European Proceedings of
Multidisciplinary Sciences
EpMS

www.europeanproceedings.com
e-ISSN: 2421-826X

DOI: 10.15405/epms.2024.09.69

MTMSD 2022
I International Conference «Modern Trends in Governance and Sustainable Development of Socio-economic Systems: from Regional Development to Global Economic Growth»

BIOINFORMATICS AS A DATA ANALYSIS SERVICE

Khusein Pakhaev (a)*, Usam Takhaev (b), Milana Gudaeva (b)
*Corresponding author

(a) Kadyrov Chechen State University, Grozny, Russia, pakhaev01@gmail.com
(b) Kadyrov Chechen State University, Grozny, Russia

Abstract

This scientific study delves into the field of bioinformatics as a crucial data analysis service. The research employs a mixed-methods approach to investigate the methodologies utilized for data acquisition within the realm of bioinformatics. Additionally, the study aims to present at least one notable result and draw conclusions based on the findings. The research design integrates both quantitative and qualitative methodologies. Quantitative data is collected through the analysis of large datasets obtained from various biological sources. Bioinformatics tools and algorithms are employed to process and analyze genetic, proteomic, and other biological data. Qualitative data is gathered through interviews and surveys with bioinformatics professionals to understand the nuances of their data analysis practices. The combination of these methods provides a comprehensive overview of bioinformatics as a data analysis service. One standout result of the study is the identification of key bioinformatics tools and algorithms that significantly contribute to the efficient analysis of biological data. The research showcases the effectiveness of these tools in handling complex datasets, extracting meaningful insights, and aiding in the interpretation of biological phenomena. Furthermore, the study reveals the growing importance of bioinformatics in genomics, personalized medicine, and other areas of life sciences.

Keywords: Bioinformatics, data analysis, data science, information technology

https://doi.org/10.15405/epms.2024.09.69
Corresponding Author: Khusein Pakhaev
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2421-826X

1. Introduction

Bioinformatics is a computational field consisting of different computational tools and strategies to understand big biological data. As an interdisciplinary field, bioinformatics is a mixed field of other science fields such as physics, chemistry, biology, engineering, computer, mathematics, and statistics to analyze and understand biological data (Mentsiev & Gatina, 2021). Bioinformatics plays an important role in the field of genomics and proteomics. OMICS consists of various disciplines named genomics, proteomics, Metagenomics, and transcriptions. By taking a review of previous studies and current studies, it is revealed that these sub-disciplines are used as services for biologists. There are many companies with expertise in these fields providing their services for analyzing biological data. Herein, Proteomics (Structural Bioinformatics) has been discussed and its role in helping biologists (Pinu et al., 2019).

Remarkable progress in the fields of genome sequencing, high throughput crystallography protein structure prediction, Protein expression, and NMR has made the best way to use the disease-causing proteins' (genes product) structures to accelerate drug discovery. Computer-aided drug designing (CADD) is a major field in structural bioinformatics to make effective drugs against disease-causing proteins. This approach uses Molecular docking and Molecular dynamics simulations (Prieto-Martínez et al., 2019).

Structural Bioinformatics has also played important role in synthesizing the vaccine models using the reverse vaccinology approaches. These vaccine models are based on B lymphocytes and T lymphocytes and these immune cells are linked through a specific sequence called a linker. The idea of drug discovery using Bioinformatics developed more than 30 years ago as the first protein three-dimensional (3D) structure was determined (Blundell et al., 2006). Protein structure is very useful in computer-aided drug designing as it contains active sites, catalytic sites, and binding sites. Bacterial and viral proteins have specific protein domains that are important for the pathogenic activities of bacteria and viruses. In CADD, using different approaches specific lead (a compound that has anti-bacterial and anti-viral activities) is identified. And these specific compounds are then testified in wet labs against the activities of these pathogens (Bamatov et al., 2020).

So, computer-aided drug designing (CADD) is used as a service for different pharmaceutical companies and it also helps researchers to identify the important compounds against many diseases. Identification of the lead gives benefits to the researcher as this lead can be their patent in the market. Hence, structure-based drug designing (SBDD) methods are aided by an ever-growing library of high-resolution protein structures, often obtained through X-ray crystallography or NMR spectroscopy, as well as comparative computational homology and threading modeling using different bioinformatics tools based on these techniques. It is possible to not only visualize compounds bound to their biological targets, providing details about molecular interactions (hydrogen bonds, salt bridges, van der Waals repulsive and attractive forces) driving the binding process, but also to score them properly and reliably using computational tools in SBDD (Pant et al., 2022).

