovering and designing the drugs of the future, ushering in a new era in which the code of life is no longer merely read but begins to be understood and, potentially, rewritten.

BIOINFORMATICS: THE LANGUAGE OF BIOLOGY IN THE DIGITAL ERA

DEFINITION AND FOUNDATIONS

Bioinformatics is the scientific discipline situated at the intersection of biology, computer science, mathematics, and statistics (Carolina Cabral da Silva; Cidinaria Silva Alves, 2024; Uesaka et al., 2022). Its fundamental purpose is to develop methods and tools for storing, analyzing, and, crucially, interpreting the vast amount of biological data generated by modern technologies (Carolina Cabral da Silva; Cidinaria Silva Alves, 2024; Diniz; Canduri, 2017; Uesaka et al., 2022). In essence, bioinformatics functions as the "linguistic component of genetics." Just as a linguist studies language patterns to understand meaning and structure, a bioinformatician analyzes patterns in DNA and protein sequences to identify genes, determine their functions, establish evolutionary relationships, and predict their three-dimensional conformation.

This discipline serves as the indispensable bridge linking raw biological data—such as the billions of bases in a sequenced genome—to applicable biological knowledge and functional understanding (Uesaka et al., 2022).

The need for bioinformatics arose from an inescapable reality: advances in technologies such as next-generation sequencing (NGS) have produced millions of biological data points (Rabbani; Tekin; Mahdieh, 2014). This volume of information exceeded the capacity of traditional analytical methods. Without a computational framework to organize, query, and analyze these data, the Human Genome Project, for example, would have resulted in an unintelligible collection of letters rather than a functional map of our species. Bioinformatics provided this framework by combining massive biological databases with sophisticated algorithms and statistical models (Clark; Lillard, 2024; Sayers et al., 2021). Therefore, the birth of "*big data* biology" was not solely a product of sequencing technologies but the result of a symbiosis between large-scale data generation and bioinformatics' ability to transform these data into structured knowledge.

THE TRANSFORMATIVE IMPACT ON SCIENCE AND MEDICINE

The catalyst that demonstrated the power of bioinformatics on a global scale was the completion of the Human Genome Project in 2004 (Hood; Rowen, 2013). This milestone was not only a triumph of molecular biology but also a bioinformatic feat that rendered manageable a task of enormous scale and complexity. Since then, its applications have permeated nearly every field of life sciences, generating a transformative impact.

Among its most notable applications are:

- **Drug Development:** Bioinformatics tools drastically accelerate the discovery of new drugs. They enable the identification of therapeutic targets at the molecular level and the execution of computational simulations to predict how a potential drug will interact with its target, reducing costs and time in the early stages of research (Behzadi; Gajdács, 2022; Bian; Xie, 2018; Xia, 2017; Zhang et al., 2025).
- **Public Health and Epidemiology:** The role of bioinformatics was crucial during the COVID-19 pandemic. Rapid computational analysis of the SARS-CoV-2 genomic sequence allowed scientists to decipher the molecular structure of its proteins, an essential step for designing and developing effective vaccines in record time (Abdelsattar et al., 2021; Al-Janabi, 2022; Ma et al., 2021; Sawyer; Free; Martin, 2021; Torrington, 2022). Additionally, bioinformatic models are used to predict the evolution and spread of epidemics.
- **Industry, Biotechnology, and Agriculture:** In the industrial sector, bioinformatics is essential for "genome mining," which involves exploring microbial genomes for genes that

produce enzymes of interest. These biocatalysts are optimized and employed to develop more efficient and environmentally friendly industrial processes for manufacturing detergents, biofuels, and food products (Beller et al., 2018; Costessi et al., 2018; Mitra et al., 2022; Van den Bogert et al., 2019).

