Journal of Current Veterinary Research

ISSN: 2636-4026

Journal home page: http://www.jcvr.journals.ekb.eg

Internal medicine & Infectious disease

Applications of Bioinformatics on Genomics and Proteomics Levels Highlighted in *Brucella* Pathogen

Alyaa Elrashedy[1]; Mohamed Nayel[1]; Akram Salama[1] and Mohamed E. Hasan[2]

(1)Department of Animal Medicine and Infectious Diseases (Infectious Diseases), Faculty of Veterinary Medicine, University of Sadat City, Egypt.
(2) Bioinformatics Department, Genetic Engineering and Biotechnology Research Institute, University of Sadat City, Egypt.

**Corresponding author:* mohamed.aboalez@vet.usc.edu.eg *Received: 25/7/2023 Accepted: 2/8/2023*

ABSTRACT

In a variety of industries, including biotechnology and medicine, bioinformatics is extremely important. It offers an extensive range of uses in the structural and sequencing analysis of biological data. Each application has distinct functions in data interpretation, as database search sequence and alignment are essential tools that are prerequisites for more advanced applications like modelling, epitope predictions, and molecular docking. Genome annotation is thought to be the key to decoding the meaningless sequence data and providing crucial details about the protein-coding genes and genomic characteristics. Additionally, comparative modelling, fold modelling, and *ab initio* modelling are the three main approaches to protein modelling, which is fundamental for various proteomics applications. Epitopes prediction for B cells and T cells using MHC I and MHCII is another significant use. This may aid in the development of novel vaccine candidates for the prevention and management of infectious illnesses. Software can virtually dock the created 3D model or predicted epitopes with ligand to construct a new medicine or vaccination quickly and affordably. Using bioinformatics tools in diagnosis and control of brucellosis has a promising vision to eliminate the disease.

Keywords: Brucellosis, Comparative analysis, Epitopes Prediction, Molecular docking, Protein Modelling.

INTRODUCTION

Recently in the era of bioinformatics and next generation sequencing (NGS); it has become easy to collect and analysis large amount of biological data from various species (Hernández-Domínguez et al., 2019). It includes many disciplines: genomic, proteomic, transcriptomic, and metabolomic that permitted predictions on different levels such as regulation of genes expression, structure, transcription and translation, and mechanisms of action or pathway of proteins. Also it can compute homology, diversion and mutations, as well as evolution relationship that responsible for structure and function alterations all over the time (Oliver et al., 2015). There are numerous applications in bioinformatics on sequence and structural analysis for biological data. In sequence analysis, database search tools, pairwise and multiple sequence alignments (MSA), phylogenetic analysis, genome assembly, genome annotation and comparative genome

analysis are considered fundamental tasks. Likewise, in structure analysis, the main significant tasks are three-dimensional (3D) structures prediction of proteins, epitopes prediction, molecular docking and dynamics simulation (Parthasarathy, 2015). In this review we will discuss some definitions, approaches and applications in bioinformatics that help in handling genomics and proteomics data as well as emphasized these applications on *brucella* pathogen.

Genomic Sequence Analysis

Sequence Database Search, Pairwise and Multiple Sequence Alignment (MSA)

Sequence databases search algorithms must fulfill three features: specificity, sensitivity, and speed. The specificity is the facility to ignore incorrect hits (false positives), the sensitivity has the power to catch several correct hits as possible (true positives), while the speed is the time it takes to search out results from databases (Hauser et al., 2013).The most important heuristic databases search tools are BLAST and FASTA. These algorithms are not ensured to discover all the homologous sequences, but they are 50 –100 times quicker than the dynamic programming techniques (Berger et al., 2021).

Alignment is essential for comparing and discovering the evolutionary process of sequences for organisms. The residues that preserved during the evolutionary process known as "conserved regions" are key structure and have functional roles, while the other residues that mutate frequently are less vital for structure and function (Slodkowicz and Goldman, 2020). DNA and amino acid (aa)sequences comparison with one another to find the identical percent or similarity degree between them is an important step conducted via pairwise sequence alignment (Bayat et al., 2019).

A natural addition of pairwise alignment ismultiple sequence alignment(MSA) that aligns multiple associated sequences to obtain the best matching among these sequences. MSA has the advantage of disclosing more biological information than various pairwise alignments. Many conserved amino acids can be recognized in MSA of proteins. It is also an important requirement to make a phylogenetic analysis of sequence families and prediction of secondary and 3D structures of proteins (Bawono et al., 2017). CLUSTALW server is MSA tool that widely used with default and editable options (Chenna, 2003). The method used in this server is based on firstly, originating a phylogenetic tree from a matrix score that obtained through a quick pairwise alignment algorithm. Secondly, the multiple alignment is completed from a groups of pairwise alignments for sequences clustering, followed by order the branches in the tree (Sofi et al., 2022).

Genome Assembly and Annotation

Genome assembly is a process of producing sequences after fragmentation of chromosomes and the gained sequences are collected. According to the sequencing technology used, the obtained reads have different size range from 20 and 30,000 bases (Zhang et al., 2022). *De novo* assemblers can assemble short reads like that produced from Illumina (50-200bp) into longer contigs followed by scaffolds without using a reference genome (Vaser et al., 2017).

