International Journal of Herbal Medicine
Available online at www.florajournal.com

E-ISSN: 2321-2187
P-ISSN: 2394-0514
www.florajournal.com
IJHM 2023; 11(3): 01-05
Received: 03-02-2023
Accepted: 08-03-2023

P Kavita
Department of Botany, Maitreyi College, University of Delhi, New Delhi, India

Swarnali
Department of Botany, Maitreyi College, University of Delhi, New Delhi, India

Anamika Singh
Department of Botany, Maitreyi College, University of Delhi, New Delhi, India

Corresponding Author:
Anamika Singh
Department of Botany, Maitreyi College, University of Delhi, New Delhi, India

Computational resources, Biological databases and tools of biological analysis

P Kavita, Swarnali and Anamika Singh

DOI: https://doi.org/10.22271/flora.2023.v11.i3a.862

Abstract
Today is the era of bioinformatics as it uses mathematical algorithms to do DNA sequence alignment to protein structure prediction and drug designing making biological research easier. Biological research has generated lots of databases may it be in form of gene sequences, protein sequences or metabolic pathways across various species. These databases are huge and without bioinformatics one can does not make sense out of it. Bioinformatics provides wide range of tools to deal with all the queries raised in biological research and to find a solution to it. In this article we tried to compile all such freely available tools as well as biological databases, one can have ease to access them and have promising prospects in the area of biological research.

Keywords: Biological databases, tools for sequence analysis, protein structure, bioinformatics, Genomic

Introduction
Data is increasing day by day in every field of research, but issues start with its arrangement, future utility, easy access for others and prospects of data in research and its applications. Experimental data collection is very important for further and increasing amount of data generated from different genomic projects has made the use of computer databases a necessity which helps in rapid assimilation. By using different computer languages and programs theses data can be further analyzed and used for research. With the remarkable increase in the results produced by different biological research, the amount of information stored in the databases doubling every 14-15 months [1] presents a huge demand for analysis and interpretation of these data [2]. Bioinformatics organizes data in a way that allows researchers to access existing information and to submit new entries as they are produced, e.g., the Protein Data Bank for 3D macromolecular structures [3, 4]. Its aims to develop tools and resources which help in the analysis of data [5]. Bioinformatics deals with biological information acquisition, processing, storage, distribution, analysis along with data interpretation that uses different tools and techniques of mathematics, computer science, and biology [1]. With the help of bioinformatics, it is easier for biologists to access data from the internet and other fitting websites and easily discovers the composition of any biological molecules such as nucleic acids and proteins. Bioinformatics is having many different branches (Figure 1) and collections of biological sequences information's in different biological databases. The two most important biological sequences databases are protein databases and nucleic acid databases, while the structural databases are separate and having 3D structural information's. The main use of the bioinformatics tools are sequence analysis of DNA and protein with the help of different programs and databases available on the web [2].
Bioinformatics can be used for analysis of gene expression, gene analysis, detection of gene regulation networks, protein structure and its function [2] along with genomic and transcriptomics data [5]. Bioinformatics constitute a wide range of scientific disciplines and genomic analysis [6].
The two main classes of nucleic acids i.e., DNA and RNA are the carrier of genetic information. As DNA double helical structure is well-known structure with defined function, this information is copied passed on to the next generation [7]. One of the most important molecules in living cells is deoxyribonucleic acid (DNA). The genetic material DNA is a polymer composed of monomeric units known as nucleotides.
A nucleotide is made up of a 5-carbon sugar, deoxyribose, a nitrogenous base, and one or more phosphate groups and the phosphate groups acidic, so the name nucleic acid was coined [7].