Similarly, in agriculture, the integration of bioinformatics with non-destructive technologies has opened new frontiers. Methods such as near-infrared spectroscopy (NIR) have shown promise for assessing fruit quality, enabling simultaneous analysis of multiple attributes with minimal sample preparation. This technique has proven effective in crops such as mango (Da Silva Alves et al., 2025; De Freitas et al., 2022), pineapple (Amuah et al., 2019), avocado (Subedi; Walsh, 2020), orange (Borba et al., 2021), nectarine (Scalisi; O'Connell, 2021), banana (Sripaurya et al., 2021), melon (Kim et al., 2021), apple (Pourdarbani et al., 2022), pear (Lu et al., 2022), tangerine (Huang et al., 2022), and kiwi (Basile; Marsico; Perniola, 2022). By combining these large volumes of spectral data with bioinformatic algorithms, it becomes possible not only to predict quality attributes but also to identify patterns related to nutrition, ripening, and even physiological disorders, bringing agriculture closer to an increasingly data-driven approach.

ARTIFICIAL INTELLIGENCE: THE COMPUTATIONAL ENGINE OF BIOLOGICAL DISCOVERY

In the context of life sciences, artificial intelligence is not an abstract consciousness but a set of advanced computational tools designed to analyze complex biological data, model natural processes, and, fundamentally, make predictions based on patterns often undetectable by human analysis. AI enables researchers to decode vast genomic datasets, simulate molecular behaviors, and model ecological interactions at an unprecedented scale and speed (Hamid Jamialahmadi et al., 2024).

To understand its application, it is useful to distinguish two key concepts:

Machine Learning: A branch of AI in which algorithms are trained with large datasets to "learn" how to perform specific tasks without being explicitly programmed for each rule (Janiesch; Zschech; Heinrich, 2021). For example, a machine learning model can be fed thousands of genetic mutation sequences, some labeled as benign and others as pathogenic. Over time, the algorithm learns to identify the distinctive patterns of each class and can predict the nature of a new mutation it has never encountered before (Diaz et al., 2023; Zhu; Ong; Huttley, 2020).

Neural Networks and Deep Learning: Deep learning is a subfield of machine learning that employs architectures called artificial neural networks, loosely inspired by the structure of the human brain[40,41]. These networks consist of multiple layers of interconnected "neurons" (hence the term "deep"), where each layer learns to recognize increasingly complex and abstract features of the data[40]. This

hierarchical structure is particularly effective for analyzing biological data, which are massive, high-dimensional, and often unstructured, such as genomic sequences or microscopic images. A key advantage of deep learning is its robustness in handling large-scale, unstructured datasets. Alharbi et al. (2022) note that "*deep learning methods have shown unprecedented performance in genomics, especially when analyzing large-scale, high-dimensional data sets*" [42]. This capability has enabled significant advances in bioinformatics, such as genomic sequence analysis, pattern recognition in biomedical images, and protein structure prediction.

Moreover, deep neural networks have achieved performance comparable to or even surpassing that of human experts. As Goodfellow et al. (2016) describe, "*deep learning has drastically improved the state of the art in speech recognition, visual object recognition, object detection, and many other domains,*" and this technological revolution now extends to predictive biology (Goodfellow; Bengio; Courville, [n.d.]).

The true revolution of AI in biology lies not merely in automation or increased analytical speed but in a fundamental paradigm shift: from a predominantly descriptive science to a predictive one. Traditional biology and classical bioinformatics focused on describing and cataloging the natural world: sequencing a genome, determining a protein structure, annotating a gene's function. AI, particularly deep learning, excels at prediction. It does not merely describe what exists but anticipates what could exist or what will occur. Examples include predicting a protein's three-dimensional structure from its linear sequence, forecasting the pathogenicity of a genetic variant, or estimating the efficacy of a pharmacological compound (Hsu; Lu; Hsu, 2024; Jumper et al., 2021; Zhou; Astore; Skolnick, 2022). This predictive power allows laboratory experiments to be guided by computationally generated hypotheses with a high probability of success, drastically optimizing the use of time and resources in research.

