Check for updates

(REVIEW ARTICLE)

Meaning, concept and application of bioinformatics

Bhaskar Mahanayak *

Department of Zoology, Berhampore Girls' College, Berhampore, Murshidabad, India.

Abstract

Bioinformatics, originating from the term coined in 1979 by Paulien Hogeweg and Ben Hesper, is a burgeoning interdisciplinary field at the nexus of biology, computer science, and information technology. It encompasses the collection, classification, storage, and analysis of vast biological datasets using computational methods, aiming to unravel the complexities of biological systems. This abstract provides a comprehensive overview of bioinformatics, emphasizing its definition, scope, branches, aims, research areas, databases, and applications, particularly highlighting its role in drug discovery through structural bioinformatics. Bioinformatics integrates diverse branches such as computational biology, genomics, proteomics, structural biology, systems biology, pharmacogenomics, and bioprogramming. Each branch applies mathematical modeling, molecular interactions analysis, or genomic data interpretation to advance understanding in biology and medicine. Key aims include applying advanced computational technologies to biological problems, presenting complex data clearly, and providing robust statistical tools for genomic analysis. The field also fosters collaborations across academic, commercial, and government sectors to leverage resources and expertise. Research areas span computational evolutionary analysis, genome annotation, gene expression analysis, cancer mutation studies, and comparative genomics, illuminating disease mechanisms and evolutionary relationships. Bioinformatics databases play a crucial role by storing and organizing biological data. Primary repositories like GenBank and the Protein Data Bank (PDB), secondary databases such as SWISS-Prot, and specialized databases support research and clinical applications by providing access to genomic, proteomic, and structural information. In drug discovery, structural bioinformatics techniques like homology modeling, molecular docking, and simulations predict protein structures, analyze interactions, and aid in designing therapies. These computational methods accelerate drug development, optimize drug-target interactions, and reduce costs associated with bringing new drugs to market. Overall, bioinformatics drives innovation in biological sciences and healthcare by leveraging computational methods to interpret biological data. As technology advances, bioinformatics will continue to play a pivotal role in addressing global health challenges, advancing personalized medicine, and facilitating groundbreaking discoveries in biology and medicine. The evolving field of bioinformatics promises continued growth and impact, shaping the future of biological research and its applications in improving human health and understanding the natural world.

Keywords: Bioinformatics; Computational biology; Structural bioinformatics; Drug discovery; Genomics

1. Introduction

The term Bioinformatics was coined by Paulien Hogeweg and Ben Hesper in 1979 for the study of informatic processes in biotic systems (Hogeweg & Hesper, 1979). It is the science of collecting and analyzing complex biological data. It involves the collection, classification, storage, and analysis of biochemical and biological information using computers, making it an interdisciplinary field that develops methods and software tools for understanding biological data (National Center for Biotechnology Information [NCBI], 2001).

* Corresponding author: Bhaskar Mahanayak Orcid id: https://orcid.org/0000-0002-1933-0638

Bioinformatics involves the application of computation to the field of biology, including data management, algorithm development, and data mining (Attwood & Parry-Smith, 1999). An algorithm is a process or set of rules to be followed in calculations or other problem-solving operations, especially by a computer. Data mining is the practice of examining large pre-existing databases to generate new information (Engelhardt & Jordan, 2005).

According to the NCBI (2001), bioinformatics is "the field of science in which biology, computer science, and information technology merge into a single discipline." The three important sub-disciplines within bioinformatics are the development of new algorithms and statistics to assess relationships among large data sets; the analysis and interpretation of various types of data, including nucleotide and amino acid sequences, protein domains, and protein structures; and the development and implementation of tools that enable efficient access and management of different types of information.

2. Branches of Bioinformatics

- **Computational Biology:** Involves the development and application of data-analytical and theoretical methods, mathematical modeling, and computational simulation techniques to study biological, ecological, behavioral, and social systems (Chou & Elrod, 2010).
- **Genomics:** Concerned with the structure, function, evolution, and mapping of genomes (Attwood & Parry-Smith, 1999).
- **Proteomics:** The study of the proteome, the complete set of proteins expressed by an organism, tissue, or cell. This includes the study of changes in protein expression patterns related to diseases and environmental conditions (Rost, 2001).
- **Bioprogramming:** Refers to a hypothetical program in the brain responsible for constructing a language from words (National Center for Biotechnology Information [NCBI], 2001).
- **Structural Biology:** The study of the molecular structure and dynamics of biological macromolecules, particularly proteins and nucleic acids, and how alterations in their structures affect their function (Berman et al., 2000).
- **Systems Biology:** The study of biological systems considering the interactions of key elements such as DNA, RNA, proteins, and cells with one another, often integrated by computer analysis (Engelhardt & Jordan, 2005).
- **Pharmacogenomics:** The study of how genes affect a person's response to drugs (Rost, 2001).

Aim of Bioinformatics

The aims of bioinformatics include

- Application of computer technology to biological problems (Attwood & Parry-Smith, 1999).
- Presentation of complex results in an understandable form (Engelhardt & Jordan, 2005).
- Providing access to powerful statistical and graphical tools for the analysis of genomic data (Berman et al., 2000).
- Encompassing the use of algorithmic tools to facilitate biological analysis (Chou & Elrod, 2010).
- Facilitating the integration of biological metadata in the analysis of experimental data (Rost, 2001).
- Establishing strategic partnerships with commercial, academic, and other government organizations engaged in bioinformatics research and development (Attwood & Parry-Smith, 1999).
- Providing at-a-glance availability of bioinformatics tools and databases through the development of software modules (Engelhardt & Jordan, 2005).