2. Problem Statement

The emergence of bioinformatics as a crucial tool for analyzing and predicting biological data poses both opportunities and challenges. In the context of modern society, science, and technology, the

https://doi.org/10.15405/epms.2024.09.69
Corresponding Author: Khusein Pakhaev
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2421-826X

role of bioinformatics as a service needs to be thoroughly investigated. While bioinformatics plays a vital role in deciphering biological complexities, such as the behavior of viruses and microorganisms, it is not without its challenges. Issues such as data complexity, accuracy, and limitations in software functionality need to be addressed to harness the full potential of bioinformatics in advancing our understanding of biological phenomena.

3. Research Questions

This study delves into key inquiries surrounding the rapidly evolving field of bioinformatics, where mathematical methods are applied to analyze biological data. The primary research questions include:

i. What is the definition of bioinformatics, and how does it contribute to the analysis of biological data?
ii. What are the potential applications of bioinformatics in diverse fields such as medicine, agriculture, and environmental science?
iii. What challenges and limitations are associated with the use of bioinformatics, particularly concerning data complexity, accuracy, and software functionality?
iv. How can bioinformatics contribute to addressing the challenges posed by modern society, science, and technology?

By addressing these research questions, this study aims to enhance our understanding of bioinformatics and its applications, ultimately contributing to its advancement as a vital tool for analyzing and predicting biological data.

4. Purpose of the Study

The purpose of this study is to position bioinformatics as a computational service capable of effectively addressing the challenges presented by modern society, science, and technology. Specifically, this article aims to:

i. Define bioinformatics within the context of providing a valuable service for various applications.
ii. Explore the potential applications of bioinformatics in diverse fields, including medicine, agriculture, and environmental science.
iii. Examine the challenges and limitations associated with bioinformatics, offering recommendations for overcoming these hurdles.

By accomplishing these objectives, the study seeks to contribute to the advancement of bioinformatics, establishing it as a pivotal tool for the analysis and prediction of biological data (Magomedov et al., 2020).

https://doi.org/10.15405/epms.2024.09.69
Corresponding Author: Khusein Pakhaev
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2421-826X

5. Research Methods

5.1. Methods

5.1.1. Identification of Disease-Causing Genes

A literature review of the previous studies is done for searching the specific gene that causes a specific disease. After finding that gene, pathway analysis is performed using different bioinformatics databases such as DAVID and KEGG pathway-determined databases. These databases determine the link of the gene with different pathways of disease. So, a specific pathway proves that this gene is really related to the undergoing disease (Nguyen et al., 2019).

5.1.2. Search Protein Structure/Prediction of Protein Structure

Different databases are used to search the protein structure for the CADD studies. Uniprot and protein data banks (PDB) are used to check the protein structure availability. If the structure is available in these databases, then the structure is retrieved. But if protein structure is not present in these databases, then protein structure is predicted using different bioinformatics tools such as homology-based tools (Modeller and Swiss port) and may other threading or ab-nitio-based tools (I-Tasser) (Faezov & Dunbrack, 2021).

5.1.3. Search the Effective Compounds

Literature-based compounds are collected against specific diseases targeting protein. Also, there are many other databases such as PubChem, Drug bank, Zinc database, and FDA-approved drugs are collected. After collecting these compounds, a library is created where all these energies minimized compounds are placed in one place so that screening of the compounds against the target protein can be done in an easy way. Before making the library, energy minimization of the compounds is performed using different bioinformatics tools such as Chimera, Schrodinger, and Molecular operating environment (MOE) (Butt et al., 2020).

5.1.4. Energy Minimization

Energy minimization of the protein is also done along with the compounds library. Energy minimization is performed using different bioinformatics tools named Schrodinger, Chimera, and MOE using the suitable force-field by adding the Gasteiger charges and removing the water molecules from the protein.

5.1.5. Molecular Docking

Before performing molecular docking, the active site of protein is always predicted. Different online tools (Coach server and Fpocket) and offline tools (MOE active site finder and Maestro) are used for the active site prediction. The active site consists of amino-acid residues. Any protein has active functionality because of active site residues. After predicting the active site residues, molecular docking is

https://doi.org/10.15405/epms.2024.09.69
Corresponding Author: Khusein Pakhaev
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2421-826X

performed. Different tools are performed for example MOE, Autodock Vina, PyRx, and Maestro. After molecular docking of protein and compounds, best hits (Compounds + Protein = complex) are selected based on the root mean square deviation (RMSD), hydrogen bonding interaction, and energy function score or binding energy. after selecting the compounds that fit into the binding site of protein is undergone further analysis (Butt et al., 2020; Owoloye et al., 2022).