FUNDAMENTAL APPLICATIONS OF AI IN BIOLOGY

The predictive capabilities of AI are already being harnessed across a wide range of biological and medical applications:

- **Genomics:** Tools such as Google's DeepVariant employ deep neural networks to analyze millions of genomic sequences and identify genetic variants with accuracy surpassing previous methods, which is crucial for diagnosing genetic diseases (Poplin et al., 2018).
- **Drug Discovery:** AI is transforming the lengthy and costly process of drug development. Algorithms can analyze vast chemical libraries to identify candidate molecules with therapeutic potential, predict their pharmacological properties, and even optimize their structures to enhance efficacy and safety (Blanco-González et al., 2023; Rehman et al., 2025).

- **Systems Biology:** AI enables the modeling of complex networks of interactions among genes, proteins, and metabolites that govern cellular processes (Dasgupta; De, 2023; Kannan et al., 2025; Kitano, 2012). These models help scientists understand how small perturbations—such as a mutation or the action of a drug—can affect the entire biological system.
- **Medical Diagnostics:** Deep learning algorithms, particularly those based on computer vision, have demonstrated remarkable ability to analyze medical images. They can, for instance, identify early signs of melanoma in skin lesion photographs, detect anomalies in radiographs, or predict Gram-staining outcomes directly from microscopy images, thereby automating and accelerating bacterial identification (McMahon et al., 2025; Seven et al., 2025).

THE SOLUTION TO THE PROTEIN FOLDING ENIGMA

In 1973, biochemist Christian Anfinsen postulated his "thermodynamic hypothesis," which states that all the information necessary for a polypeptide chain to fold into its unique, functional three-dimensional structure is contained within its own sequence (Anfinsen, 1973). This principle laid the foundation for one of the greatest challenges in computational biology: predicting the three-dimensional (3D) structure of a protein based solely on its sequence. The problem is overwhelmingly complex; a protein can theoretically adopt a number of possible conformations that exceeds the number of atoms in the universe (Hou et al., 2024).

Solving this enigma was of paramount importance. A protein's biological function—whether catalyzing a reaction, transporting a molecule, or transmitting a signal—depends intimately on its 3D shape. Therefore, knowing the structure is a prerequisite for understanding disease mechanisms at the molecular level, designing drugs that bind to specific targets, and engineering new enzymes for industrial applications. For decades, the "gold standard" for determining these structures relied on experimental techniques such as X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy (Rahimi et al., 2022; Smyth; Martin, 2000). However, these methods are slow, extremely costly, and not applicable to all proteins (Bertoline et al., 2023). This created a massive knowledge gap: while databases contained hundreds of millions of protein sequences, only a small fraction—about 200,000—had experimentally resolved structures (Bertoline et al., 2023).

ALPHAFOLD 2: THE REVOLUTION IN STRUCTURAL BIOLOGY

In 2020, during the 14th edition of the biennial Critical Assessment of protein Structure Prediction (CASP14), the AI system AlphaFold 2, developed by Google's DeepMind, delivered results described as transformational and highly accurate (Jumper et al., 2021). For the first time, a computational method was

capable of predicting protein structures with an accuracy comparable to experimental methods, effectively solving a 50-year-old problem.

The functioning of AlphaFold 2 represents a milestone in the application of deep learning to scientific challenges. Unlike previous models, it is an end-to-end system that integrates physical and biological knowledge directly into the design of its deep neural network architecture (Jumper et al., 2021). The process can be summarized as follows:

1. The system takes as input the amino acid sequence of the target protein.
2. It searches massive public databases for genetically related sequences to construct a multiple sequence alignment (MSA). The MSA reveals positions of amino acids that have tended to mutate together throughout evolution—a strong signal that these residues, although distant in the linear sequence, are likely in physical contact in the folded 3D structure (Jumper et al., 2021).
3. A neural network based on an "attention mechanism" (a concept borrowed from natural language processing) analyzes the complex relationships within the MSA to infer distances and orientations between amino acid pairs.
4. This information is used to build a "spatial graph" representing the protein's structure. The system iteratively refines this graph, passing information back and forth between the 1D sequence representation, a 2D distance map, and the final 3D structure until it converges on a high-confidence prediction.

