



Juveriya Israr, Shabroz Alam, Sahabjada Siddiqui, Sankalp Misra, Indrajeet Singh, and Ajay Kumar

Abstract

Structural bioinformatics is a captivating discipline that delves into the intricate realm related to proteins, RNA, and DNA, the macromolecules of life. Its primary focus lies in comprehending and foreseeing the enigmatic three-dimensional (3D) architecture of these fundamental entities. By employing cutting-edge computational techniques and advanced algorithms, structural bioinformatics unravels the complex interplay between structure and function, shedding light on the inner workings of life's molecular machinery. Bioinformatics is an interdisciplinary field that combines experimental and computational approaches to investigate various aspects of macromolecular 3D structure. By utilizing experimentally determined structures and computational models, bioinformatics aims to explore diverse inquiries related to macromolecules. These inquiries encompass understanding the distinctions and similarities between macro and micro structures, understanding the rules of molecular interaction, evolution, and folding, and revealing the complexity of structure-function correlations. Structural

J. Israr

Faculty of Biosciences, Institute of Biosciences and Technology, Shri Ramswaroop Memorial University, Barabanki, Uttar Pradesh, India

Department of Biotechnology, Era University, Lucknow, Uttar Pradesh, India

S. Alam · S. Siddiqui

Department of Biotechnology, Era University, Lucknow, Uttar Pradesh, India

S. Misra

Faculty of Biosciences, Institute of Biosciences and Technology, Shri Ramswaroop Memorial University, Barabanki, Uttar Pradesh, India

I. Singh · A. Kumar (✉)

Department of Biotechnology, Rama University, Kanpur, Uttar Pradesh, India

e-mail: drajay.fet@ramauniversity.ac.in

bioinformatics, a specialized domain within the realm of computational structural biology, encompasses the study and analysis of biological structures. The term “structural” in this context aligns with its definition in the field of structural biology. The field of structural bioinformatics is dedicated to addressing biological challenges and unveiling novel insights through the development of innovative methodologies for the analysis of data pertaining to biological macromolecules.

Keywords

Structural biology · Protein data bank · Drug discovery · Molecular modeling

2.1 Introduction

The growing domain of structural bioinformatics encompasses the dynamic realm of prognosticating and scrutinizing protein architectures. Utilizing state-of-the-art methodologies and computational algorithms, bioinformatics researchers have achieved remarkable advancements in the realm of protein modeling. By employing these cutting-edge techniques, they can effectively simulate intricate protein structures, forecast protein-protein interactions, and discern critical binding sites crucial in order to create innovative medicines. This document aims to delve into the latest breakthroughs in the realm of bioinformatics. Important biological macromolecules including proteins, RNA, and DNA are the subject of structural bioinformatics, a subfield of bioinformatics that analyzes and predicts their three-dimensional structures (Patel et al. 2019). Bioinformatics is a multidisciplinary field that encompasses the study of macromolecular 3D structures. Analyzing and generalizing about these structures includes doing things like comparing overall folds and local motifs, learning the principles of molecular folding, learning about evolutionary relationships, learning about binding interactions, and learning how these structures are put together to perform specific tasks. (Gauthier et al. 2019). This comprehensive approach utilizes both experimentally determined structures and computational models to gain insights into the intricate world of macromolecules. In structural bioinformatics, the word “structural” has the same meaning as it does in structural biology. As a subfield of computational structural biology, structural bioinformatics is essentially an essential part of the field. Structural bioinformatics, as a field, is primarily focused on the development and implementation of innovative methodologies for the analysis and manipulation of biological macromolecular data. By harnessing these advanced techniques, researchers aim to address complex biological challenges and uncover novel insights into the intricate workings of living systems. The overarching goal is to not only expand our understanding of biology but also pave the way for the generation of transformative knowledge that can revolutionize various facets of scientific inquiry (Gu and Bourne 2011; Wei et al. 2014).

2.2 Protein Structure

In bioinformatics, understanding how a protein's structure contributes to its function is crucial. The intricate three-dimensional arrangement of amino acids within a protein dictates its ability to carry out specific biological tasks. This fundamental principle, known as structure-function relationship, forms the cornerstone of our understanding of protein biology (Rigden 2009). By deciphering the structural characteristics of proteins, bioinformaticians can unravel the underlying mechanisms that govern their diverse functions, ranging from enzymatic catalysis to molecular recognition. Consequently, the Proteins, through the strategic arrangement of distinct chemical moieties, exhibit enzymatic properties that facilitate the acceleration of diverse biochemical reactions. Primary, secondary, tertiary, and quaternary structures are the usual divisions into a protein's four stages of organization explained in Fig. 2.1 (Kocincová et al. 2017).