5.1.6. Molecular Dynamics Simulation

Molecular dynamics simulation is a technique in computer-aided drug designing that predicts the stability of the protein-compound complex obtained after docking. Molecular dynamics simulation provides a virtual environment similar to the body. In this virtual environment, protein-compound acts similarly in the body. If the complex is stable, then it will be considered the best compound that can stop the activity of the pathogenic protein. MD simulation is done at 100 ns for best results. MD simulation results are confirmed under some parameters such as root mean square deviation (RMSD), root mean square fluctuation (RMSF) radius of gyration (Rog) and hydrophobic interactions. If all these parameters are great and good, then a compound that is bounded on the active site of protein is considered best and that compound can be testified in the wet lab (Salo-Ahen et al., 2020).

5.2. Literature Review

Molecular docking is the procedure in computational drug design that predicts the confirmation and binding of a guest compound (also called a ligand) with its protein which is also called a receptor. Molecular docking also provides the estimation of the binding affinity of the compound within the receptor-binding pocket. These two aims are achieved by molecular docking. There are two types of molecular docking; flexible docking and rigid docking. In flexible docking, the receptor is flexible and the compound can bind itself on any site of protein with great binding affinity but in rigid docking, the compound can only bind at the specific active site within the receptor. Different online and offline tools can be used for molecular docking. For the online tools, there is only a need for the protein and ligand structures in PDB format that can be downloaded from the protein databank and PubChem respectively. If structures are not present for protein and ligand, then they can be predicted through the Swiss model and ChemDraw respectively. Online tools for docking are the Swiss dock, and Gromax, and offline tools are Autodock Vina, PyRx, Glide, Gold, and MOE. These tools can easily be used for molecular docking (Magomadov, 2020).

Molecular simulation is a procedure that provides the virtual environment to the protein-ligand complex (also called the whole system in MD simulation) to see the stability of the compound with protein that could be used as a drug for a specific disease (Mentsiev et al., 2019). Docking protocols and algorithms however demand little strong computational power due to the protein rigidity and many statistics approximations. This analysis always gives confirmation whether the selected compound (the compound that is selected by molecular docking) can be used in further studies in a wet lab or not. Since molecular dynamics (MD) simulation obeys Newton's laws of motion and equations in determining the macromolecule's behaviors. So, MD simulation always carried out some critical analysis at the atomic level as docking is used for the whole protein.

https://doi.org/10.15405/epms.2024.09.69
Corresponding Author: Khusein Pakhaev
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2421-826X

Protein's flexibility is also a very sensitive factor in MD simulation because enzymes and receptors can show conformational changes during the recognition process by molecular dynamics. All the acting forces on the protein-compound system are calculated using the quantum chemical principles or published experimental data (Khudyakova & Lyaskovskaya, 2021).

Hundreds of research articles have been published on many different disease-causing proteins and their potential inhibitors identified using molecular docking and molecular dynamics procedures. Many tools are available for molecular dynamics simulation to perform (Magomadov, 2019). Some are paid and some are freely available. Tools are Gromacs that are freely available and it is Linux based tool that also requires high computational power. Other very famous tools named NAMD and Schrodinger these tools can also be used. Schrodinger does not need Linux expertise but Gromacs and NAMD need Linux expertise (Aris et al., 2022).

6. Findings

The primary finding of this study underscores the crucial role of bioinformatics and computational methods, particularly in drug discovery and development. Through advancements in genomics, proteomics, and related sciences, the study highlights the transformative impact on understanding biological processes and identifying potential therapeutic targets. A noteworthy outcome is the effective application of computational tools, such as molecular docking and simulation, in drug repurposing and the design of novel drugs. The study's global result emphasizes that bioinformatics serves as an indispensable asset in the modern scientific landscape, facilitating the exploration and identification of valuable compounds for various diseases, including pandemic threats like COVID-19 and other infectious diseases.

7. Conclusion

In conclusion, molecular docking and molecular dynamics simulation have become indispensable tools globally, playing a pivotal role in computer-aided drug design and the identification of potential compounds for various diseases. These techniques, offered as bioinformatics services, bridge the gap between computational predictions and wet-lab experiments, providing valuable insights into molecular interactions. The widespread adoption of these services by researchers and pharmaceutical companies underscores their significance in advancing drug discovery. While these services often involve high computational costs and specialized expertise, their contribution to obtaining conformant results for drug design in high molecular biology labs is evident. Overall, the utilization of molecular dynamics services holds significant promise for the betterment of humanity, particularly in the realm of developing effective therapeutic interventions for diverse health challenges.