The impact of this breakthrough was monumental. The paper describing the method became one of the most influential scientific works in recent history, accumulating more than 40,000 citations by September 2025 (Jumper et al., 2021). However, the strategic decision by DeepMind to make the source code publicly available and, in collaboration with the European Bioinformatics Institute (EMBL-EBI), to create the AlphaFold Protein Structure Database truly catalyzed the revolution (Varadi et al., 2022). This open-access database democratized access to more than 300,000 structural predictions. This open-science strategy functioned as a massive feedback loop: it allowed millions of researchers worldwide to use, validate, and build upon the technology. This not only accelerated discoveries in areas as diverse as malaria vaccine development, enzyme engineering for plastic degradation, and the fight against antibiotic resistance (Behling et al., 2023; Ko et al., 2022; Liu et al., 2024; Yang et al., 2023), but also established AlphaFold as the gold standard far more rapidly and robustly than a proprietary approach could have achieved.

ALPHAFOLD 3: THE ARCHITECTURE OF MOLECULAR INTERACTION

If AlphaFold 2 solved the problem of protein structure as individual actors, AlphaFold 3, introduced in 2024 by Abramson et al., represents the logical and conceptual next step: understanding proteins as components of an interactive molecular ecosystem (Abramson et al., 2024). Biological function rarely results from a single molecule; proteins constantly interact with other biomolecules to perform their tasks. The evolution from AlphaFold 2 to AlphaFold 3 is not merely a technical improvement but a computational reflection of a maturing biological perspective—shifting from a "part-centered" approach to one focused on the complete "molecular machine."

The improvement in accuracy is remarkable: the model demonstrates at least a 50% enhancement in predicting interactions between proteins and other types of molecules compared to existing methods (Abramson et al., 2024). This is achieved through a completely redesigned neural network architecture based on a "diffusion model." This approach begins with a disordered "cloud" of atoms and, through an iterative refinement process, converges on the most probable three-dimensional structure of the entire molecular complex (Abramson et al., 2024).

CASE STUDY: THE HUNT FOR ANTIBIOTICS IN THE "THIRD DOMAIN" OF LIFE

Antimicrobial resistance (AMR) is a silent pandemic advancing at an alarming pace. According to 2019 data, drug-resistant infections were the direct cause of at least 1.27 million deaths worldwide and were associated with nearly 5 million additional deaths (Murray et al., 2022b). Projections are even more dire: a United Nations–backed report warns that, without drastic measures, annual deaths from AMR could reach 10 million by 2050 (Naddaf, 2024). The economic impact is equally devastating, with costs including longer hospital stays, the need for more expensive medications, and a potential reduction of $3.4 trillion in global GDP (Ahmed et al., 2024; Naddaf, 2024).

Faced with the escalating AMR crisis, the need to discover new classes of antibiotics is desperate. Traditional searches, focused on soil bacteria and fungi, have yielded diminishing returns, prompting researchers to explore more unconventional biological frontiers. One of the most promising is the domain of Archaea—the "third domain" of life—comprising unicellular organisms with genetic and biochemical characteristics that distinguish them from both bacteria and eukaryotes, such as their unique cell wall and membrane structures (Baker et al., 2020; Spang; Offre, 2019).

Archaea are known for thriving in some of Earth's most extreme environments, from boiling hot springs to hypersaline waters and deep-sea hydrothermal vents (Baker et al., 2020; Chaban; Ng; Jarrell, 2011). Their ability to survive under such conditions suggests a unique biochemistry. However, their relevance is not limited to these niches, as they are also part of complex microbiomes such as the human gut, making them a potentially significant and directly relevant source of new bioactive compounds

(Guerra, 2024). Despite this potential, many archaea are extremely difficult to culture, a barrier that has limited their study.