Structural bioinformatics is a field that primarily focuses on the analysis and understanding of interactions between biomolecular structures, with a particular emphasis on their spatial coordinates. The analysis of the primary structure is commonly undertaken within the purview of conventional bioinformatics disciplines. Specific constraints in the underlying genetic code of the supplied sequence allow for the formation of conserved regional topologies within the polypeptide chain, such as alpha-helices, beta-sheets, and loops. In the realm of bioinformatics, it is worth noting that the protein fold is fortified by a series of feeble interactions, including but not limited to hydrogen bonds. The stability of the protein

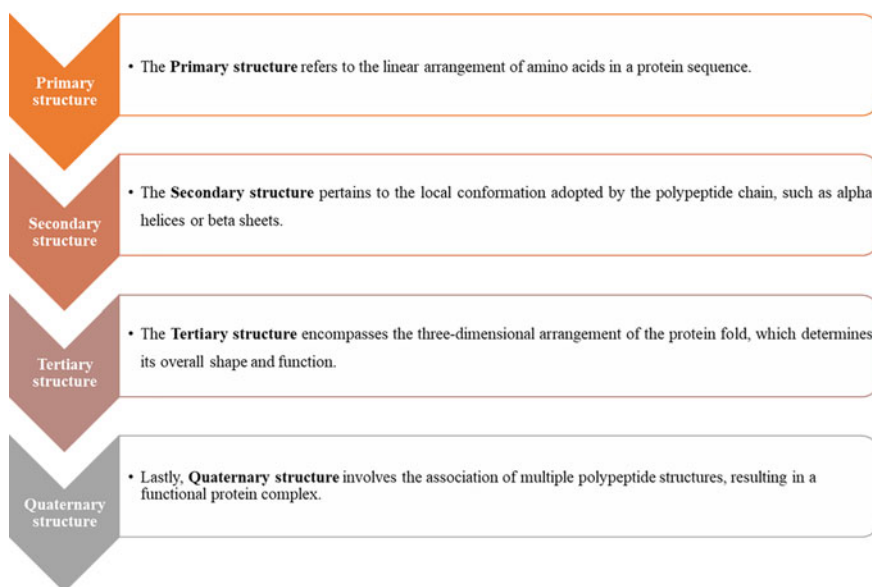


Fig. 2.1 Types of protein structure

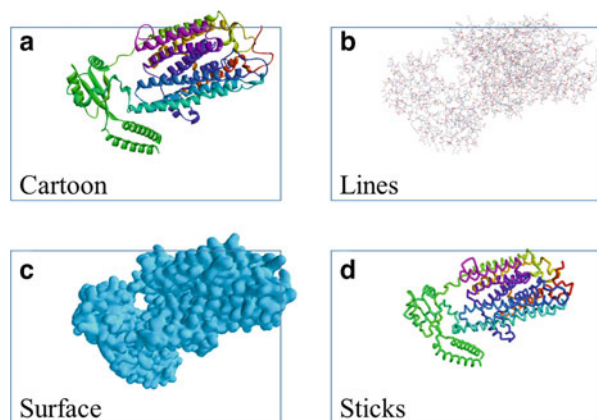
structure depends on these interactions. In the realm of bioinformatics, interactions can manifest in two distinct ways: intrachain and interchain. Intrachain interactions transpire within the confines of a single protein monomer, specifically within its tertiary structure. On the other hand, interchain interactions take place between diverse structures, commonly referred to as the quaternary structure. In the realm of structural bioinformatics, researchers are currently engaged in the comprehensive examination of the intricate organization of interactions, encompassing both robust and delicate connections, as well as the interwoven complexities. This captivating domain employs cutting-edge methodologies, including circuit topology frameworks, to unravel the profound mysteries underlying the topological configuration of biological systems.

