Computational Molecular Biology: Interdisciplinary Technological Integration and New Advances

Jessi White [1] ✉, Garen Lee [2]

1 Cuixi Academy of Biotechnology, Zhuji, 311800, China

2 Hangzhou Lin'an Aikang Biotechnology Co., Ltd., Lin'an, 311300, China

✉ Corresponding author email: jessi.j.zhang@foxmail.com

Abstract This review presents the development, fundamental techniques, applications, and future directions of Computational Molecular Biology. Computational Molecular Biology is an interdisciplinary field that integrates knowledge from computer science, statistics, and biology to study molecular biology problems. The review emphasizes the fundamental techniques in Computational Molecular Biology and discusses its applications in biomedical research. Through the use of Computational Molecular Biology approaches, researchers can gain better insights into the molecular structures and functions within organisms, leading to the design of more effective drugs and treatment strategies, as well as the discovery of new therapeutic targets and pathways. Lastly, the review explores the future directions of Computational Molecular Biology. As these techniques continue to evolve, Computational Molecular Biology will further expand its application scope, bringing about more innovations and breakthroughs in biomedical research.

Keywords Computational Molecular Biology; Biomedical Research; Data Analysis; Risk Prediction

Computational molecular biology is an interdisciplinary field that combines computer science, statistics, and biology (Pevzner, 2000). It emerged from the intersection of biology and information science with the aim of applying computational and mathematical tools and techniques to address biological problems (Zhang and Min, 2012). With the advancement of technology and the deepening of biological research, computational molecular biology has made significant progress over the past few decades and become a vital force driving advancements in biological research and medicine. It provides powerful tools and methods for deciphering the structure, function, and interactions of biological molecules, promoting rapid developments in genomics, proteomics, metabolomics, systems biology, and other fields (Li et al., 2006; Ma et al., 2021). In the information age, the interdisciplinary nature and integration of technologies in computational molecular biology have increasingly played a crucial role in the field of life sciences.

Computational molecular biology is an interdisciplinary field that combines computer science, statistics, and biology. It aims to develop and apply tools and techniques from computer and mathematics (Hong et al., 2006) to address biological problems, particularly those involving molecular-level research. Computational molecular biology can be regarded as a branch of bioinformatics (Zhang, 2000), but it places greater emphasis on the computational simulation, structure prediction, functional analysis, and interactions of biological molecules.

In Computational molecular biology, researchers apply computer algorithms, statistical methods, and mathematical models to process and analyze vast amounts of biological data, such as genomics, proteomics, and metabolomics data (Yu et al., 2008). The data can originate from laboratory techniques such as sequencing, microarrays, and mass spectrometry, or they can come from existing information in public databases. By processing and analyzing the data, computational molecular biology can reveal important information about the interrelationships, signaling pathways, and associations between the structure and function of biological molecules, thereby deepening our understanding of biology.

As an interdisciplinary field, computational molecular biology combines knowledge and techniques from computer science, statistics, and biology. Through computation and simulation, it deciphers the structure, function, and interactions of biological molecules, driving advancements in biological research and bringing about new breakthroughs and progress in medicine and the life sciences.

The review will comprehensively showcase the new technologies, methods, and advancements in the field of computational molecular biology, with a particular emphasis on its significance and unique value in interdisciplinary cross-fertilization, providing readers with a comprehensive and in-depth overview of computational molecular biology, and allowing them to understand the latest progress, technological integration, and application prospects in this interdisciplinary field. By fostering collaboration and communication between different disciplines, it seeks to drive further development and application of computational molecular biology in the life sciences.

1 Development of Computational Molecular Biology

Computational molecular biology has undergone a development process from early protein and DNA sequence analysis to the advancements in genomics, proteomics, and metabolomics, progressively encompassing a broader range of biological molecules and processes. Its interdisciplinary feature has made it an important approach for addressing biological questions, leading to new breakthroughs and advancements in biology and medicine. With the continuous innovation of technology and the promotion of applications, the future prospects of computational molecular biology remain promising,and it is expected to play a more significant role in the fields of biological science and medicine.

1.1 Early research in computational molecular biology

Early research in computational molecular biology can be traced back to the 1950s and 1960s when the technological conditions and computer equipment were far more limited compared to today. Nonetheless, researchers still explored the integration of computation with biology actively. During this period, computational molecular biology primarily focused on protein sequence analysis and DNA sequence analysis, laying the groundwork for the subsequent development of computational biology (Gao et al., 2001).

Protein sequence analysis was one of the main research directions in early computational molecular biology. Scientists have started using simple computational methods and manual calculations to analyze amino acid sequences of known protein, searching for patterns and regularities in protein structure and function. DNA sequence analysis was another important research direction. Scientists have started comparing and classifying DNA sequences to explore the genetic relationships between different species. At that time, sequence alignment tasks were mostly performed manually, with computers serving as auxiliary tools. Additionally, early computational molecular biology also involved computational simulations of protein structures.