To overcome this barrier, a paradigm shift has emerged: digital bioprospecting. Instead of relying on the constraints of microbial cultivation, this approach directly explores genomic and proteomic information stored in databases. Artificial intelligence acts as a virtual "prospector," analyzing these vast repositories for sequences with antimicrobial potential. This methodology not only radically accelerates discovery but also reveals the potential of previously inaccessible organisms.

A recent study demonstrates the power of this approach. Using a machine learning platform (APEX 1.1), researchers conducted a systematic search across the proteomes of 233 archaeal species (Torres; Wan; De La Fuente-Nunez, 2025). The results were extraordinary: the AI model identified more than 12,000 peptide sequences with high antimicrobial potential, which the authors named "archaeasins." To validate these predictions, the team synthesized 80 of these archaeasins, and in vitro tests showed that an astonishing 93% exhibited activity against drug-resistant bacteria. The tests progressed to animal models (in vivo), where the most promising candidate, archaeasin-73, demonstrated efficacy in treating infections comparable to that of polymyxin B, a potent last-resort antibiotic (Torres; Wan; De La Fuente-Nunez, 2025). This pioneering work not only validates archaea as a new and rich source of antibiotics but also establishes the power of AI to unlock the hidden potential encoded in the genomes of life's most unexplored corners.

THE FUTURE OF BIOLOGY IS COMPUTATIONAL

The synergy between artificial intelligence and bioinformatics has consolidated a fundamental paradigm shift, transforming biology from a descriptive science into a predominantly predictive one. The success of AlphaFold in solving the protein folding problem does not represent an endpoint but rather the catalyst for a new era of applied research, whose potential is exemplified by its use in prospecting new antibiotics from underexplored biological lineages to combat the antimicrobial resistance crisis. The ability to predict with high precision the structure and interactions of molecular complexes now allows laboratory experimentation to be guided by computationally generated hypotheses, drastically optimizing discovery cycles. This predictive power lays the foundation for future applications, such as the design of personalized treatments based on the structural dynamics of proteins and the development of synthetic biology to create biomolecules with tailored functions. The convergence of the biological and computational codes has equipped science with the tools to interpret and, potentially, rewrite the molecular logic of life, establishing that the future of biological research is, unequivocally, computational.

REFERENCES

1. 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.

2. ABRAMSON, Josh et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature, v. 630, n. 8016, p. 493–500, 13 jun. 2024.

3. 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.

4. AL-JANABI, Aisha. Has DeepMind's AlphaFold Solved the Protein Folding Problem? BioTechniques, v. 72, n. 3, p. 73–76, 4 mar. 2022.

5. 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.

6. ANFINSEN, Christian B. Principles that Govern the Folding of Protein Chains. Science, v. 181, n. 4096, p. 223–230, 20 jul. 1973a.

7. BAKER, Brett J. et al. Diversity, ecology and evolution of Archaea. Nature Microbiology, v. 5, n. 7, p. 887–900, 1 jul. 2020.

8. 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.

9. 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.

10. BEHZADI, Payam; GAJDÁCS, Mário. 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.

11. 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.

12. BERTOLINE, Letícia M. F. et al. Before and after AlphaFold2: An overview of protein structure prediction. Frontiers in Bioinformatics, v. 3, p. 1120370, 28 fev. 2023.

13. 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.

14. 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.

15. 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.

16. 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.

17. 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.

18. 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.

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

20. 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.

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

22. 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.

23. 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.

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

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

26. 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.

27. 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.

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

29. 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.

30. 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.

31. 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.

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

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

34. 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.

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

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

37. 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.

38. 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.

39. 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.

40. 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.

41. 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.

42. 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.

43. 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.

44. MURRAY, Christopher JL 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.

45. 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.

46. 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.

47. POURDARBANI, Razieh 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.

48. 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.

49. 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.

50. 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.

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

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

53. 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.

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

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

56. 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.

57. 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.

58. 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.

59. 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.

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

61. 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.

62. 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.

63. 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.

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

65. 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.

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

67. 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.

68. 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.