2.3 Structure Visualization

The visualization of protein structures holds significant importance within the field of structural bioinformatics. The user's text is not provided. The platform facilitates the visualization of both static and dynamic molecular representations, enabling the identification of molecular interactions that can be leveraged for the inference of underlying molecular mechanisms. In the field of bioinformatics, a plethora of visualization techniques have emerged to aid when it comes to deciphering and analyzing complex biological data. Among the most prevalent and widely utilized types of visualization are (Fig. 2.2) (Shi et al. 2017):

Cartoon: The utilization of cartoon representations in protein visualization serves as a valuable tool for highlighting variations in secondary structure. In the realm of bioinformatics, the α -helix, a fundamental structural motif, is often symbolically depicted as a helical screw, embodying its characteristic spiral conformation. Similarly, β -strands, another crucial element of protein structure, are commonly represented as arrows, symbolizing their linear arrangement. Furthermore, the flexible regions known as loops are typically denoted by straight lines, capturing their

Fig. 2.2 Structural visualization of transmembrane protein 63B (PDB ID: 8EHX). (a) Cartoon; (b) lines; (c) surface; (d) sticks



inherent flexibility and variability. The study and analysis of protein structures can be aided by these visual representations, enabling researchers to comprehend and interpret the intricate three-dimensional arrangements of these vital biomolecules.

Lines: In this bioinformatics representation, the amino acid residues are elegantly depicted as slender lines, enabling efficient and cost-effective graphic rendering.

Surface: The visualization depicts the molecular structure's external shape, providing insights into its surface characteristics.

Sticks: Stick diagrams are a common way for bioinformaticians to visualize the complex world of molecular structures. The framework's slender sticks representing the covalent links between amino acid atoms perfectly capture the essence of this chemical connection. This method of visualization is typically used to better understand and portray the complex web of interactions that takes place between amino acids.

2.4 DNA Structure Background

The seminal elucidation of the DNA duplex structure was first expounded by the esteemed scientific duo of Watson and Crick, with notable contributions from the esteemed researcher Rosalind Franklin. The DNA molecule is a complex construction made composed of a phosphate group, a pentose sugar, and a nitrogenous base. These constituents, when combined, form the fundamental building blocks of DNA. The phosphate group serves as a crucial backbone, providing structural stability to the molecule. The pentose sugar, a five-carbon sugar, acts as a central framework, connecting the various components of DNA. Lastly, the nitrogenous bases, which include adenine, thymine, cytosine, and guanine, play a pivotal role in encoding genetic information within the DNA molecule (Travers and Muskhelishvili 2015). Together, these three substances harmoniously unite to form the remarkable DNA molecule, the cornerstone of life's genetic blueprint. Hydrogen bonding between complementary base pairs are responsible for maintaining the DNA double helix's structural integrity. Adenine and thymine (A-T) and cytosine and guanine (C-G) form hydrogen bonds in DNA. These hydrogen bonds play a crucial role in stabilizing the overall structure of DNA. Structural bioinformatics research endeavors have predominantly concentrated on elucidating the intricate interplay between deoxyribonucleic acid (DNA) and diminutive chemical entities. This captivating area of investigation has garnered considerable attention in the realm of drug design, with numerous studies dedicated to unravelling the underlying mechanisms governing these interactions (Fig. 2.3).

2.5 Interactions

Interactions encompass the intricate network of contacts that are established between various components of molecules operating at distinct hierarchical levels. As a vital component of the intricate world of bioinformatics, they assume the crucial role of



Fig. 2.3 DNA structure

ensuring the stability of protein structures while engaging in a diverse array of functional activities. In bioinformatics, the study of molecular interactions entails the identification, classification, and evaluation of sets of atoms or molecular areas that have an effect on one another. The hydrophobic effect, hydrogen bonding, and electrostatic forces are all possible explanations for these phenomena. Proteins engage in a wide variety of interactions, such as those between themselves and other proteins, between themselves and peptides, between themselves and ligands, and between themselves and DNA (Stanfield and Wilson 1995; Klebe 2015).

2.6 Calculating Contacts

The computation of contacts holds significant significance within the realm of structural bioinformatics. Docking and molecular dynamics analyses are made much easier using it, and it also helps predict protein structure and folding, thermodynamic stability, protein-protein and protein-ligand interactions, and more. In the realm of bioinformatics, conventional approaches have relied upon computational methodologies that leverage the concept of threshold distance, commonly referred to as cut-off, in order to identify potential interactions among atoms. The detection methodology employed in this study relies on the calculation of Euclidean distances and angles between atoms of specific types. In the realm of bioinformatics, it has been observed that a majority of the methodologies relying on the straightforward Euclidean distance principle tend to fall short in effectively identifying occluded

Fig. 2.4 Distance criteria for contact definition

Distance criteria for contact definition

Type	Max distance criteria
Hydrogen bond	3,9 Å
Hydrophobic interaction	5 Å
Ionic interaction	6 Å
Aromatic Stacking	6 Å

contacts. In recent years, the utilization of cut-off-free methodologies, such as Delaunay triangulation, has emerged as a prominent approach in the field of bioinformatics. Furthermore, the integration of a diverse range of parameters, such as physicochemical attributes, spatial proximity, molecular structure, and bond orientations, has been leveraged to enhance the accuracy of contact identification. Distance criteria for contact definition are explained in Fig. 2.4 (Martins et al. 2018; da Silveira et al. 2009).