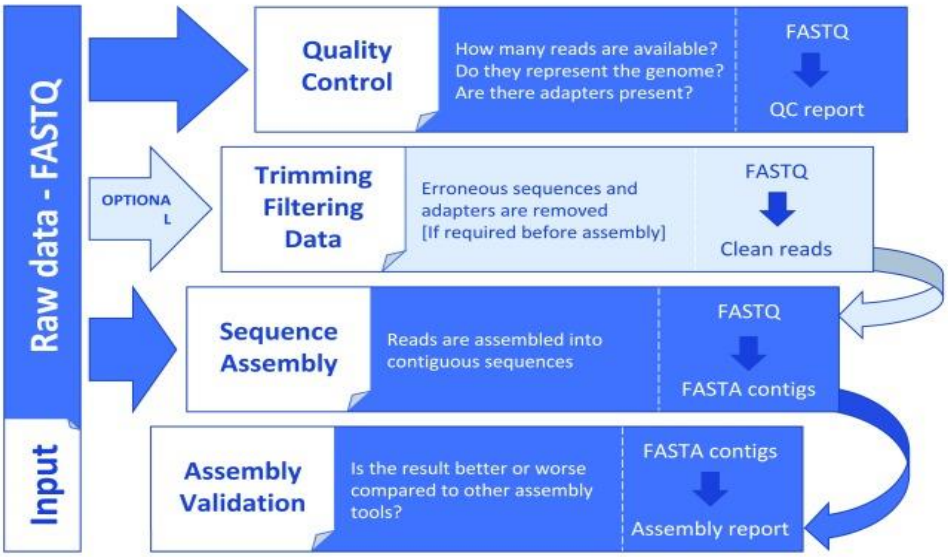


Figure (1): Roadmap of common genome assembly steps (Lantz et al., 2018).

For biologists, rawdata of genomic sequence is without profound value. For that reason, genome annotation is considered the decipher for those meaningless sequences (Ejigu and Jung, 2020). It includes characterizing the most biological significant features in the genomic sequence, and the greatest attention is directed for identifying protein coding genes. The approach of detecting the correct structure and location of the protein coding genes is called gene prediction. In general, genes prediction has three major methods: *ab initio* or intrinsic, extrinsic as well as the combiners (Mishra et al., 2019). The intrinsic method concentrate on data that can be gotten from the genome sequence like the splice site prediction and coding potential, while the extrinsic approach employs resemblance to other types of sequence (RNA or amino acid) as amaterial(Lantz et al., 2018).

Comparative Genomics Analysis

Comparative genomics is an extensive branch of bioinformatics. It studies variations between genomes and decide which one of them is reliable for phenotypical alterations in species (Hu et al., 2011). Comparing to conventional genomics research that emphasis on one genome in study, the comparative genomics gives much supplementary detailed information than that from one analyzed genome (Edwards and Holt, 2013). Additionally, it allows a good interpretation of how species have evolved (Karthik et al., 2021). In this context, pangenome refers to all genes that are found in the analyzed dataset in contrast to core genome is the genes conserved by all analyzed organisms as well as accessory genome is the collection of genes that are exist in 2 or more genomes but not all (Tettelin et al., 2008).

Proteomic Sequence Alignment

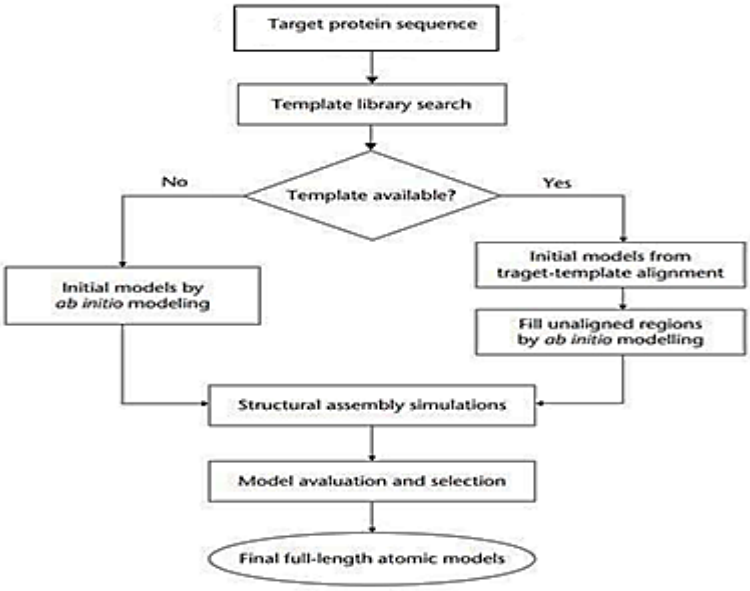
Protein Modeling

Establishing three-dimensional structure of proteins is essential to create new proteins and medications. Nowadays, most of the protein structures are obtained through nuclear magnetic resonance (NMR) spectroscopy, X-ray crystallography, or cryo-EM methods, although the high cost and longtime it requires (Cheung and Yu, 2018). The atomic structure of proteins has been modelled through comparison with identified protein structures “templates”. If the target protein displayed a homology with an existing structure, the modelling become simpler and given model with high

resolution, unless that the homology present, the modeling is created from beginning (Lee et al., 2017). Ab initio modeling is dependent on protein assembly with short peptide fragments (less than 100 aa) (Cheung and Yu, 2018).