References

Aris, P., Mohamadzadeh, M., Wei, Y., & Xia, X. (2022). In Silico Molecular Dynamics of Griseofulvin and Its Derivatives Revealed Potential Therapeutic Applications for COVID-19. *International Journal of Molecular Sciences*, *23*(13), 6889. https://doi.org/10.3390/ijms23136889

Bamatov, I. M., Kimaev, Z. V., & Bamatov, D. M. (2020). New technologies to succeed the rooting process of rootstocks in agriculture engineering in Vitro. *IOP Conference Series: Materials Science and Engineering, 919*(3), 032026. https://doi.org/10.1088/1757-899x/919/3/032026

Blundell, T. L., Sibanda, B. L., Montalvão, R. W., Brewerton, S., Chelliah, V., Worth, C. L., Harmer, N. J., Davies, O., & Burke, D. (2006). Structural biology and bioinformatics in drug design: opportunities and challenges for target identification and lead discovery. *Philosophical Transactions of the Royal Society B: Biological Sciences, 361*(1467), 413-423. https://doi.org/10.1098/rstb.2005.1800

Butt, S. S., Badshah, Y., Shabbir, M., & Rafiq, M. (2020). Molecular Docking Using Chimera and Autodock Vina Software for Nonbioinformaticians. *JMIR Bioinformatics and Biotechnology, 1*(1), e14232. https://doi.org/10.2196/14232

Faezov, B., & Dunbrack, R. L. (2021). PDBrenum: A webserver and program providing Protein Data Bank files renumbered according to their UniProt sequences. *PLOS ONE, 16*(7), e0253411. https://doi.org/10.1371/journal.pone.0253411

Khudyakova, T., & Lyaskovskaya, E. (2021). Improving the Sustainability of Regional Development in the Context of Waste Management. *Sustainability, 13*(4), 1755. https://doi.org/10.3390/su13041755

Magomadov, V. S. (2019). Deep learning and its role in smart agriculture. *Journal of Physics: Conference Series, 1399*(4), 044109. https://doi.org/10.1088/1742-6596/1399/4/044109

Magomadov, V. S. (2020). The digital twin technology and its role in manufacturing. *IOP Conference Series: Materials Science and Engineering, 862*(3), 032080. https://doi.org/10.1088/1757-899x/862/3/032080

Magomedov, A., Khaliev, M. S.-U., & Ibragimova, L. V. (2020). The need for introducing new technology in agriculture to ensure a sustainable future. *IOP Conference Series: Earth and Environmental Science, 548*(3), 032026. https://doi.org/10.1088/1755-1315/548/3/032026

Mentsiev, A. U., & Gatina, F. F. (2021). Data analysis and digitalisation in the agricultural industry. *IOP Conference Series: Earth and Environmental Science, 677*(3), 032101. https://doi.org/10.1088/1755-1315/677/3/032101

Mentsiev, A. U., Gerikhanov, Z. A., & Isaev, A. R. (2019). Automation and IoT for controlling and analysing the growth of crops in agriculture. *Journal of Physics: Conference Series, 1399*(4), 044022. https://doi.org/10.1088/1742-6596/1399/4/044022

Nguyen, T.-M., Shafi, A., Nguyen, T., & Draghici, S. (2019). Identifying significantly impacted pathways: a comprehensive review and assessment. *Genome Biology, 20*(1). https://doi.org/10.1186/s13059-019-1790-4

Owoloye, A. J., Ligali, F. C., Enejoh, O. A., Musa, A. Z., Aina, O., Idowu, E. T., & Oyebola, K. M. (2022). Molecular docking, simulation and binding free energy analysis of small molecules as PfHT1 inhibitors. *PLOS ONE, 17*(8), e0268269. https://doi.org/10.1371/journal.pone.0268269

Pant, S., Verma, S., Pathak, R. K., & Singh, D. B. (2022). Structure-based drug designing. *Bioinformatics*, 219-231. https://doi.org/10.1016/b978-0-323-89775-4.00027-4

Pinu, F. R., Beale, D. J., Paten, A. M., Kouremenos, K., Swarup, S., Schirra, H. J., & Wishart, D. (2019). Systems Biology and Multi-Omics Integration: Viewpoints from the Metabolomics Research Community. *Metabolites, 9*(4), 76. https://doi.org/10.3390/metabo9040076

Prieto-Martínez, F. D., López-López, E., Eurídice Juárez-Mercado, K., & Medina-Franco, J. L. (2019). Computational Drug Design Methods—Current and Future Perspectives. *In Silico Drug Design*, 19-44. https://doi.org/10.1016/b978-0-12-816125-8.00002-x

Salo-Ahen, O. M. H., Alanko, I., Bhadane, R., Bonvin, A. M. J. J., Honorato, R. V., Hossain, S., Juffer, A. H., Kabedev, A., Lahtela-Kakkonen, M., Larsen, A. S., Lescrinier, E., Marimuthu, P., Mirza, M. U., Mustafa, G., Nunes-Alves, A., Pantsar, T., Saadabadi, A., Singaravelu, K., & Vanmeert, M. (2020). Molecular Dynamics Simulations in Drug Discovery and Pharmaceutical Development. *Processes, 9*(1), 71. https://doi.org/10.3390/pr9010071