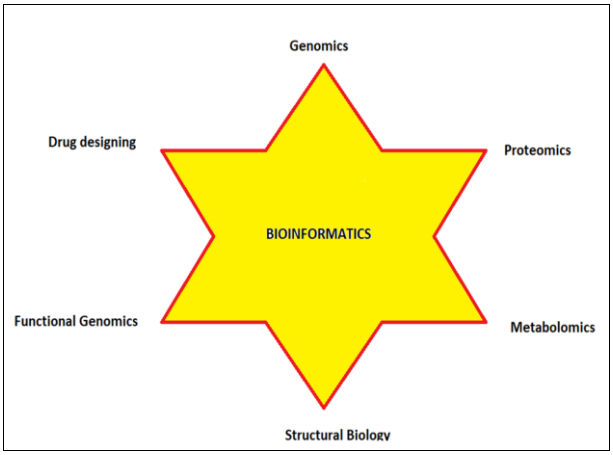

Fig 1: Showing different branches of Bioinformatics

Similarly, RNA is chemically identical to DNA as it is a chain of similar monomers. RNA is further of three types: mRNA, tRNA and rRNA. RNA molecules are required at all stages of protein synthesis. Protein structure is known to have a higher degree of conservation compared to sequences due to large variations in sequence within the protein family which can still result in very similar three-dimensional structures [8]. The structure of any protein molecule helps in determining its function [9]. There are many important such findings and, we can analyze it by using different tools used in bioinformatics. Here we are discussing about some most common biological databases and tools used in sequence analysis of protein and nucleic acids.

Table 1: Data bases for protein

Databases	Use	Links
PROSITE	It is first secondary database having information of domains, functional and all documentation entries.	
Pfam	Database of protein with multiple sequence alignments by using hidden Markov models.	http://pfam.xfam.org/
PIR	Contain comprehensive, annotated and non-redundant protein sequence information.	https://proteininformationresource.org/
TrEMBL	Curated protein sequence database with high level of annotation, computer annotated supplement of Swiss-Prot.	http://www.bioinfo.pte.hu/more/TrEMBL.htm
PDB	It contains3-Dstructuresprotein, DNA, RNA Obtained by NMR and X-Ray crystallography methods.	https://www.rcsb.org/
PRINTS	It contains protein fingerprints that is used to assign family and functional attributes.	http://130.88.97.239/PRINTS/index.php
Swiss-Prot	It contains protein descriptions, domain structure, function, subcellular location and post-translational modifications.	https://www.expasy.org/resources/uniprotkb-swiss-prot

Table 2: Data base for nucleic acid

Databases	Use	Links
DDBJ	DNA Data bank of Japan is a DNA sequences database and member of INSDC.	https://www.ddbj.nig.ac.jp/index-e.html
EMBL	It is European Molecular Biology Laboratory nucleotide sequence data base.	https://www.ebi.ac.uk/
GenBank	Annotated collection of all publicly available DNA sequences also part of INSDC.	https://www.ncbi.nlm.nih.gov/

Table 3: Tool for protein

Name of tools	Use	Websitelink	References
CPH Models	Web server used for predicting 3D structure of protein by using single template homology modeling.	http://www.cbs.dtu.dk/services/CPHmodels/	[10]
ESyPred3D	It is automated homology modeling program alignment performances increased of any new alignment strategy with the help of neutral networks. MODELLER is used finally for 3D structure prediction.	https://www.unamur.be/sciences/biologie/urbm/bioinfo/esypred/	[11]
PHYRE2	Contains collection of tools prediction and analysis of protein structure.	http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index	[12]
SWISSMODEL	Homology server for protein 3D structure prediction.	https://swissmodel.expasy.org/	[13]
I-TASSER	Hierarchical approach based protein structure prediction and structure based function annotation.	https://zhanggroup.org/I-TASSER/	[14]
Robetta	Comparative modeling based protein structure prediction server.	http://robetta.bakerlab.org/	[15]
PEP-FOLD 3	Follows de novo approach for predicts peptide structures prediction.	https://bioserv.rpbs.univ-paris-diderot.fr/services/PEP-FOLD3/	[16]
(PS)2	Automatic server for protein structure prediction by using MODELLER.	http://ps2.life.nctu.edu.tw/	[17]
AS2TSsystem	It is used for protein structure comparison and 3D modelling.	http://proteinmodel.org/AS2TS/AS2TS_MB/index.html	[18]
Raptor X	Used for the calculation of protein secondary, tertiary structure, contact and distance map, solvent accessibility, disorder regions, functional annotation and building site analysis.	http://raptorx.uchicago.edu/	[19]
Modeller	Used for homology or comparative modelling of protein 3-dimensionalstructures.	https://salilab.org/modeller/	[20]
ProtSkin	Used to converts protein sequence alignment in BLAST, CLUSTAL, or MSF based on identity or similarity on to a	https://sbgrid.org/software/titles/protskin	[21]