Although early research in computational molecular biology was limited by technological conditions, the efforts of researchers laid the foundation for its development. Through exploration and accumulation during that time, computational molecular biology gradually developed into an indispensable and important field in modern biology, covering multiple fields including genomics, proteomics, metabolomics, and systems biology, providing powerful tools and methods to address biological problems and promote medical research (Cock et al., 2009). With the continuous advancement of computing power and technology, the development of computational molecular biology has entered a new stage, continuously driving progress and advancement in the field of biological science.

1.2 The rise of genomics and transcriptomics

The emergence of genomics and transcriptomics marks a new stage in computational molecular biology (Brutlag, 1998), providing revolutionary means for a comprehensive understanding of the genomic structure and function of organisms. The rise of these two fields is attributed to the rapid development of high-throughput sequencing technologies, which enables scientists to quickly and accurately obtain genomic and transcriptomic information from organisms.

The earliest milestone in the field of genomics was the completion of the Human Genome Project in 2001, which took 13 years and ultimately completed the sequencing of the human genome. Subsequently, with the continuous advancement of sequencing technologies and a substantial reduction in costs, genome sequencing of many organisms, including various plants, animals, and microorganisms, has been accomplished (Ejigu and Jung, 2020). Transcriptomics focuses on the gene expression profile of an organism under specific conditions, known as transcriptomics, which represents the gene expression patterns of the genome in a particular physiological state. Transcriptomic studies, through the analysis of changes in RNA sequences, reveals the transcriptional profiles of cells and tissues under different physiological conditions, including which genes are activated or repressed, and which biological processes are involved (Hackl et al., 2016).

The emergence of genomics and transcriptomics has presented new challenges and opportunities for computational molecular biology. The vast amount of genomic and transcriptomic data requires powerful computational capabilities and efficient algorithms for processing and analysis. Researchers in computational molecular biology are actively engaged in developing more efficient and accurate bioinformatics tools and software to address this challenge. Meanwhile, the development of genomics and transcriptomics has also brought new research directions to computational molecular biology, such as single-cell transcriptomics, epigenomics, etc. These emerging fields provide more detailed information to deeply understand the molecular mechanisms of organisms (Pellegrini et al., 2022). Additionally, the rise of genomics and transcriptomics has facilitated the integration of computational molecular biology with other disciplines like biology and medicine, forming a more comprehensive and interdisciplinary research model, providing more powerful and diverse tools and methods for solving biological problems.

1.3 Development of proteomics and metabolomics
The development of proteomics and metabolomics represents two important milestones in the field of computational molecular biology. They respectively focus on the composition, structure, function, and interactions of proteins and metabolites in organisms, providing powerful means to delve into the complex molecular networks in organisms.

Proteomics is the discipline that studies the composition and function of proteins in organisms. With the continuous advancement of protein mass spectrometry technology, researchers can identify and quantify a large number of proteins in a high-throughput manner, thereby revealing the structure and function of proteins in organisms (Taylor et al., 2021). The development of computational molecular biology in the field of proteomics primarily focuses on data analysis and processing, protein structure prediction, function prediction, and network analysis, offering essential theoretical and technical support to decipher complex protein interaction networks and biological processes.

Metabolomics is the discipline that studies the composition and changes of metabolites in organisms. By using nuclear magnetic resonance spectroscopy and mass spectrometry techniques, researchers can obtain a large amount of metabolomic data, thereby comprehensively understanding the activity of metabolic pathways in organisms. The development of computational molecular biology in the field of metabolomics primarily involves data processing and analysis, metabolic pathway analysis, and identification of metabolite biomarkers, providing critical computational methods and tools for in-depth insights into the regulation in organisms.

The development of proteomics and metabolomics has taken an important step in computational molecular biology in studying complex molecular mechanisms and biological processes within organisms. They provide a comprehensive perspective for studying the functions and regulations of biological systems, expanding the application scope of computational molecular biology in biological and medical research. The research achievements in proteomics and metabolomics have provided valuable data resources for complex molecular networks in organisms (Ramalingam et al., 2015), and facilitated the deep integration of computational molecular biology with disciplines such as biology and medicine, leading to novel breakthroughs and advancements in addressing biological problems and promote medical research.

1.4 The interdisciplinary integration of computational molecular biology

As an interdisciplinary field, computational molecular biology benefits from the technological integration of multiple disciplines such as computer science, statistics, mathematics, and biology. With the rise of big data technology and the application of artificial intelligence, computational molecular biology has made significant progress in data processing and analysis, model construction, and biological simulations. Researchers in computational molecular biology have begun to adopt emerging technologies such as deep learning, machine learning, and data mining to handle massive biological data, predict protein structures and functions, and identify new drug targets. Moreover, computational molecular biology also plays a crucial role in the field of biomedical, helping medical researchers better understand the molecular mechanisms of diseases, and promoting the development of personalized medicine and precision treatments.