The 3D protein structure has provided information at the molecular level, properties, functions, mechanism, design and development ligands, and interaction macro-molecules with proteins (Kuhlman and Bradley, 2019). Proteins are classified at various hierarchical levels such as super family, family, as well as folding structure. All proteins that are grouped into the same family have a high similarity percent. It is recommended that the different families which preserve functions and structure show shared ancestor as well as they clustered into super families. The variance occurred because of the secondary structure or folds that having (Kelm et al., 2014).

Protein structure prediction are categorized into three main lines: comparative modeling (homology), fold recognition (threading) and *de novo* or *ab initio* prediction for a given target protein with an unknown structure (Roy and Zhang, 2012).



Figure(2): A workflow for the commonly used protein modelling approach (Roy and Zhang, 2012).

Comparative Modeling

Comparative modeling (CM) also named "homology", produces information about many numbers of proteins structural. One of the requisites for building a model is the presence of one template at least has significant similarity of aa sequence against the target sequence (Haddad et al., 2020). The main steps mandatory to build a homology-based model summarized (1) searching for the template; (2) alignment between target sequence and template; (3) alignment results help in designing the model; (4) modeling assessment (Muhammed and Aki-Yalcin, 2019).

Root-Mean-Square Deviation (RMSD) is used to model evaluation. It measures the mean distance between the corresponding atoms in two structures after superimposing. When the target proteins have templates with identity less than 50% are public in the protein data bank (PDB), the homologous templates are identified easily during the alignment step. The accuracy of the backbone models created using the CM approach is up to 1-2 A˚ RMSD from the original configuration (Kufareva and Abagyan, 2011). On the other hand, the alignment

accuracy decreases for target proteins with templates whose identity varies between 30 and 50%, although the models contain 85% of their conserved elements with RMSD of 2 to 4 A˚ from the natural structure, with some errors occurring often in the loop areas. Nevertheless, when the identity falls under 30%, the accuracy of modelling by CM sharply declines due to alignment errors and absence of important template hits. For these modelling steps, structural biology knowledge and specialized computer tools are necessary (Trindade Maia et al., 2021).

Protein Threading (Fold Recognition)

Threading (fold recognition) indicates identifying templates for target protein in the PDB with fold or motif similarity. Folding approaches are effective and extensively used due to there are few number of protein folds in native structure because of evolution and also due to limited imposed with the physics and chemistry of polypeptide chains, that verify the resulting protein structure by this technique (Outeiral et al., 2022). Fold recognition is applied by the following steps: (1) Template building. (2) Evaluate the quality of sequences-template alignment by designing of scoring function "the lowest value of the scoring function relates to the optimal alignment". (3) Optimizing the scoring function to attain the best alignment. (4) Choosing the best template for the target sequence "fold recognition or template selection". (5) Building the 3D structure for the target protein concentrate on the alignment (Kumar et al., 2012).

De Novo Or Ab Initio Modeling

Also, is identified as free modelling. In case of lacking templates from the PDB, the models demand to be created from scratch. It is regarded as the type of protein structure prediction that is the most complicated. The modulation phase space of sampling rises as protein size increases, making the *Ab initio* modelling of larger proteins particularly challenging. Understanding the physicochemical theory of how proteins folding in nature is one of its benefits (Kong et al., 2020). At present, *Ab initio* modeling's precision is poor, and it can only be used for tiny proteins. Less than 100 aa until one decade ago. With the advances in pipelines, trRosetta, I-TASSER and QUARK servers generate correct folds for targets with lengths above 100 aa in CASP 11 (Zhang et al., 2016).

General to all *Ab Initio* approaches that: (1) Protein representation and protein conformation space in that representation suitably. (2) Energy functions corresponding to the protein representation. (3) Effective and trustworthy algorithms to explore the conformational space to reduce the energy function, and these conformations are guided to be the structures that the protein is approved at native conditions. The quality of the models is still low and it is challenging to specify which parts of which model are correct. Also, the *Ab-initio* methods are quite challenging to use and need expertise to explain the results into biologically meaningful predictions (Pearce and Zhang, 2021).

Table (1): List of Tools for predicting protein structure that are freely available (Roy and Zhang, 2012).

Name	Web address	Methods[a]
On-line protein structure prediction servers		
I-TASSER	http://zhanglab.ccmb.med.umich.edu/I-TASSER/	TBM + FM
Robetta	http://robetta.bakerlab.org/	FM
ModWeb	https://modbase.compbio.ucsf.edu/scgi/modweb.cgi	TBM
SwissModel	http://swissmodel.expasy.org/	TBM
HHpred	http://hhpred.tuebingen.mpg.de/hhpred	TBM
chunk-TASSER	http://cssb.biology.gatech.edu/skolnick/webservice/chunk-TASSER/index.html	TBM + FM
QUARK	http://zhanglab.ccmb.med.umich.edu/QUARK/	FM
Phyre	http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id = index	TBM
SAM-T08	http://compbio.soe.ucsc.edu/SAM_T08/T08-query.html	TBM
3D-Jury	http://meta.bioinfo.pl	TBM (meta–server)
LOMETS	http://zhanglab.ccmb.med.umich.edu/LOMETS/	TBM (meta–server)
PSIpred	http://bioinf.cs.ucl.ac.uk/psipred/	TBM + SS
Freely downloadable software for protein structure prediction		
Modeller	http://salilab.org/modeller/	TBM
I-TASSER	http://zhanglab.ccmb.med.umich.edu/I-TASSER/download/	TBM + FM
Rosetta	http://www.rosettacommons.org/software/	FM
HHsearch	ftp://toolkit.lmb.uni-muenchen.de/HHsearch/	TBM
Scwrl4	http://dunbrack.fccc.edu/scwrl4/	SC

[a]TBM, template-based modelling; FM, free modelling; SS, secondary structure prediction; SC, Side-chain structure modelling.