2.7 Protein Data Bank (PDB)

Proteins, DNA, and RNA are just a few examples of the complex biomolecules that are represented in PDB, which is a repository for three-dimensional structural information relevant to macromolecules of biological significance. PDB is achieved and maintained by the acclaimed multinational cooperation known as the Worldwide Protein Data Bank (wwPDB). The PDB is managed and curated by a collaboration that includes PDBe, PDBj, the RCSB, and BMRB, to name a few of the regional organizations involved. The individual plays a crucial role in ensuring that online copies of PDB (Protein Data Bank) data are publicly available to anyone who wants to use them. Bioinformatics has come a long way, as seen by the exponential increase in the quantity of structural data stored in the Protein Data Bank (PDB). These invaluable databases have been growing steadily each year as cutting-edge techniques like X-ray crystallography, NMR spectroscopy, and cryo-electron microscopy have been used to collect them (Berman et al. 2000).

2.8 Data Format

The PDB format, also known as the PDB format, is a textual file format that has been widely adopted in the field of bioinformatics. It serves as a repository for storing crucial information pertaining to the three-dimensional structures of macromolecules. This format has been extensively utilized by the esteemed PDB, which is a renowned source for researchers and scientists worldwide. The PDB

format, due to inherent limitations in its structural framework, imposes restrictions on the representation of expansive molecular structures. More specifically, it can't handle structures with more than 99,999 atoms or more than 62 chains. The macromolecular Crystallographic Information File (mmCIF), or Protein Data Bank exchange (PDBx), is a widely adopted and standardized text file format that serves as a comprehensive representation of crystallographic information. This format is specifically designed to capture and convey essential data pertaining to macromolecular structures obtained through crystallography techniques. By adhering to a consistent and well-defined structure, the PDBx/mmCIF format enables efficient storage, exchange, and analysis of crystallographic information, facilitating seamless collaboration and interoperability within the bioinformatics community. PDBx/mmCIF (Protein Data Bank exchange/macromolecular Crystallographic Information File) has been the primary way of exchanging data within the PDB archive since its introduction in 2014. This transition has allowed for enhanced data representation and improved interoperability within the field of bioinformatics. The PDBx/mmCIF file format, denoted by (.cif) extension, offers a more comprehensive and structured approach to storing and exchanging macromolecular structural information. By adopting this standardized format, researchers and scientists can seamlessly access and analyze protein structure data, facilitating advancements in various areas of bioinformatics research. The PDB (Protein Data Bank) format is a standardized file format that comprises a collection of records denoted by a keyword consisting of a maximum of 6 characters. In contrast, the PDBx/mmCIF format employs a distinct structure that relies on a key-value system. In this system, the key represents a specific name that serves to identify a particular attribute, while the value corresponds to the variable information associated with that attribute.

2.9 Other Structural Databases

Alongside the PDB, a multitude of databases housing comprehensive repositories of protein structures and diverse macromolecules have been established. In the realm of bioinformatics, a multitude of examples can be found that showcase the application of computational techniques to biological data analysis. These instances include:

The **Molecular Modeling Database (MMDB)** is an extensive repository that contains experimentally derived three-dimensional structures of biomolecules. These structures are derived from the Protein Data Bank (PDB), a repository of protein and nucleic acid structures. MMDB provides a valuable platform for researchers in the field of bioinformatics to access and analyze these structures, enabling a deeper understanding of the molecular architecture and function of biomolecules. By leveraging the wealth of information contained within MMDB, scientists can unravel the intricate relationships between structure and function, paving the way for advancements in fields such as drug discovery,

The **Nucleic Acid Database (NDB)** is a comprehensive depository of experimentally derived information pertaining to nucleic acids, including DNA and RNA.