3. Scope of Bioinformatics

Bioinformatics is broadly defined as the study of the inherent structure of biological information (NCBI, 2001). It merges biology and information sciences, with current research including the analysis of gene and protein sequences to reveal protein evolution and alternative splicing, the development of computational approaches to study and predict protein structure, and the analysis of mass spectrometry data to understand the connection between phosphorylation and cancer (Rost, 2001). Graduates in bioinformatics can engage in research, teaching, clinical service, and consultation, with growing demand in universities, research centers, biotechnology, and pharmaceutical companies (Benson et al., 2011).

4. Research Areas of Bioinformatics

- **Computational Evolutionary Analysis:** Enables researchers to trace the evolution of organisms by measuring changes in their DNA (Xie, Bourne, & Zhou, 2005).
- **Nucleotide and Protein Sequence Analysis:** Facilitates large-scale sequencing of nucleotide and protein data, which has significant applications in biology (Benson et al., 2011).
- **Genome Annotation:** Involves marking genes and other biological features in a DNA sequence, often requiring automation due to the large size of genomes (Ashburner et al., 2000).
- **Analysis of Gene Expression:** Determines the expression of many genes by measuring mRNA levels using techniques like microarrays and RNA-Seq (Xie, Bourne, & Zhou, 2005).
- **Analysis of Mutations in Cancer:** Identifying specific mutations aids in proper diagnosis and treatment of cancer (Chou & Elrod, 2010).
- **Comparative Genomics:** Establishes correspondence between genes in different organisms to trace evolutionary processes responsible for genomic divergence (Engelhardt & Jordan, 2005).

5. Bioinformatics Databases

Bioinformatics databases are computerized archives used to store and organize data for easy retrieval. They include:

- **Primary Databases:** Contain original biological data, such as GenBank and Protein Data Bank (PDB) (Benson et al., 2011).
- **Secondary Databases:** Contain computationally processed or manually curated information based on primary databases, like SWISS-Prot and Protein Information Resources (PIR) (UniProt Consortium, 2019).
- **Specialized Databases:** Cater to specific research interests, such as FlyBase and the HIV sequence database (NCBI, 2001).

6. Structural Bioinformatics in Drug Discovery

Structural bioinformatics leverages computational methods to predict the 3D structure of proteins, aiding in drug discovery by modeling protein-ligand interactions and facilitating the design of new therapies. Homology modeling, molecular docking, and simulations are key techniques in this process (Sali & Blundell, 1993). Tools like MODELLER, SWISS-MODEL, and TASSER are used to predict protein structures and analyze their interactions with potential drug candidates (Marti-Renom et al., 2000). Visualization tools such as UCSF Chimera and PyMOL help analyze and refine these structures (Zhang, 2008). Databases documenting drug-target interactions, like SwissDock and SwissTargetPrediction, support drug development by providing insights into protein-ligand interactions (Kitchen et al., 2004).

7. Conclusion

Bioinformatics is a dynamic field integrating biology, computer science, and information technology to analyze large biological datasets. It aims to develop tools and algorithms for understanding biological data, with applications in various branches such as genomics, proteomics, and drug discovery. As technological advances continue, bioinformatics is poised to play a critical role in addressing global health challenges and advancing personalized medicine.

Compliance with ethical standards

Disclosure of conflict of interest

The Authors proclaim no conflict of interest.

References

[1] Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., ... & Sherlock, G. (2000). Gene Ontology: tool for the unification of biology. Nature Genetics, 25(1), 25-29.

[2] Attwood, T. K., & Parry-Smith, D. J. (1999). Introduction to Bioinformatics. Pearson Education.

[3]   Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2011). GenBank. Nucleic Acids Research, 39(Database issue), D32-D37.

[4]   Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., ... & Bourne, P. E. (2000). The Protein Data Bank. Nucleic Acids Research, 28(1), 235-242.

[5]   Chou, K. C., & Elrod, D. W. (2010). Bioinformatics and its applications. Annual Review of Biophysics and Biomolecular Structure, 29(1), 239-263.

[6]   Engelhardt, B. E., & Jordan, M. I. (2005). Analysis of microarray gene expression data. Statistical Methods in Medical Research, 14(3), 321-336.

[7]   Hogeweg, P., & Hesper, B. (1979). Interactive instruction on population interactions. Computers & Education, 3(2), 111-119.

[8]   Kitchen, D. B., Decornez, H., Furr, J. R., &Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: methods and applications. Nature Reviews Drug Discovery, 3(11), 935-949.

[9]   Marti-Renom, M. A., Stuart, A. C., Fiser, A., Sánchez, R., Melo, F., & Sali, A. (2000). Comparative protein structure modeling of genes and genomes. Annual Review of Biophysics and Biomolecular Structure, 29(1), 291-325.

[10]   National Center for Biotechnology Information (NCBI). (2001). Overview of Bioinformatics. Retrieved from https://www.ncbi.nlm.nih.gov

[11]   Rost, B. (2001). Review: Protein secondary structure prediction continues to rise. Journal of Structural Biology, 134(2-3), 204-218.

[12]   Sali, A., & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. Journal of Molecular Biology, 234(3), 779-815.

[13]   UniProt Consortium. (2019). UniProt: a worldwide hub of protein knowledge. Nucleic Acids Research, 47(D1), D506-D515.

[14]   Xie, L., Bourne, P. E., & Zhou, H. (2005). Bioinformatics in the post-genomic era. Annual Review of Genomics and Human Genetics, 6(1), 403-425.

[15]   Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. BMC Bioinformatics, 9(1), 40.