	protein structure to visualize also help to analyses conserved/non-conserved regions within proteins.		
GeNMR	Generating 3D structure of protein by using amino acid sequence data.	http://www.genmr.ca/index.php	[22]
FATCAT	Used to compare two PDB-form at protein structures.	https://fatcat.godziklab.org/fatcat/fatcat_pair.html	[23]
SuperPose	Calculates protein super position using a modified quaternion approach.	http://superpose.wishartlab.com/	[24]
PSIPRED	PSIPRED used to predict protein structure.	http://bioinf.cs.ucl.ac.uk/psipred/#	[25]
ProteinPeeling	Used to predict protein structure descriptions between protein domain and secondary structure.	https://www.dsimb.inserm.fr/dsimb_tools/peeling/	[26]
InterProSurf	Predicts interactive amino acid residues within a protein that are interactive with other proteins also give 3D structures of subunits of protein complex.	http://curie.utmb.edu/pdbcomplex.html	[27]
MulPBA	Tool used for multiple structures comparison of available protein.	https://www.dsimb.inserm.fr/dsimb_tools/mulpba/	**[28]**
WHATIF	Generate 3D model for a pre-aligned sequence template structure.	https://swift.cmbi.umcn.nl/whatif/	**[29]**
MAPSCI	Multiple alignment of protein structure and consensus identification.	http://www.geom-comp.umn.edu/mapsci/	**[30]**
iPBA	Tool used for comparison of protein structure by sequence alignment methods.	https://www.dsimb.inserm.fr/dsimb_tools/ipba/	[31]
CEP	Predict conformational epitope for protein antigen.	https://www.hsls.pitt.edu/obrc/index.php?page=URL1127484564	**[32]**
SAS	Tool for applying structural information to a given protein sequence.	https://www.ebi.ac.uk/thornton-srv/databases/sas/	**[33]**
ORION	Optimized protein fold recognition is a sensitive method for protein template detection.	https://www.dsimb.inserm.fr/ORION/	**[34]**
Auto-mute	A server for predicting functional amino acid mutation in protein.		[35]
Movie Maker	A web server used to generates small movies of protein dynamics studies.	http://wishart.biology.ualberta.ca/moviemaker/	[36]
DUET	A server that helps in analysis of mutations effects on protein.	http://biosig.unimelb.edu.au/duet/stability	[37]
ClustalOmega	It is a multiple sequence alignment program based on HMM profile-profile techniques that helps to generate alignment between three or more sequences.	https://www.ebi.ac.uk/Tools/msa/clustalo/	[38]