Epitopes Prediction

Immunoinformatics is a discipline that assists in establishing considerable information of immunity via bioinformatics software as well as servers. From its main significant applications is prediction a collection of specific B and T cells epitopes via MHC class I and II (Tarrahimofrad et al., 2022). Compared to laboratory tests, this procedure is less expensive and time-consuming. With that method, immunogenic regions from the genomes of pathogens can be chosen. The optimal locations could be created as possible vaccination candidates to stimulate the hosts' protective immune responses (Raoufi et al., 2020). B cell epitopes can be either linear (continuous) or conformational (discontinuous). The target protein's amino acid sequence is used as an epitope in linear epitopes, and there are various criteria to choose the best epitope (Sharon et al., 2014). X-ray crystallographic and NMR procedures are the two practices for verifying the count of B cell epitopes which are time-consuming and very expensive. These days, epitopes prediction of B cell becomes more accurate. Most antigenic epitopes are 3D structures made up of several protein parts rather than linear amino acid sequences. As a result, their bioinformatics modelling is necessary for the proper creation of antigenic regions when predicting the B cell epitopes (Potocnakova et al., 2016).

.

Table (2): Prediction of B-cell epitopes are using different servers.

Server	Link	Type of prediction
ABCPred	http://www.imtech.res.in/raghava/abcpred/	continuous
BepiPred	http://www.cbs.dtu.dk/services/BepiPred/	continuous
Bcepred	http://www.imtech.res.in/raghava/bcepred/	continuous
BEST	http://biomine.ece.ualberta.ca/BEST/	continuous
MIMOX	http://immunet.cn/mimox/	discontinuous
EpiSearch	http://curie.utmb.edu/episearch.html	discontinuous
Discotope	http://www.cbs.dtu.dk/services/DiscoTope/	discontinuous
SEPPA	http://lifecenter.sgst.cn/seppa/index.php	discontinuous
MimoPro	http://informatics.nenu.edu.cn/MimoPro	discontinuous
BEPro (PEPITO)	http://pepito.proteomics.ics.uci.edu/	discontinuous
Pep-3D-Search	http://kyc.nenu.edu.cn/Pep3DSearch	discontinuous
ElliPro	http://tools.immuneepitope.org/tools/ElliPro/iedb_input	continuous and discontinuous
PepSurf	http://pepitope.tau.ac.il	Continuous and discontinuous
Epitopia	http://epitopia.tau.ac.il/	continuous and discontinuous

Molecular Docking

Virtual Screening In Structure-Based Drug Discovery (SBDD)

Without knowing in advance, the chemical composition of other target modulators, molecular docking promotes the discovery of novel therapeutically relevant compounds and the molecular prediction of ligand-target interactions. Furthermore, Understanding how small and large molecules recognize one another is helpful (Pinzi and Rastelli, 2019).

In fact, in silico methods today use virtual screening for millions of molecules in a short amount of time, lowering the initial expenses of hit detection and increasing chances of discovering the desired drug candidates. Presently, structure-based and ligand-based methods have two categories into which various molecular docking methods for drug discovery activities are divided (Cheng et al., 2012). Structure-based drug discovery (SBDD) begins with modelling the 3D structure of the target protein. Finding ligand receptors on targets in biology is becoming more and more crucial. Following receptor and library setup, every molecule in the library is virtually docked into binding receptor of the target via a docking software (Yu and Mackerell, 2017).

By identifying the conformational space of the ligands within the binding position of the candidate protein, docking attempts to anticipate the structure of the ligand protein interaction. The energy binding between the ligand and the protein for each docking pose is then roughly calculated using a scoring algorithm. Chemical diversity, binding scores, lead-likeness, and the validity of the produced pose, metabolic liabilities, unwanted chemical moieties, and score-producing ranked compounds are all calculated after docking and ranking the compounds. Further processing results of the selected compounds are exposed to experimental trials (Nitulescu et al., 2023).

Protein Preparation For SBVS

The target protein's and the ligand's designed architectures determine whether SBVS will be successful. Only heavy atoms compose a traditional PDB structure

file, which may also contain water molecules, ligands, cofactors, metal ions, activating agents, and different protein subunits (Lionta et al., 2014). The overall steps for protein preparation involve; firstly, determining the protonation conditions of the aa in the target protein via accessible software. Secondly, assigning and optimizing hydrogen atoms and bonds in protein. Finally, To reduce steric conflicts, the protein structure is minimized, partial charges are assigned, residues are capped, metals are treated, lacking loops and side chains are filled in, and metals are treated (Madhavi Sastry et al., 2013).

Compound Dataset Preparation

The next important stage in the development of the SBVS is the generation of compound datasets. Databases for SBVS contain drug-like small molecules such as ZINC database and Drug Bank database, often free or purchase available or even synthesis, that have desirable characteristics like solubility and stability, presence of proper functional sets to cooperate with targets and lack of toxic and unwanted moiety(Maia et al., 2020). The molecular weight (MW) of drug-like compounds should be less than 500, the lipophilicity (logP) value should be less than 5, and there should be less than five and ten hydrogen bond donors and acceptors, respectively. It should be emphasized that most commercial substances identified in chemical libraries are more hydrophobic and have higher molecular weights than orally accessible medications (Roskoski, 2023).