The **Structural Classification of Proteins (SCOP)** is a comprehensive and intricate framework that elucidates the intricate structural and evolutionary connections among proteins that have been experimentally determined. By meticulously analyzing the three-dimensional structures of proteins, SCOP provides a detailed and systematic classification system that allows researchers to comprehend the intricate relationships and evolutionary patterns within the protein universe. The advancement of our knowledge of the complicated molecular machinery that underpins life has been greatly aided by this wonderful resource, which has become a cornerstone in the science of bioinformatics.

TOPOFIT-DB is a comprehensive database that specializes in protein structural alignments utilizing the cutting-edge TOPOFIT methodology. This innovative approach enables the accurate comparison and analysis of protein structures, facilitating the identification of conserved regions and functional motifs. By leveraging TOPOFIT-DB, researchers can gain valuable insights into the structural relationships between proteins, unravelling intricate molecular mechanisms and aiding in the discovery of novel therapeutic targets (Ilyin et al. 2004).

The **Electron Density Server (EDS)** is a cutting-edge bioinformatics tool that provides researchers with invaluable insights into the electron-density maps and statistical analyses pertaining to the fit of crystal structures and their corresponding maps. By leveraging advanced computational algorithms, EDS provides a robust environment for the study and interpretation of electron-density data, helping researchers gain insight into the structural features and molecular interactions of biological macromolecules. EDS's intuitive design and powerful features allow scientists to make sound judgments and propel innovative discoveries in structural biology.

The **Critical Assessment of Protein Structure Prediction (CASP)** is an internationally acclaimed program that encourages scientists everywhere to work together. It's a hub for massive-scale studies to determine how proteins fold in three dimensions. CASP's global reach allows it to convene specialists from a wide range of disciplines to develop protein structure prediction. The Critical Assessment of Protein Structure Prediction (CASP) is the industry benchmark for evaluating protein structure prediction methods. It allows scientists to evaluate the efficacy of their methods for predicting the three-dimensional structure of proteins. Redundancy in protein databases is a problem for protein structure analysis in bioinformatics. To access the PISCES server, the bioinformatics platform used here may compile a vetted set of Protein Data Bank (PDB) entries according to predefined sequence identity and structural quality thresholds. This program swiftly sifts through the millions of PDB structures in the database, using sophisticated algorithms and data mining techniques to zero in on entries that fulfil the user-specified sequence identity criterion and display excellent structural quality. This method guarantees that the resulting list contains only the most pertinent and trustworthy PDB items.

The **Structural Biology Knowledgebase (SBK)** is a comprehensive platform that offers a wide range of sophisticated tools and resources specifically tailored to facilitate and enhance protein research design. With a focus on structural biology, SBK provides invaluable assistance to scientists and researchers in their pursuit of

unravelling the intricate details of protein structures and functions. Equipped with cutting-edge computational algorithms and advanced data analysis techniques, SBK empowers users to explore, analyze, and manipulate protein structures with utmost precision and efficiency. Its extensive repertoire.

ProtCID is an invaluable resource in the field of bioinformatics, specifically designed to cater to the needs of researchers studying protein-protein interactions. The Protein Common Interface Database, or ProtCID, is an extensive database of crystal structures that include homologous proteins that have comparable protein-protein interfaces. By meticulously curating and organizing a vast collection of crystal structures, ProtCID enables scientists to explore and analyze the intricate details of protein-protein interactions. This database offers a unique opportunity to investigate the similarities and differences in the interfaces of homologous proteins, shedding light on the underlying principles governing these crucial interactions. With its user-friendly interface and powerful search capabilities, ProtCID empowers researchers to delve into the wealth of information.

Alpha Fold is a cutting-edge bioinformatics tool that has revolutionized the field of protein structure prediction. Developed by DeepMind, Alpha Fold utilizes advanced machine learning algorithms to accurately predict protein three-dimensional structure prediction using only amino acid sequences by leveraging vast amounts of genomic and proteomic data.