2 Techniques and Methods of Computational Molecular Biology

In recent years, the field of computational molecular biology has witnessed the emergence of numerous new technologies, tools, and methods. Among them, the cross application of bioinformatics and computer science, the development of databases and algorithms, the application of artificial intelligence and machine learning, as well as the cutting-edge progress in supercomputing and cloud computing, have particularly captured attention. These emerging technologies and methods have brought unprecedented opportunities and challenges to computational molecular biology, driving its continuous development in biological research and medical applications. However, with the constant updates of technology and data, computational molecular biology still faces many challenges, such as data integration and sharing, algorithm interpretability and reliability, etc. In the future, we look forward to further expanding the technology boundaries in the field of computational molecular biology, address real-world biological problems, and make outstanding contributions to the advancement of biological sciences.

2.1 Bioinformatics and computer science

The cross application of bioinformatics and computer science plays an important role in the field of computational molecular biology. In recent research, significant progress has been made in the cross application of bioinformatics and computer science. For instance, deep learning based algorithms for protein structure prediction (Zhang et al., 2017) have become a major highlight in computational molecular biology. These algorithms can accurately predict the three-dimensional structure of proteins by utilizing neural networks and large-scale protein structure databases, providing important references for studying protein functions and interactions.

Additionally, significant advancements has also been made in the fields of genomics and transcriptomics. The high-throughput sequencing technology of genomic data provides rich data resources for genome comparisons, gene identification, and regulatory network analysis. By utilizing algorithms and data processing techniques in computer science, genes can be identified rapidly and accurately, gene functions and regulatory mechanisms can be inferred, and association between genes and diseases can be explored.

Moreover, it also involves the analysis of protein interaction networks and data mining in metabolomics. Utilizing network analysis algorithms to reveal the topological structure and functional modules of protein interaction networks, in order to gain a deeper insights into the regulatory mechanisms of protein interactions. Simultaneously, by applying data mining and machine learning techniques, researchers can extract regulatory mechanisms of metabolic pathways and disease-related metabolic biomarkers from large-scale metabolomics data, offering new insights and strategies for disease diagnosis and treatment.

A review article published on Trends in Biotechnology introduces the application of bioinformatics and computer science in tumor immunotherapy (Eggermont et al., 2014). The paper discusses how bioinformatics methods are used to predict tumor neoantigens and immune checkpoints, as well as how computer science methods are applied to analyze the response and drug resistance of tumor immunotherapy. Furthermore, another review article published on Nature intruduces the application of bioinformatics and computer science in the research of COVID-19 (Figure 1). The paper describes how to use bioinformatics methods to carry out genomics and evolutionary analysis of COVID-19, and how to use computer science methods to develop new strategies for drugs and vaccines (Wang et al., 2022).

nature > nature communications > articles > article

Article | Open Access | Published: 27 July 2022

Differential neutralization and inhibition of SARS-CoV-2 variants by antibodies elicited by COVID-19 mRNA vaccines

Li Wang, Markus H. Kainulainen, Nannan Jiang, Han Di, Gaston Bonenfant, Lisa Mills, Michael Currier, Punya Shrivastava-Ranjan, Brenda M. Calderon, Mili Sheth, Brian R. Mann, Jaber Hossain, Xudong Lin, Sandra Lester, Elizabeth A. Pusch, Joyce Jones, Dan Cui, Payel Chatterjee, M. Harley Jenks, Esther K. Morantz, Gloria P. Larson, Masato Hatta, Jennifer L. Harcourt, Azaibi Tamin, SSEV Bioinformatics Working Group, ... Bin Zhou ✉ + Show authors

Nature Communications **13**, Article number: 4350 (2022) | Cite this article

Figure 1 The review paper published in Nature

2.2 Databases and algorithms

With the rapid accumulation of biological data, the construction and maintenance of biological databases have become an important task in computational molecular biology. Biological databases play a vital role in storing, managing, and sharing biological data, providing researchers with valuable data resources. In recent years, many new biological databases have emerged, covering various fields such as genomics, proteins, and metabolites (Guéguen et al., 2013). These databases not only store a large amount of biological data, but also provide rich data query and analysis tools, providing convenience for researchers in computational molecular biology.

At the same time, new algorithms and data mining techniques are also continuously emerging to better handle and analyze biological data. For example, network analysis algorithms can help reveal protein interaction networks and gene regulatory networks, thereby gaining a deeper understanding of the complex molecular regulatory mechanisms within organisms. Data mining and pattern recognition techniques can extract hidden patterns and regularities from large-scale biological data, assisting researchers in discovering novel biological patterns and biomarkers.