Bioinformatics Applications Applied in Brucellosis.

Brucellosis is one from the top listed zoonotic infectious diseases that cause huge economic losses all over the world (Mousa et al., 2022). *Brucella* is immune evasion organism that escape from the immune system and has the ability to intracellular replication and survival in the host cells (Elrashedy et al., 2022). The application of protein modelling and molecular docking were used in *Brucella*. The BvrR/BvrR system is associated with genes responsible for virulence factors, membrane transport and metabolism. In study conducted by Ramírez-González et al., they performed 3D structure prediction of BvrR protein using I-TASSER server, one of the best servers in 3D stuructre prediction according to CASP15, and knew the mechanism of interaction between BvrR protein and DNA through molecular docking and molecular dynamics simulation (MD)(Ramírez-González et al., 2019).

Moreover, Proteomic annotation is employed on *B. suis* to decode the sequence, discover new proteins can be used as a drug candidate. It displayed different features including exclusive metabolic pathway, essential, non-homologous, virulence, drug like and resistance proteins. Through this information and previous prior art, they selected isocitratelyase as a powerful drug candidate and collected around 18,000 Zinc compound for docking virtually, then they selected few compounds from the ranked dataset and examine them for the absorption, distribution, metabolism, and excretion (ADMET) properties (Khan et al., 2022).

Although the obstacles of S19 and RB51 vaccines strains in *B. abortus* and REV1 vaccine in *B. melitensis*, they are widely used, also there is no recombinant vaccine used against more than one species (Heidary et al., 2022). Therefore, it is vital to design new candidate vaccine that trigger the immune response with the advanced bioinformatics tools. Comparative genomic analysis between the isolated field strains and vaccine strains of *Brucella* gives new insights on the level of the genome feature to detect the evolutionary relationship between these strains and discover new target vaccine candidates (Wang et al., 2020,

2022). They reveled that OMP2b protein c;arified three epitopes forB cell, three epitopes for CD4+ T cell, and five CD8+ T cell epitopes, P39 protein recognized three epitopes for linear B cell, two epitopes for CD4+ T cell and two epitopes for CD8+ T cell, and BLS protein detect one B cell epitope, two CD4+ T cell epitopes and one CD8+ T cell epitope (Sha et al., 2020).

Table (3): Prediction of T-cell epitopes are using different servers.

Server	**Link**	**Predictive**	
		for	**method**
nHLAPred	http://www.imtech.res.in/raghava/nhlapred/	MHC I	Artificial Neural Networks
EpiJen	http://www.ddgpharmfac.net/epijen/EpiJen/EpiJen.htm	MHC I	Multi-step algorithm
NetCTL	http://www.cbs.dtu.dk/services/NetCTL	MHC I	ANN-regression
NetMHC	http://www.cbs.dtu.dk/services/NetMHC/	MHC I	ANN based method
KISS	http://cbio.ensmp.fr/kiss/	MHC I	SVM based method
BIMAS	http://www-bimas.cit.nih.gov/molbio/hla_bind/	MHC I	Published coefficient tables
MMBPred	http://www.imtech.res.in/raghava/mmbpred/	MHC I	Quantitative matrix
PREDEP	http://margalit.huji.ac.il/Teppred/mhc-bind/index.html	MHC I	Published coefficient tables
ANNPRED	http://www.imtech.res.in/raghava/nhlapred/neural.html	MHC I	ANN-regression
ProPred I	http://www.imtech.res.in/raghava/propred1/	MHC I	Quantitative matrix
ProPred	http://www.imtech.res.in/raghava/propred/	MHC II	Quantitative matrix
IMTECH	http://www.imtech.res.in/raghava/mhc	MHC II	Quantitative matrix
MHC2Pred	http://www.imtech.res.in/raghava/mhc2pred/	MHC II	SVM-based method
IMTECH	http://www.imtech.res.in/raghava/mhc	MHC II	Quantitative matrix
IEDB	https://www.iedb.org/	MHC I and II	ANN and SMM method
SVRMHC	http://svrmhc.biolead.org/	MHC I and II	SVM-based method
SYFPEITHI	http://www.syfpeithi.de/bin/MHCServer.dll/EpitopePrediction.htm	MHC I and II	Published motifs
SVMHC	http://abi.inf.uni-tuebingen.de/Services/SVMHC	MHC I and II	SVM-based method
MHCPred	http://www.ddg-pharmfac.net/mhcpred/MHCPred/	MHC I and II	Additive method
EpiVax	http://www.epivax.com/	MHC I and II	Epimatrix algorithm
RANKPEP	http://bio.dfci.harvard.edu/RANKPEP/	MHC I and II	PSSM

SVM: support vector machine, *ANN:* artificial neural networks, *PSSM:* position-specific scoring matrix.