Table 4: Tools for DNA analysis

Name of tool	Use	Link	References
DNAproDB	It is a tool that is used for structural analysis of DNA-protein complexes.	https://dnaprodb.usc.edu/search.html	[39]
RNA fold webserver	Used for the prediction of secondary structures of single stranded RNA or DNA sequences.	http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi	[40]
DSSR	It stands for Dissecting the Spatial Structure of RNA. It is an integrated software tool for the analysis/annotation, model building, and schematic visualization of 3D nucleic acid structures.	http://forum.x3dna.org/rna-structures/	[41]
R-chie	A web server for visualizing cis and trans RNA-RNA, RNA-DNA, DNA-DNA interactions.	https://www.e-rna.org/r-chie/	**[42]**
WebFR3D	It is a suite of programs that is designed to search RNA 3Dstructures.	http://rna.bgsu.edu/webfr3d/fr3d.php	**[43]**
FRASS	A web-server used for comparison of 3D structures of two RNA molecules.	https://sourceforge.net/projects/frass/	**[44]**
SARA	SARA stands for Structure Alignment of Ribonucleic Acids. It is an automated server used for aligning two RNA structures.	https://structure.biofold.org/sara/	[45]
w3DNA	A tool used for analysis, reconstruction and visualization of nucleic acid structures.	http://w3dna.rutgers.edu/	**[46]**
MOLProbity	Validation of macromolecular 3Dstructures for nucleic acid and their complexes.	http://molprobity.biochem.duke.edu/index.php?MolProbSID=eli71k2nfofd1u261asaho69i5&eventID=2	**[47]**
SETTER webserver	Webserver used to caparison of RNA structure.	http://setter.projekty.ms.mff.cuni.cz/	[48]
Ribo Vision	ribosome information viewer open source website	http://apollo.chemistry.gatech.edu/RiboVision/	**[49]**
R3D Align	An application used for detailed nucleotide to nucleotide pairwise alignment of 3D RNA structure based on a method.	http://rna.bgsu.edu/r3dalign/	[50]
RNAviewer	A programme for visualizing the two-dimensional structure of RNA, as well as base pairs and RNA patterns.	http://ndbserver.rutgers.edu/ndbmodule/services/rna_viewer/rnaViewerIndex.html	[51]
SARA	Structure Alignment of Ribonucleic Acids is a fully automated approach for aligning different RNA structures.	https://structure.biofold.org/sara/	[45]
Forna	It is an RNA secondary structure visualization, feature rich and easy to use web based tool which enables the users to easily change sequences or RNA secondary structures for clean, concise and personalized	a.tbi.univie.ac.at/forna/	[52]

	visualization without the need of installing any software.		
Clustal Omega	It is a multiple sequence alignment program which uses HMM profile-profile techniques for generating alignment between more than three sequences.	https://www.ebi.ac.uk/Tools/msa/clustalo/	[38]

Discussion

Bioinformatics is a discipline of biology that grows extensively in last few years. Sequences analysis and identification of new gene, proteins and structure are few important applications of bioinformatics [52]. The most important application is designing of 3D structure of proteins whose structures were not predicted by Nuclear magnetic resonance (NMR) and crystallographic method, due to protein bulky size and other limitations [53]. Genome analysis and sequencing of genome of new varieties is possible only because of extensive computational applications of bioinformatics. In this article we tried to compile most of the resources related to protein and nucleic acids that gives new insight to biological research.

Conclusion

Bioinformatics is a young discipline, which is widely used for analysis of genome, prediction of protein and gene structures, cell modelling, analysis of molecular pathways etc. As per the requirement of these tasks, various tools mentioned here have been successfully curated and provide an easy access and analysis platform for the study of complex genomes. These tools can be used for various tasks like retrieval of structures, prediction and formation of new structures, comparison of different structures etc. that could be helpful for research of a new macromolecule. All these tools are easy to use and most of them are freely available to access.