An article published in Nature introduces a gene network analysis method based on large-scale transcriptome data (Van de Sande et al., 2020). This method can be used to predict gene regulatory networks and cell signaling transduction pathways, revealing interactions between biomolecules (Figure 2). Additionally, some databases and tools have also been widely used in the field of computational molecular biology. For instance, the genome browsers can be used to visualize genome sequences and gene annotation information, BLAST can be used for sequence similarity searches, KEGG can be utilized for metabolic and signaling pathway analysis, and so on.

nature > nature protocols > protocols > article

Protocol | Published: 19 June 2020

A scalable SCENIC workflow for single-cell gene regulatory network analysis

Bram Van de Sande, Christopher Flerin, Kristofer Davie, Maxime De Waegeneer, Gert Hulselmans, Sara Aibar, Ruth Seurinck, Wouter Saelens, Robrecht Cannoodt, Quentin Rouchon, Toni Verbeiren, Dries De Maeyer, Joke Reumers, Yvan Saeys & Stein Aerts

Nature Protocols **15**, 2247–2276 (2020) | Cite this article

41k Accesses | **238** Citations | **51** Altmetric | Metrics

Figure 2 presents an article published in Nature that introduces gene network analysis methods based on large-scale transcriptomic data

2.3 Artificial intelligence and machine learnings

The application of Artificial Intelligence (AI) and Machine Learning (ML) in computational molecular biology has become increasingly widespread. AI and ML technologies can learn and identify patterns from complex biological data, thus assisting researchers in data analysis and decision-making.

In the fields of genomics and proteomics, AI and ML technologies have made significant progress in protein structure prediction, gene recognition, and protein functional annotation. For example, by applying ML algorithms to large-scale protein sequence and structure data, researchers can achieve more accurate predictions of protein functions and interactions, providing a deeper understanding for biological research. A paper published in Nature Methods introduces a deep learning based method for protein structure prediction (Marx, 2022). This method utilizes neural network models to predict protein structures from protein sequences, with high accuracy and efficiency (Figure 3).

Figure 3 presents an article published in Nature Methods that introduces protein structure prediction methods based on deep learning

In addition, AI and ML technologies are widely applied in the field of metabolomics. By clustering and classifying metabolomic data, researchers can discover metabolic patterns associated with specific physiological states or diseases, thereby providing new ideas and strategies for disease diagnosis and treatment.

However, the application of artificial intelligence and machine learning technologies in computational molecular biology also faces some challenges. For instance, the interpretability and transparency of algorithms are significant issues in current research. When using machine learning algorithms for prediction and decision-making, researchers often need to understand how the algorithms make predictions and decisions in order to explain the results and verify the reliability of the algorithms. Additionally, data quality and data annotation are also key issues in research on artificial intelligence and machine learning. High-quality data and accurate data annotation are crucial for algorithm training and performance, but in biological research, these issues still need to be addressed due to the diversity of data sources and the instability of data quality.

2.4 Supercomputing and cloud computing

With the continuous expansion of biological data scale and the increasing demand for computing, supercomputing and cloud computing technologies are playing an increasingly important role in computational molecular biology. The powerful computational power of supercomputers enables complex biological simulations and modeling, such as protein molecular dynamics simulations, genome assembly, etc. The high performance and efficiency of supercomputers have provided significant acceleration for biological research.

An article published in Nature introduced simulating the folding process of proteins through supercomputing (Snow et al., 2002). The research utilized supercomputers to simulate the folding process of a protein, revealing the dynamics and structural characteristics of protein folding. Additionally, some cloud computing platforms have also been widely applied in the field of computational molecular biology.

On the other hand, cloud computing technology has made the storage and sharing of large-scale biological data more convenient. Through cloud computing platforms, researchers can quickly access and process remotely stored biological data, avoiding the challenges of data transmission and storage. Moreover, cloud computing provides researchers in computational molecular biology with powerful computing resources, supporting high-performance computing and large-scale data processing. For example, cloud computing platforms such as Amazon Web Services and Google Cloud Platform can provide services such as high-performance computing and large-scale data storage, providing new support and possibilities for biological research and applications, accelerating the progress of computational molecular biology research.

3 Application of Computational Molecular Biology

Computational molecular biology, as an interdisciplinary field, its applications cover various aspects of life sciences, making significant contributions from basic research to applied studies. It has made remarkable progress in disease diagnosis and prediction, providing new diagnostic methods and treatment strategies for clinical medicine. By analyzing genomic and metabolomic data, computational molecular biology can identify potential biomarkers, predict individual reactions and resistance to drugs, forecast the risk of hereditary diseases, and track the spread of infectious diseases. These advancements provide new ways for personalized medicine and precision healthcare, as well as crucial support for early diagnosis and prevention of diseases. Emerging technologies, tools, and methods have brought new breakthroughs and advancements to biological research, propelling the development of life sciences. However, computational molecular biology still faces a series of challenges, such as data processing and interpretation. Therefore, further strengthening interdisciplinary cooperation is necessary to drive the continuous development of this field.