CONCLUSION

In conclusion, bioinformatics has immense importance in different fields such as biotechnology and medicine. It has different many applications on sequence and structural analysis for biological data. Each application has its role for interpreting the data. Database search sequence and alignment are the fundamental tools that prerequisite for further

applications such as modelling, epitope predictions and molecular docking. Genome annotation is considered the decipher to understand the meaningless information of sequences and give valuable information of the protein coding genes and genomic features. Moreover, protein modelling is cornerstone for other proteomics applications, and it has three different approaches: comparative modelling, fold modelling and *Ab initio* modelling. Another one important application is epitopes prediction for B cells and T cells through MHC I and MHCII. This can help for generate new candidate vaccine for preventing and controlling infectious diseases. The produced 3D model or the predicted epitopes can be docked virtually with ligand by software to design new drug or vaccine with low cost and short time. In last years, there were great attention to use these applications (comparative genomic analysis, annotation, protein modelling, epitopes prediction, and molecular docking) in diagnosis and control of brucellosis.

REFERENCES

Bawono, P., Dijkstra, M., Pirovano, W., Feenstra, A., Abeln, S., and Heringa, J. (2017). Multiple Sequence Alignment. *Methods in Molecular Biology (Clifton, N.J.)*, *1525*, 167–189. https://doi.org/10.1007/978-1-4939-6622-6_8

Bayat, A., Gaëta, B., Ignjatovic, A., and Parameswaran, S. (2019). Pairwise alignment of nucleotide sequences using maximal exact matches. *BMC Bioinformatics*, *20*(1), 1–15. https://doi.org/10.1186/S12859-019-2827-0/FIGURES/11

Berger, B., Waterman, M. S., and Yu, Y. W. (2021). Levenshtein Distance, Sequence Comparison and Biological Database Search. *IEEE Transactions on Information Theory*, *67*(6), 3287–3294. https://doi.org/10.1109/TIT.2020.2996543

Cheng, T., Li, Q., Zhou, Z., Wang, Y., and Bryant, S. H. (2012). Structure-Based Virtual Screening for Drug Discovery: a Problem-Centric Review. *The AAPS Journal*, *14*(1), 133–141. https://doi.org/10.1208/s12248-012-9322-0

Chenna, R. (2003). Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Research*, *31*(13), 3497–3500. https://doi.org/10.1093/nar/gkg500

Cheung, N. J., and Yu, W. (2018). De novo protein structure prediction using ultra-fast molecular dynamics simulation. *PLOS ONE*, *13*(11), e0205819. https://doi.org/10.1371/journal.pone.0205819

Edwards, D. J., and Holt, K. E. (2013). Beginner's guide to comparative bacterial genome analysis using next-generation sequence data. *Microbial Informatics and Experimentation*, *3*(1), 2. https://doi.org/10.1186/2042-5783-3-2

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

Elrashedy, A., Gaafar, M., Mousa, W., Nayel, M., Salama, A., Zaghawa, A., Elsify, A., and Dawood, A. S. (2022). Immune response and recent advances in diagnosis and control of brucellosis. *German Journal of Veterinary Research*, *2*(1), 10–24. https://doi.org/10.51585/gjvr.2022.1.0033

Haddad, Y., Adam, V., and Heger, Z. (2020). Ten quick tips for homology modeling of high-resolution protein 3D structures. *PLoS Computational Biology*, *16*(4). https://doi.org/10.1371/JOURNAL.PCBI.1007449

Hauser, M., Mayer, C. E., and Söding, J.

(2013). kClust: fast and sensitive clustering of large protein sequence databases. *BMC Bioinformatics*, *14*(1), 248. https://doi.org/10.1186/1471-2105-14-248

Heidary, M., Dashtbin, S., Ghanavati, R., Mahdizade Ari, M., Bostanghadiri, N., Darbandi, A., Navidifar, T., and Talebi, M. (2022). Evaluation of Brucellosis Vaccines: A Comprehensive Review. *Frontiers in Veterinary Science*, *9*, 925773. https://doi.org/10.3389/FVETS.2022.925773

Hernández-Domínguez, E. M., Castillo-Ortega, L. S., García-Esquivel, Y., Mandujano-González, V., Díaz-Godínez, G., Álvarez-Cervantes, J., Hernández-Domínguez, E. M., Castillo-Ortega, L. S., García-Esquivel, Y., Mandujano-González, V., Díaz-Godínez, G., and Álvarez-Cervantes, J. (2019). Bioinformatics as a Tool for the Structural and Evolutionary Analysis of Proteins. *Computational Biology and Chemistry*. https://doi.org/10.5772/INTECHOPEN.89594

Hu, B., Xie, G., Lo, C.-C., Starkenburg, S. R., and Chain, P. S. G. (2011). Pathogen comparative genomics in the next-generation sequencing era: genome alignments, pangenomics and metagenomics. *Briefings in Functional Genomics*, *10*(6), 322–333. https://doi.org/10.1093/bfgp/elr042

Karthik, K., Anbazhagan, S., Thomas, P., Ananda Chitra, M., Senthilkumar, T. M. A., Sridhar, R., and Dhinakar Raj, G. (2021). Genome Sequencing and Comparative Genomics of Indian Isolates of Brucella melitensis. *Frontiers in Microbiology*, *12*. https://doi.org/10.3389/FMICB.2021.698069

Kelm, S., Choi, Y., and Deane, C. M. (2014). Protein Modeling and Structural Prediction. In *Springer Handbook of Bio-/Neuroinformatics* (pp. 171–182). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-30574-0_11