2.10 Structure Comparison

2.10.1 Structural Alignment

Structural alignment, a fundamental technique in bioinformatics, facilitates the comparison of three-dimensional (3D) structures by evaluating their shape and conformation. The user, identified by the numerical identifier, has expressed a desire for their text to be Bioinformatics has revolutionized the field of evolutionary biology by enabling the inference of intricate evolutionary relationships among proteins, even in cases where their sequence similarity is relatively low. Through the application of sophisticated computational algorithms and statistical models, bioinformatics tools have unlocked the potential to unravel the evolutionary history of protein sequences, shedding light on their shared ancestry and divergent paths. By harnessing the power of bioinformatics, researchers can delve into the intricate tapestry of protein evolution, uncovering hidden connections and gaining valuable insights into the complex dynamics that shape the molecular world. Structural alignment is a fundamental technique in bioinformatics that involves the precise alignment of two or more protein structures in three-dimensional space. The atoms in identical places are rotated and translated so that one structure can be superimposed on top of another. Alpha carbon atoms or the backbone heavy atoms (carbon, nitrogen, oxygen) are typically used for the alignment. Protein structures can be compared and analyzed with this alignment method, revealing previously hidden information on evolutionary links, functional domains, and conserved areas. The

root-mean-square deviation (RMSD) of atomic locations is commonly used in bioinformatics to evaluate the quality of an alignment. This index calculates the typical separation between stacked atoms.

The distance between atom i and either a reference atom that corresponds to the same atom in another structure or the mean coordinate of N similar atoms is denoted by the variable δ_i in structural bioinformatics. This distance measurement is crucial for analyzing and comparing the spatial arrangement of atoms in different structures, aiding in the identification of structural similarities and differences. By quantifying the spatial relationships between atoms, δ_i provides valuable insights into the structural characteristics and dynamics of biomolecules. The measurement of the root mean square deviation (RMSD) outcome is typically expressed in the Ångström (Å) unit, a widely used scale in bioinformatics and structural biology. This unit is equivalent to 10^{-10} m, providing a precise and standardized representation of the structural differences between biomolecules. In bioinformatics, the root mean square deviation (RMSD) is a widely used metric to quantify the similarity between protein or nucleic acid structures. The higher the similarity between the two structures, the smaller the RMSD number will be.

$$\text{RMSD} = \sqrt{\frac{1}{N} \sum_{i=1}^N \delta_i^2}$$

2.10.2 Graph-Based Structural Signatures

Structural signatures, commonly referred to as fingerprints, serve as representations of macromolecule patterns, enabling the inference of similarities and distinctions. Comparing a large number of proteins using the Root Mean Square Deviation (RMSD) approach is complicated by the high computing expense of creating structural alignments. The growth of structural signatures in bioinformatics has been aided by the use of graph distance patterns among atom pairs. These signatures have proven to be instrumental in the identification of protein vectors and the detection of non-trivial information within protein structures. In addition, the integration of linear algebra and machine learning techniques has proven to be invaluable in the field of bioinformatics. Protein signature clustering, ligand identification, free energy prediction, and Euclidean distance-based mutation recommendation are just some of the many applications that have benefited from the use of these effective technologies (Pires et al. 2011; Mariano et al. 2019).

2.10.3 Structure Prediction

Bioinformatics work relies heavily on the determination of molecular structures. Methods including X-ray crystallography (XRC), nuclear magnetic resonance

(NMR) spectroscopy, and three-dimensional electron microscopy are used for this purpose. These techniques enable scientists to gain insights into the intricate arrangements of atoms within molecules. However, it is important to note that these methods can be resource-intensive, both in terms of financial costs and experimental efforts. Additionally, certain molecular structures, particularly those of membrane proteins, pose unique challenges and may require specialized approaches for successful establishment. Therefore, the utilization of computational methodologies becomes imperative in the determination of three-dimensional (3D) structures of macromolecules. In the field of bioinformatics, the study of protein structure prediction encompasses various methodologies, primarily categorized into two main approaches: comparative modeling and de novo modeling.

2.10.4 Comparative Modeling

Protein 3D structure prediction is a common use of comparative modeling, also known as homology modeling, in the field of bioinformatics. To do this, they compare the target protein's amino acid sequence to that of a known-structure template protein. By comparing the similarities between the target and template sequences, computational algorithms can predict the three-dimensional structure of the target protein. Comparative modeling is a valuable tool in the field of bioinformatics, enabling researchers to gain insights into protein structure and function. The scientific literature extensively documents the phenomenon wherein proteins that share evolutionary ancestry exhibit a remarkably preserved three-dimensional conformation. The user, identified as, is seeking assistance in rewriting their text to align with the field Furthermore, it is worth noting that protein sequences exhibiting a degree of dissimilarity greater than 20% may exhibit distinct structural conformations (Kaczanowski and Zielenkiewicz 2010; Chothia and Lesk 1986).