3.1 Genetic function research

Gene function research is a significant application direction in the field of computational molecular biology, which mainly focuses on the functions and regulatory mechanisms of genes in organisms. Computational molecular biology utilizes methods such as analyzing gene regulatory networks, predicting gene functions, guiding gene editing, and uncovering gene-disease associations to help researchers elucidate the roles and interactions of genes in biological processes.

By integrating genomics and transcriptomics data, a gene regulatory network has been constructed to reveal the regulatory relationships and signaling pathways between genes. A study published in Cell has constructed a transcription factor regulatory network in human epithelial cells by integrating large-scale gene expression data, unveiling the regulatory relationships of hundreds of transcription factors, providing important clues for understanding the development and function of epithelial cells (Lambert et al., 2018).

By utilizing genomics and proteomics data, it is possible to predict the functions of genes and the biological processes involved. For example, a study published in Cell used machine learning algorithms to predict a set of potential immune related genes (Swanson et al., 2023), which were subsequently experimentally confirmed to be closely related to immune responses.

By combining gene editing technologies such as CRISPR-Cas9 and computational simulation methods, researchers can predict the efficiency and accuracy of gene editing, providing theoretical support for genome modification. For example, a report published in Nature Biotechnology used computational simulations to predict the editing efficiency of multiple target genes (Koeppel et al., 2023), guiding the gene editing strategies in experiments and improving the success rate of gene editing (Figure 4).

nature > nature biotechnology > articles > article

Article | Open Access | Published: 16 February 2023

Prediction of prime editing insertion efficiencies using sequence features and DNA repair determinants

Jonas Koeppel, Juliane Weller, Elin Madli Peets, Ananth Pallaseni, Ivan Kuzmin, Uku Raudvere, Hedi Peterson, Fabio Giuseppe Liberante & Leopold Parts ✉

Nature Biotechnology (2023) | Cite this article

Figure 4 The paper published in Nature Biotechnology

3.2 Drug development and treatment

Drug development and treatment are important application directions in the field of computational molecular biology. Computational molecular biology provides new technologies and methods for drug development and treatment. By predicting drug targets, optimizing drug doses, and personalized drug design, it accelerates the discovery and development process of new drugs, bringing revolutionary changes to the pharmaceutical field and making greater contributions to improving the effectiveness of disease treatments and the quality of life for patients.

By integrating genomics, proteomics, and chemical information, new drugs can be discovered. For instance, a study published in Nature Reviews Drug Discovery utilizes computational simulation methods to explore the potential of novel antibiotics in bacteria, and genomic mining can provide breakthroughs in novel antibiotics (Cully, 2018) (Figure 5).

nature > nature reviews drug discovery > biobusiness briefs > article

Published: 01 December 2017

Deal watch

Roche taps potential antibiotics mine with Warp Drive Bio

Megan Cully

Nature Reviews Drug Discovery **17**, 9 (2018) | Cite this article

Figure 5 The paper published in Nature Reviews Drug Discovery

By analyzing patients' genomic data and pathological information, it is possible to predict their reactions to specific drugs, thereby enabling personalized drug therapy. For example, a study used computational simulation methods to predict the sensitivity of breast cancer patients to chemotherapy drugs, assisting doctors in developing personalized chemotherapy regimens and improving the treatment effect.

By utilizing computational simulation methods, it is possible to rapidly evaluate the activity and efficacy of drug candidate compounds, accelerating the drug development process. For instance, using computational simulations methods to screen a group of compounds with anti-cancer activity, multiple of which have already entered clinical trial stage.

3.3 Disease diagnosis and prediction

The application of computational molecular biology in disease diagnosis and prediction has also made significant progress. Based on the analysis of genomics and metabolomics data, it helps identify gene mutations and metabolic abnormalities related to disease, thereby providing early disease diagnosis and risk prediction.

By analyzing the genomic data of cancer patients, it is possible to identify gene mutations and protein expression abnormalities associated with tumors. For example, in thr study of breast cancer, researchers have discovered novel biomarkers related to the occurrence and metastasis of breast cancer through the analysis of genomics and proteomics data. These biomarkers provide new ideas for early diagnosis and treatment of breast cancer.

Analyzing individual genomic data can help predict the risk of hereditary diseases. For example, in predictive research on congenital heart disease, computational molecular biology methods can predict whether individuals carry pathogenic genes associated with congenital heart disease, thereby helping families understand genetic risk and take appropriate intervention measures.