References

1. Benton D. Bioinformatics principles and potential of a new multidisciplinary tool. Trends In Biotechnology. 1996;14(8):261-272.
2. Bayat A. Science, medicine, and the future: Bioinformatics. BMJ. 2002;324(7344):1018-1022.
3. Bernstein FC, Koetzle TF, Williams GJ, Meyer EF, Jr Brice MD, Rodgers JR, *et al*. The Protein Data Bank. A computer-based archival file for macromolecular structures. European journal of biochemistry. 1977;80(2):319-324.
4. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, *et al*. The Protein Data Bank. Nucleic acids research. 2000;28(1):235-242.
5. Jorge NA, Ferreira CG, Passetti F. Bioinformatics of Cancer ncRNA in High Throughput Sequencing: Present State and Challenges. Frontiers in genetics. 2012;3:287.
6. Zagursky RJ, Russel D. Bioinformatics: Use in Bacterial Vaccine Discovery. Bio techniques. 2001;31:636-659.
7. Minchin S, Lodge J. Understanding biochemistry: Structure and function of nucleic acids. Essays in Biochemistry. 2019;63(4):433-456.
8. Al-Karadaghi S, Hansson M, Nikonov S, Jönsson B, Hederstedt L. Crystal structure of ferrochelatase: the terminal enzyme in heme biosynthesis. Structure. 1997;5(11):1501-1510.
9. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. The Shape and Structure of Proteins. Ncbi.nlm.nih.gov 2000.
10. Nielsen M, Lundegaard C, Lund O, Petersen TN. CPHmodels-3.0--remote homology modeling using structure-guided sequence profiles. Nucleic acids research. 2010;38:W576–W581.
11. Lambert C, Leonard N, De Bolle X, Depiereux E. ESyPred3D: Prediction of proteins 3D structures. Bioinformatics (Oxford, England) 2002;18(9):1250-1256.
12. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. Nature protocols. 2015;10(6):845-858.
13. Guex N, Peitsch MC, Schwede T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. Electrophoresis. 2009;30(1):S162-S173.
14. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. Nature protocols. 2010;5(4):725-738.
15. Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server. Nucleic acids research. 2004;32:W526-W531.
16. Lamiable A, Thévenet P, Rey J, Vavrusa M, Derreumaux P, Tufféry P. PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex. Nucleic acids research 2016.;44(W1):W449-W454.
17. Chen CC, Hwang JK, Yang JM. $(PS)^2$: protein structure prediction server. NAR. 2006;34:W152-W157.
18. Zemla A, Zhou CE, Slezak T, Kuczmarski T, Rama D, Torres C, *et al*. AS2TS system for protein structure modeling and analysis. Nucleic acids research. 2005;33:W111-W115.
19. Källberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, *et al*., JTemplate-based protein structure modeling using the RaptorX web server. Nature Protocols. 2012;7(8):1511-1522.
20. Fiser A, Do RK, Sali A. Modeling of loops in protein structures. Protein science: a publication of the Protein Society. 2000;9(9):1753-1773.
21. Deprez C, Lloubès R, Gavioli M, Marion D, Guerlesquin F, Blanchard L. Solution structure of the E.coli TolA C-terminal domain reveals conformational changes upon binding to the phage g3p N-terminal domain. Journal of molecular biology. 2005;346(4):1047-1057.
22. Berjanskii M, Tang P, Liang J, Cruz JA, Zhou J, Zhou Y, *et al*. GeNMR: a web server for rapid NMR-based protein structure determination. Nucleic acids research. 2009;37:W670-W677.
23. Ye Y, Godzik A. Flexible structure alignment by chaining aligned fragment pairs allowing twists. Bioinformatics. 2003;19(2):ii246-ii255.
24. Maiti R, Van Domselaar GH, Zhang H, Wishart DS. SuperPose: a simple server for sophisticated structural superposition. Nucleic acids research. 2004;32:W590-W594.
25. Buchan DW, Minneci F, Nugent TC, Bryson K, Jones DT. Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic acids research. 2013;41W349-W357.
26. Gelly J.-C, Brevern AG, Hazout S. Protein peeling: An approach for splitting a 3D protein structure into compact fragments. Bioinformatics. 2006;22(2):129-133.
27. Negi S, Schein C, Oezguen N, Power T, Braun W. Inter