2.10.5 De Novo Modeling

In the realm of structural bioinformatics, the concept of de novo modeling, or ab initio modeling, pertains to methodologies employed to derive three-dimensional structures from sequences, obviating the requirement for a pre-existing homologous 3D structure. The field of de novo protein structure prediction continues to captivate the scientific community, as it grapples with the persistent challenge of unravelling protein structures through the application of novel algorithms and methodologies (Meyers et al. 2021). Despite the notable advancements witnessed in recent years, this pursuit remains an unresolved frontier in the realm of modern science.

2.10.6 Structure Validation

Following the process of structure modeling, it becomes imperative to incorporate an essential subsequent phase of structure validation. This is due to the fact that numerous algorithms and tools employed in both comparative and “de novo” modeling rely on heuristic approaches to assemble the three-dimensional structure, thereby leading to the potential generation of a multitude of errors. In the realm of bioinformatics, a plethora of validation strategies are employed to ascertain the accuracy and reliability of computational models. One such approach involves the calculation of energy scores, which are subsequently juxtaposed against experimentally determined structures. This comparative analysis serves as a means to evaluate the fidelity and plausibility of the computational predictions, enabling researchers to gain valuable insights into the structural integrity and functional characteristics of biomolecules. The DOPE score, a widely employed energy score in bioinformatics, plays a pivotal role within the MODELLER tool. Its primary function revolves around the assessment and selection of optimal models. One additional validation strategy involves the computation of φ and ψ backbone dihedral angles for each residue, followed by the construction of a Ramachandran plot. The conformational space of amino acids is influenced by the side-chain properties and the interplay of interactions within the backbone. Consequently, the Ramachandran plot serves as a valuable tool for visualizing the permissible conformations by constraining these two angles. The presence of a substantial abundance of amino acids positioned in non-permissive locations within the chart serves as an indicative characteristic of a modeling outcome of inferior quality (Webb and Sali 2014).

2.10.7 Prediction Tools

The compendium of protein structure prediction software encompasses a comprehensive array of frequently employed computational tools. De novo protein structure prediction, protein threading, comparison modeling, and secondary structure prediction are only few of the methods included in this collection.

2.10.8 Molecular Docking

Among the most often used computational methods in bioinformatics is molecular docking, is employed to forecast the precise spatial arrangement and coordinates of a ligand molecule upon its binding to a receptor or target molecule. The binding phenomenon predominantly occurs via non-covalent interactions, although investigations into covalently linked binding are also conducted. Molecular docking is a computational approach utilized in bioinformatics to forecast potential conformations, also known as binding modes, of a ligand as it engages with distinct regions on a receptor. Bioinformatics software applications employ molecular docking algorithms that leverage the principles of force fields to ascertain an

objective score, thereby facilitating the ranking of optimal molecular conformations. This scoring mechanism is designed to prioritize poses that exhibit enhanced intermolecular interactions between the two biomolecules under investigation.

Docking protocols are widely employed in the field of bioinformatics to computationally forecast the intricate interplay between minute chemical compounds and proteins. Protein docking, peptide docking, DNA/RNA docking, lipid docking, and carbohydrate docking are just a few examples of how docking is used as a powerful computational tool in bioinformatics to decipher complex connections and binding patterns in macromolecules.

2.10.9 Virtual Screening

Virtual screening is an indispensable computational methodology employed in the realm of bioinformatics to expedite the screening process of vast compound libraries, thereby facilitating the discovery of potential drug candidates. In the realm of bioinformatics, virtual screening is a widely employed technique that leverages the power of docking algorithms to prioritize small molecules based on their affinity towards a specific target receptor.

In the realm of contemporary scientific inquiry, a multitude of cutting-edge computational tools have been harnessed to assess the efficacy and potential of virtual screening methodologies within the realm of pharmaceutical drug discovery. In the realm of bioinformatics, the docking process encounters various challenges that impede its efficacy. These obstacles encompass issues such as incomplete data, flawed comprehension of drug-like molecular characteristics, suboptimal scoring functions, and inadequate docking strategies. The current body of literature indicates that the technology in question is not yet regarded as fully developed or mature within the field of bioinformatics (Dhasmana et al. 2019; Wermuth et al. 2015).