By utilizing metabolomics data from organisms, potential biomarkers can be identified, which may change in the early stages of diseases. For example, in the early diagnosis of diabetes, computational molecular biology methods can detect abnormal changes in specific metabolites in the urine of diabetes patients. These metabolites can serve as early markers for disease.

3.4 Frontier research in life sciences

The application of computational molecular biology in cutting-edge research in life sciences involves multiple domains, such as revealing biological evolution, studying biodiversity, exploring drug mechanisms of action, and delving into genomics and transcriptomics. These advancements have led to new breakthroughs in the field of life sciences. With the continuous development of computational molecular biology technology, researchers are now able to delve deeper into the intricate biological processes within organisms, promoting the development and application of biology.

By analyzing genomic data and comparing genomics methods, researchers can unveil the genetic variation and evolutionary relationships between different species. For instance, through the comparison of genomes from multiple species, a set of genes related to the evolution of human brain has been discovered, which may play a crucial role in the evolution of human intelligence.

By utilizing genomic data from environmental samples, high-throughput analysis and classification of microbial communities in the environment can be carried out. For instance, computational analysis methods have been used to conduct in-depth research on microbial communities in the ocean, revealing the microbial diversity and functionalities in marine ecosystems.

Analyzing large-scale genomics and transcriptomics data can identify genes and regulatory networks related to diseases. For example, using computational methods to study the transcriptome of breast cancer cells, researchers discovered a set of new genes related to tumor proliferation and metastasis, providing novel targets for the treatment and prevention of breast cancer.

4 Prospects and Challenges of Computational Molecular Biology

As a rapidly advancing field, computational molecular biology will continue to play a crucial role in the future. We will witness the emergence of numerous new technologies, methods, and progress, driving the widespread application of computational molecular biology in biological research. At the same time, it is also necessary to confront challenges in terms of technology and methodology and carefully consider ethical and legal considerations to ensure that the development and application of computational molecular biology remain safe, reliable, and legal.

4.1 Development trends

As an interdisciplinary field, computational molecular biology holds immense promise for the future. With ongoing advancements in biological research and technology, computational molecular biology will continue to play a crucial role and propelling biological research to new heights.

The arrival of the big data era: With the continuous development in high-throughput sequencing, mass spectrometry, single-cell sequencing, and other technologies, the rate of biological data generation is increasing

exponentially. Computational molecular biology will face challenges in handling and analyzing large-scale data, but at the same time, it will provide a more profound and comprehensive perspective for biological research.

The advancement of personalized healthcare: Computational molecular biology can analyze individual genomic data, disease characteristics, and other information to help doctors achieve personalized treatment plans. Predicting drug sensitivity, diagnosing diseases, and preventing diseases will become important applications of computational molecular biology in the medical field.

The integration of machine learning and artificial intelligence: With the development of machine learning and artificial intelligence technology, computational molecular biology will better explore hidden patterns and information in biological data. By constructing intelligent algorithms and models, it can accelerate the progress of biological research.

The development of systems biology: Computational molecular biology will place more emphasis on studying biological systems at a holistic level, integrating multiple omics data from genomics, transcriptomics, proteomics, to metabolomics. The integration will reveal complex regulatory networks and biological processes within organisms.

Multi-scale modeling and simulation: Computational molecular biology will develop more refined and realistic biological simulation and modeling techniques, including molecular dynamics simulations, cellular simulations, etc., to study the interactions and dynamic changes at the molecular and cellular levels within organisms.

4.2 Challenges faced

With the rapid development of computational molecular biology, there are also some technical and methodological challenges that will impact the application and promotion of computational molecular biology in biological research.

Data quality and standardization: The quality and degree of standardization of biological data directly affect the analysis and interpretation results of computational molecular biology. Ensuring data accuracy and consistency and establishing unified data standards and databases are currently urgent issues to be addressed.

The reliability of algorithm and models: Computational molecular biology involves numerous algorithms and models, and their reliability and reproducibility are crucial for the credibility of research results. It is necessary to strengthening algorithm validation and result reproducibility to ensure the scientific rigor and reliability of computational molecular biology studies.

Cross disciplinary cooperation and talent cultivation: Computational molecular biology requires interdisciplinary integration, covering professional knowledge in multiple disciplines, including biology, computer science, statistics, and others. Strengthening interdisciplinary collaboration and talent cultivation, cultivating professional talents with comprehensive knowledge backgrounds is the key to promoting the advancement of computational molecular biology.

4.3 Consideration of ethical and legal issues

The development of computational molecular biology has also brought about some ethical and legal considerations. With the widespread collection and application of individual genomic data, issues concerning personal privacy and data security are becoming more prominent. Protecting personal privacy and data security, ensuring the legality and compliance of scientific research, is a issue that need continuous attention throughout the development of computational molecular biology.