ProSurf: A web server for predicting interacting sites on protein surfaces. Bioinformatics 2007;23(24):3397-3399.
28. Léonard S, Joseph A, Srinivasan N, Gelly J, de Brevern A. mulPBA: An efficient multiple protein structure alignment method based on a structural alphabet. Journal of Biomolecular Structure and Dynamics. 2013;32(4):661-668.
29. Vriend G. WHAT IF: A molecular modeling and drug design program. Journal of molecular graphics. 1990;8(1):52-29.
30. Ilinkin I, Ye J, Janardan R. Multiple structure alignment and consensus identification for proteins. BMC Bioinformatics. 2010, 11(1).
31. Gelly JC, Joseph AP, Srinivasan N, de Brevern AG. iPBA: a tool for protein structure comparison using sequence alignment strategies. Nucleic acids research. 2011;39:W18-W23.
32. Kulkarni-Kale U, Bhosle S, Kolaskar AS. CEP: A conformational epitope prediction server. Nucleic Acids Research. 2005;33:W168-W171.
33. Milburn D, Laskowski RA, Thornton JM. Sequences annotated by structure: A tool to facilitate the use of structural information in sequence analysis. Protein engineering. 1998;11(10):855-859.
34. Ghouzam Y, Postic G, Guerin P.-EG, de Brevern AG, Gelly J.-C. ORION: A web server for protein fold recognition and structure prediction using evolutionary hybrid profiles. Scientific Reports. 2016, 6(1).
35. Masso M, Vaisman II. AUTO-MUTE: web-based tools for predicting stability changes in proteins due to single amino acid replacements. Protein engineering, design & selection: PEDS. 2010;23(8):683-687.
36. Maiti R, Van Domselaar GH, Wishart DS. Movie Maker: A web server for rapid rendering of protein motions and interactions. Nucleic Acids Research. 2005;33:W358-W362.
37. Pires DE, Ascher DB, Blundell TL. DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach. Nucleic acids research. 2014;42:W314-W319.
38. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, *et al*. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular systems biology. 2011;7(1):539.
39. Sagendorf JM, Berman HM, Rohs R. DNAproDB: An interactive tool for structural analysis of DNA–protein complexes. Nucleic Acids Research. 2017 45(W1):W89-W97.
40. Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL. The Vienna RNA website. Nucleic acids research. 2008;36:W70-W74.
41. Lu X, Bussemaker H, Olson WDSSR: an integrated software tool for dissecting the spatial structure of RNA. Nucleic Acids Research. 2015, gkv716.
42. Lai D, Proctor JR, Zhu JY, Meyer IMR-CHIE: A web server and R package for visualizing RNA secondary structures. Nucleic acids research. 2012;40(12):e95.
43. Sarver M, Zirbel CL, Stombaugh J, Mokdad A, Leontis NB. FR3D: Finding local and composite recurrent structural motifs in RNA 3D structures. Journal of mathematical biology. 2008;56(1-2):215-252.
44. Kirillova S, Tosatto SC, Carugo O. FRASS: The web-server for RNA structural comparison. BMC bioinformatics. 2010;11:327.
45. Capriotti E, Marti-Renom MA. SARA: A server for function annotation of RNA structures. Nucleic Acids Research. 2009;37(2):W260-W265.
46. Lu XJ, Olson WK. 3DNA: A software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. Nucleic acids research. 2003;31(17):5108-5121.
47. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, *et al*. MolProbity: all-atom structure validation for macromolecular crystallography. Acta crystallographica. Section D, Biological crystallography. 2010;66(1):12-21.
48. CEch P, Svozil D, Hoksza D. SETTER: Web server for RNA structure comparison. Nucleic Acids Research 2012;40(W1):W42-W48.
49. Petrov AS, Bernier CR, Hershkovits E, Xue Y, Waterbury CC, Hsiao C, *et al*. Secondary structure and domain architecture of the 23S and 5S rRNAs. Nucleic acids research. 2013;41(15):7522-7535.
50. Rahrig RR, Petrov AI, Leontis NB, Zirbel CL. R3D Align web server for global nucleotide to nucleotide alignments of RNA 3D structures. Nucleic acids research. 2013;41:W15-W21.
51. Yang H, Jossinet F, Leontis N, Li C, Westbrook J, Berman H, *et al*. Article navigation tools for the automatic identification and classification of RNA base pairs. Nucleic Acids Research. 2003;31(13):3450–3460.
52. Kerpedjiev P, Hammer S, Hofacker IL. Forna (force-directed RNA): Simple and effective online RNA secondary structure diagrams. Bioinformatics. 2015;31(20):3377-3379.