Khan, K., Alhar, M. S. O., Abbas, M. N., Abbas, S. Q., Kazi, M., Khan, S. A., Sadiq, A., Hassan, S. S. ul, Bungau, S., and Jalal, K. (2022). Integrated Bioinformatics-Based Subtractive Genomics Approach to Decipher the Therapeutic Drug Target and Its Possible Intervention against Brucellosis. *Bioengineering*, *9*(11), 633. https://doi.org/10.3390/BIOENGINEERING9110633/S1

Kong, R., Liu, R. R., Xu, X. M., Zhang, D. W., Xu, X. S., Shi, H., and Chang, S. (2020). Template-based modeling and ab-initio docking using CoDock in CAPRI. *Proteins*, *88*(8), 1100–1109. https://doi.org/10.1002/PROT.25892

Kufareva, I., and Abagyan, R. (2011). *Methods of Protein Structure Comparison* (pp. 231–257). https://doi.org/10.1007/978-1-61779-588-6_10

Kuhlman, B., and Bradley, P. (2019). Advances in protein structure prediction and design. *Nature Reviews Molecular Cell Biology 2019 20:11*, *20*(11), 681–697. https://doi.org/10.1038/s41580-019-0163-x

Kumar, S., Demo, G., Koca, J., and Wimmerova, M. (2012). In Silico Engineering of Proteins That Recognize Small Molecules. In *Protein Engineering*. InTech. https://doi.org/10.5772/28001

Lantz, H., Dominguez Del Angel, V., Hjerde, E., Sterck, L., Capella-Gutierrez, S., Notredame, C., Vinnere Pettersson, O., Amselem, J., Bouri, L., Bocs, S., Klopp, C., Gibrat, J. F., Vlasova, A., Leskosek, B. L., Soler, L., and Binzer-Panchal, M.

(2018). Ten steps to get started in Genome Assembly and Annotation. *F1000Research*, *7*. https://doi.org/10.12688/F1000RESEARCH.13598.1

Lee, J., Freddolino, P. L., and Zhang, Y. (2017). Ab Initio Protein Structure Prediction. In *From Protein Structure to Function with Bioinformatics* (pp. 3–35). Springer Netherlands. https://doi.org/10.1007/978-94-024-1069-3_1

Lionta, E., Spyrou, G., Vassilatis, D., and Cournia, Z. (2014). Structure-Based Virtual Screening for Drug Discovery: Principles, Applications and Recent Advances. *Current Topics in Medicinal Chemistry*, *14*(16), 1923–1938. https://doi.org/10.2174/1568026614666140929124445

Madhavi Sastry, G., Adzhigirey, M., Day, T., Annabhimoju, R., and Sherman, W. (2013). Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. *Journal of Computer-Aided Molecular Design*, *27*(3), 221–234. https://doi.org/10.1007/S10822-013-9644-8

Maia, E. H. B., Assis, L. C., de Oliveira, T. A., da Silva, A. M., and Taranto, A. G. (2020). Structure-Based Virtual Screening: From Classical to Artificial Intelligence. *Frontiers in Chemistry*, *8*, 343. https://doi.org/10.3389/FCHEM.2020.00343/BIBTEX

Mishra, S., Rastogi, Y. P., Jabin, S., Kaur, P., Amir, M., and Khatoon, S. (2019). A bacterial phyla dataset for protein function prediction. *Data in Brief*, *28*. https://doi.org/10.1016/J.DIB.2019.105002

Mousa, W. S., Gaafar, M., Abdel, A., Zaghawa, M., Nayel, M. A., Elsify, A. M., Elsobky, Y. A., Ramadan, E. S., Elrashedy, A., Arbaga, A. A., and Salama, A. (2022). *Journal of Current Veterinary Research*.

Muhammed, M. T., and Aki-Yalcin, E. (2019). Homology modeling in drug discovery: Overview, current applications, and future perspectives. *Chemical Biology and Drug Design*, *93*(1), 12–20. https://doi.org/10.1111/CBDD.13388

Nitulescu, M., Alves de Oliveira, T., Pires da Silva, M., Habib Bechelane Maia, E., Marques da Silva, A., and Gutterres Taranto, A. (2023). Virtual Screening Algorithms in Drug Discovery: A Review Focused on Machine and Deep Learning Methods. *Drugs and Drug Candidates 2023, Vol. 2, Pages 311-334*, *2*(2), 311–334. https://doi.org/10.3390/DDC2020017

Oliver, G. R., Hart, S. N., and Klee, E. W. (2015). Bioinformatics for clinical next generation sequencing. *Clinical Chemistry*, *61*(1), 124–135. https://doi.org/10.1373/CLINCHEM.2014.224360

Outeiral, C., Nissley, D. A., and Deane, C. M. (2022). Current structure predictors are not learning the physics of protein folding. *Bioinformatics*, *38*(7), 1881. https://doi.org/10.1093/BIOINFORMATICS/BTAB881

Parthasarathy, S. (2015). Bioinformatics: Application to genomics. *Plant Biology and Biotechnology: Volume II: Plant Genomics and Biotechnology*, 279–300. https://doi.org/10.1007/978-81-322-2283-5_13/FIGURES/9

Pearce, R., and Zhang, Y. (2021). Deep learning techniques have significantly impacted protein structure prediction and protein design. *Current Opinion in Structural Biology*, *68*, 194–207. https://doi.org/10.1016/j.sbi.2021.01.007

Pinzi, L., and Rastelli, G. (2019). Molecular Docking: Shifting Paradigms in Drug Discovery. *International Journal of*