5 Prospect

In this review article, we comprehensively reviewed the development and achievements of computational molecular biology. Since the 1950s, computational molecular biology has been closely integrated with computer science, statistics, and biology, becoming an interdisciplinary field. From the initial analysis of protein and DNA

sequences to the rise of genomics and transcriptomics, and then the development of proteomics and metabolomics, computational molecular biology has been dricving the progress of biological research.

With the advancement of technology and the continuous development of computational molecular biology, we can foresee the emergence of new technologies, methods, and breakthroughs in the future. Firstly, genome editing technologies like CRISPR-Cas9 will continue to advance and become key tools for precise gene editing and gene therapy. Single-cell sequencing technologies will further improve resolution and accuracy, enabling us to gain deeper insights into cellular heterogeneity and developmental processes. Protein structure prediction techniques will be continuously enhanced through the application of artificial intelligence and other methods, accelerating research on protein function and drug development.

Secondly, network biology will become an important research direction, revealing the roles of protein-protein interaction networks, transcriptional regulatory networks, and others in biological processes. AI-assisted drug development will further accelerate the speed and efficiency of drug research and development. Supercomputing and cloud computing will provide computational molecular biology with more powerful computing and storage capabilities, facilitating the processing and analysis of large-scale data.

This review paper provides an in-depth analysis and evaluation of new technologies, methods, and advancements in the field of computational molecular biology, with a focus on its significant impact on biological research. It also delves into the prospects and challenges faced by computational molecular biology. Computational molecular biology will continue to play a crucial role in biological research, bringing more opportunities for scientific progress and medical development. However, with the rapid development of technology, there are also a series of challenges to address, such as the complexity of data processing and analysis, the reliability of algorithms and models, and ethical and legal considerations. It is essential to call for continuous strengthening of cooperation in the field of computational molecular biology, jointly addressing these challenges and promoting the development and application of computational molecular biology to be safer, more reliable, and more legitimate.

Overall, computational molecular biology, as an interdisciplinary field, has achieved significant accomplishments and has broad vast prospects for future development. We believe that with continuous innovation of technology and methods, computational molecular biology is expected to continue playing a crucial role in the fields of biology and medicine. It will contribute to gaining a deeper understanding of the structure, function, and regulatory mechanisms of organisms, driving advancements in medicine, agriculture, environment, and other areas.

Authors' contributions

Jessi White is the lead researcher of this study, responsible for writing and revising the article, and participating in literature compilation. Garen Lee is responsible for collecting and organizing the literature. Both authors read and approved the final manuscript.

References

Brutlag D.L., 1998, Genomics and computational molecular biology, Current Opinion in Microbiology, 1(3): 340-345.
https://doi.org/10.1016/S1369-5274(98)80039-8

Cock P.J., Antao T., Chang J.T., Chapman B.A., Cox C.J., Dalke A., Friedberg I., Hamelryck T., Kauff F., Wilczynski B., and De Hoon M.J., 2009, Biopython: freely available Python tools for computational molecular biology and bioinformatics, Bioinformatics, 25(11): 1422.
https://doi.org/10.1093/bioinformatics/btp163

Cully M., 2018, Deal watch: Roche taps potential antibiotics mine with Warp Drive Bio, Nature Reviews Drug Discovery, 17(1): 8-9.
https://doi.org/10.1038/nrd.2017.247

Eggermont L.J., Paulis L.E., Tel J., and Figdor C.G., 2014, Towards efficient cancer immunotherapy: advances in developing artificial antigen-presenting cells, Trends in biotechnology, 32(9): 456-465.

https://doi.org/10.1016/j.tibtech.2014.06.007

Ejigu G.F., and Jung J., 2020, Review on the computational genome annotation of sequences obtained by next-generation sequencing, Biology, 9(9): 295.
https://doi.org/10.3390/biology9090295

Gao L., Xu J., and Zhang J.Y., 2001, Research Progress and Prospect of DNA Computing, Dianzi Xuebao (Journal of Electronics), 29(7): 973.

Guéguen L., Gaillard S., Boussau B., Gouy M., Groussin M., Rochette N.C., ... and Dutheil J.Y. , 2013, Bio++: efficient extensible libraries and tools for computational molecular evolution, Molecular biology and evolution, 30(8): 1745-1750.
https://doi.org/10.1093/molbev/mst097

Hackl H., Charoentong P., Finotello F., and Trajanoski Z., 2016, Computational genomics tools for dissecting tumour–immune cell interactions, Nature Reviews Genetics, 17(8): 441-458.
https://doi.org/10.1038/nrg.2016.67

Hong L., Zhu W.J., and Wang J.D., 2006, Discussion on the Hierarchical Structure of Molecular Computers and Basic Concepts of Molecular Computing, Nanjing Youdian Daxue Xuebao (Journal of Nanjing University of Posts and Telecommunications: Natural Science), 26(6): 43-47.

Koeppel J., Weller J., Peets E.M., Pallaseni A., Kuzmin I., Raudvere U., Peterson H., Liberante F.G., and Parts L., 2023, Prediction of prime editing insertion efficiencies using sequence features and DNA repair determinants, Nature Biotechnology, 1-11.
https://doi.org/10.1038/s41587-023-01678-y

Lambert S.A., Jolma A., Campitelli L.F., Das P.K., Yin Y., Albu M., Chen X., Taipale J., Hughes,T.R., and Weirauch M.T., 2018, The human transcription factors, Cell, 172(4): 650-665.
https://doi.org/10.1016/j.cell.2018.01.029

Li J., Wu X.J., Liu C.X., and Yuan Y.J., 2006, Application of New Data Processing Methods in Metabolomics Research, Yaoxue Xuebao (Acta Pharmaceutica Sinica), 41(1): 47-53.

Ma J.J., Wang X.C., and Nie X.J., 2021, Advances in the Application of Bioinformatics in Proteomics Research, Shengwu Xinxixue (Bioinformatics), 19(2): 85-91.

Marx V., 2022, Method of the Year: protein structure prediction, Nature methods, 19(1): 5-10.
https://doi.org/10.1038/s41592-021-01359-1

Pellegrini M., Emes R.D., Pirooznia M., and Zou Q., 2022, Insights in Computational Genomics: 2022, Frontiers in Genetics, 14: 1256011.
https://doi.org/10.3389/fgene.2023.1256011

Pevzner P., 2000, Computational molecular biology: an algorithmic approach, MIT press.
https://doi.org/10.7551/mitpress/2022.001.0001

Ramalingam A., Kudapa H., Pazhamala L.T., Weckwerth W., and Varshney R.K., 2015, Proteomics and metabolomics: two emerging areas for legume improvement, Frontiers in Plant Science, 6: 1116.
https://doi.org/10.3389/fpls.2015.01116

Snow C.D., Nguyen H., Pande V.S., and Gruebele M., 2002, Absolute comparison of simulated and experimental protein-folding dynamics, Nature, 420(6911): 102-106.
https://doi.org/10.1038/nature01160

Swanson K., Wu E., Zhang A., Alizadeh A.A., and Zou J., 2023, From patterns to patients: Advances in clinical machine learning for cancer diagnosis, prognosis, and treatment, Cell.
https://doi.org/10.1016/j.cell.2023.01.035

Taylor M.J., Lukowski J.K., and Anderton C.R., 2021, Spatially resolved mass spectrometry at the single cell: recent innovations in proteomics and metabolomics, Journal of the American Society for Mass Spectrometry, 32(4): 872-894.
https://doi.org/10.1021/jasms.0c00439

Van de Sande B., Flerin C., Davie K., De Waegeneer M., Hulselmans G., Aibar S., Seurinck R., Saelens W., Cannoodt R., Rouchon Q., Verbeiren T., De Maeyer D., Reumers J., Saeys Y., and Aerts S., 2020, A scalable SCENIC workflow for single-cell gene regulatory network analysis, Nature Protocols, 15(7): 2247-2276.
https://doi.org/10.1038/s41596-020-0336-2

Wang L., Kainulainen M.H., Jiang N., Di H., Bonenfant G., Mills L., Currier M., Shrivastava-Ranjan P., Calderon B.M., Sheth M., Mann B.R., Hossain J., Lin X., Lester S., Pusch E.A., Jones J., Cui D., Chatterjee P., Jenks M.H., Morantz E.K., Larson G.P., Hatta M., Harcourt J.L., Tamin A., and Zhou B., 2022, Differential neutralization and inhibition of SARS-CoV-2 variants by antibodies elicited by COVID-19 mRNA vaccines, Nature communications, 13(1): 4350.
https://doi.org/10.1101/2021.11.24.469906

Yu Y., Miao H.K., Song S.P., and Fan C.H., 2008, Research Progress on DNA Molecular Computing and DNA Computers, Kexue Tongbao (Chinese Science Bulletin), 53(5): 497-502.
https://doi.org/10.1360/csb2008-53-5-497

Zhang C.T., 2000, The Current Situation and Prospect of Bioinformatics, Shijie Keji Yanjiu Yu Fazhan (World Science and Technology Research and Development), 22(6): 17-20.

Zhang L., Li Z., Zheng F.B., and Yang W., 2017, Eight Classes of Protein Secondary Structure Prediction Algorithm Based on Deep Learning, Jisuanji Yingyong (Computer Applications), 37(5): 1512-1515.

Zhang Y., and Min X.J., 2012, Computational Molecular Biology: Integration of Experimental Molecular and Genomic Biology with Computational Techniques,

Computational Molecular Biology, 1.

https://doi.org/10.5376/2011.cmb.01.0001