Molecular Sciences, *20*(18). https://doi.org/10.3390/IJMS20184331

Potocnakova, L., Bhide, M., and Pulzova, L. B. (2016). An Introduction to B-Cell Epitope Mapping and In Silico Epitope Prediction. *Journal of Immunology Research*, *2016*, 1–11. https://doi.org/10.1155/2016/6760830

Ramírez-González, E. A., Moreno-Lafont, M. C., Méndez-Tenorio, A., Cancino-Díaz, M. E., Estrada-García, I., and López-Santiago, R. (2019). Prediction of Structure and Molecular Interaction with DNA of BvrR, a Virulence-Associated Regulatory Protein of Brucella. *Molecules*, *24*(17), 3137. https://doi.org/10.3390/molecules24173137

Raoufi, E., Hemmati, M., Eftekhari, S., Khaksaran, K., Mahmodi, Z., Farajollahi, M. M., and Mohsenzadegan, M. (2020). Epitope Prediction by Novel Immunoinformatics Approach: A State-of-the-art Review. *International Journal of Peptide Research and Therapeutics*, *26*(2), 1155–1163. https://doi.org/10.1007/s10989-019-09918-z

Roskoski, R. (2023). Rule of five violations among the FDA-approved small molecule protein kinase inhibitors. *Pharmacological Research*, *191*, 106774. https://doi.org/10.1016/J.PHRS.2023.106774

Roy, A., and Zhang, Y. (2012). Protein Structure Prediction. In *eLS*. Wiley. https://doi.org/10.1002/9780470015902.a0003031.pub2

Sha, T., Li, Z., Zhang, C., Zhao, X., Chen, Z., Zhang, F., and Ding, J. (2020). Bioinformatics analysis of candidate proteins Omp2b, P39 and BLS for Brucella multivalent epitope vaccines. *Microbial Pathogenesis*, *147*, 104318. https://doi.org/10.1016/J.MICPATH.2020.104318

Sharon, J., Rynkiewicz, M. J., Lu, Z., and Yang, C.-Y. (2014). Discovery of protective B-cell epitopes for development of antimicrobial vaccines and antibody therapeutics. *Immunology*, *142*(1), 1–23. https://doi.org/10.1111/imm.12213

Slodkowicz, G., and Goldman, N. (2020). Integrated structural and evolutionary analysis reveals common mechanisms underlying adaptive evolution in mammals. *Proceedings of the National Academy of Sciences*, *117*(11), 5977–5986. https://doi.org/10.1073/pnas.1916786117

Sofi, M. Y., Shafi, A., and Masoodi, K. Z. (2022). CLUSTALW software. *Bioinformatics for Everyone*, 75–84. https://doi.org/10.1016/B978-0-323-91128-3.00003-3

Tarrahimofrad, H., Zamani, J., Hamblin, M. R., Darvish, M., and Mirzaei, H. (2022). A designed peptide-based vaccine to combat Brucella melitensis, B. suis and B. abortus: Harnessing an epitope mapping and immunoinformatics approach. *Biomedicine and Pharmacotherapy*, *155*, 113557. https://doi.org/10.1016/J.BIOPHA.2022.113557

Tettelin, H., Riley, D., Cattuto, C., and Medini, D. (2008). Comparative genomics: the bacterial pan-genome. *Current Opinion in Microbiology*, *11*(5), 472–477. https://doi.org/10.1016/J.MIB.2008.09.006

Trindade Maia, R., de Araújo Campos, M., and Maciel de Moraes Filho, R. (2021). Introductory Chapter: Homology Modeling. In *Homology Molecular Modeling - Perspectives and Applications*. IntechOpen. https://doi.org/10.5772/intechopen.9544

6

Vaser, R., Sović, I., Nagarajan, N., and Šikić, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Research*, *27*(5), 737–746. https://doi.org/10.1101/GR.214270.116

Wang, S., Wang, W., Sun, K., Bateer, H., and Zhao, X. (2020). Comparative genomic analysis between newly sequenced Brucella abortus vaccine strain A19 and another Brucella abortus vaccine S19. *Genomics*, *112*(2), 1444–1453. https://doi.org/10.1016/J.YGENO.2019.08.015

Wang, S., Zhao, X., Sun, K., Bateer, H., and Wang, W. (2022). The Genome Sequence of Brucella abortus vaccine strain A19 provides insights on its virulence attenuation compared to Brucella abortus strain 9-941. *Gene*, *830*, 146521. https://doi.org/10.1016/J.GENE.2022.146521

Yu, W., and Mackerell, A. D. (2017). Computer-Aided Drug Design Methods. *Methods in Molecular Biology (Clifton, N.J.)*, *1520*, 85–106. https://doi.org/10.1007/978-1-4939-6634-9_5

Zhang, T., Zhou, J., Gao, W., Jia, Y., Wei, Y., and Wang, G. (2022). Complex genome assembly based on long-read sequencing. *Briefings in Bioinformatics*, *23*(5). https://doi.org/10.1093/bib/bbac305

Zhang, W., Yang, J., He, B., Walker, S. E., Zhang, H., Govindarajoo, B., Virtanen, J., Xue, Z., Shen, H.-B., and Zhang, Y. (2016). Integration of QUARK and I-TASSER for Ab Initio Protein Structure Prediction in CASP11. *Proteins: Structure, Function, and Bioinformatics*, *84*, 76–86. https://doi.org/10.1002/prot.24930