2.10.10 Molecular Dynamics

Molecular dynamics (MD) is a widely employed computational technique in the field of bioinformatics that enables the simulation and analysis of molecular systems by studying the dynamic behavior of molecules and their constituent atoms over a specified time frame (Pagadala et al. 2017). Through MD simulations, intricate details of molecular interactions and their effects can be elucidated, providing valuable insights into the underlying mechanisms governing various biological processes. By leveraging the principles of classical mechanics and statistical physics, MD offers a powerful tool for investigating the structural and functional properties of biomolecules, facilitating the exploration of complex biological phenomena at the atomic level. This bioinformatics approach enables the comprehensive examination of molecular dynamics and intermolecular interactions within a holistic system. In the realm of bioinformatics, the elucidation of system behavior and trajectory

determination is achieved through the utilization of molecular dynamics (MD). This powerful computational technique relies on the application of Newton's equation of motion, coupled with molecular mechanic's methodologies, to estimate the inter-particle forces, commonly referred to as force fields. By harnessing these fundamental principles, MD enables the comprehensive analysis of complex biological systems, shedding light on their dynamic behavior and facilitating a deeper understanding of their intricate molecular interactions (Costa et al. 2019; Alder and Wainwright 1959; Yousif 2020).

2.11 Applications

To better comprehend the three-dimensional structures of biological macromolecules like proteins and nucleic acids, structural bioinformatics employs methods from biology, computer science, and mathematics. In this field, informatics approaches play a crucial role in the analysis and interpretation of structural data. One of the key information, the process of target selection in bioinformatics involves the identification of potential targets through a comprehensive comparison with databases containing both structural and sequence information. By leveraging these databases, researchers can effectively evaluate the suitability of various targets for further investigation and analysis. The determination of a target's significance can be predicated upon an extensive analysis of the existing body of published scientific literature. Target selection can be guided by the identification of protein domains present within the target. Protein domains, the fundamental units of protein structure, possess the remarkable ability to undergo rearrangements, thereby facilitating the generation of novel proteins with diverse functionalities. Isolation of these entities can be undertaken as an initial step in their study (Gong et al. 2011).

The utilization of X-ray crystallography in the field of bioinformatics enables the elucidation of the intricate three-dimensional architecture of proteins. In the realm of bioinformatics, the utilization of X-ray technology for the examination of protein crystals necessitates the prior formation of highly refined and unadulterated protein crystals. This intricate process often entails a substantial number of experimental iterations to achieve the desired outcome. The imperative to monitor the conditions and outcomes of experiments arises from the inherent complexity of scientific investigations, necessitating a comprehensive approach to data management and analysis. Moreover, the utilization of supervised machine learning algorithms enables the analysis of the accumulated data to discern potential factors that could enhance the production of pristine crystals.

The investigation and interpretation of X-ray crystallographic data is a fundamental aspect of bioinformatics research. X-ray crystallography is used to discover the three-dimensional structures of biological macromolecules like proteins and nucleic acids, and then scientists use cutting-edge computer methods and tools to study these structures. The Fourier transform of the electron density distribution can be used to make sense of the diffraction pattern generated by shining X-rays on electrons. The demand for bioinformatics algorithms capable of deconvolving

Fourier transforms in the presence of partial information is evident. This arises from the inherent limitations in phase information, as detectors are only able to measure the amplitude of diffracted X-rays, while the phase shifts remain elusive. The utilization of advanced computational methods, such as the Multiwavelength Anomalous Dispersion technique, holds immense potential in the field of bioinformatics. By employing this technique, researchers can effectively generate electron density maps, which serve as invaluable tools for deciphering the intricate structures of biological macromolecules. In particular, the precise positioning of selenium atoms within these maps serves as a crucial reference point, enabling the accurate determination of the remaining components of the molecular architecture. The generation of the standard Ball-and-stick model involves the utilization of the electron density map.

Nuclear magnetic resonance (NMR) spectroscopy data is a key component of the inquiry and will be analyzed and interpreted. NMR spectroscopy experiments yield multi-dimensional datasets, wherein each discernible peak represents a distinct chemical moiety present within the analyzed sample. The application of optimization techniques is pivotal in the transformation of spectral data into intricate three-dimensional molecular structures.

The integration of structural and functional information is a fundamental aspect of bioinformatics research. By correlating structural data with functional insights, researchers can gain valuable insights into the intricate relationship between a biomolecule's structure and its biological activity. Structural studies serve as powerful tools to probe and elucidate the structural-functional relationship, shedding light on the underlying mechanisms governing molecular function. Through this interdisciplinary approach, bioinformatics researchers strive to unravel the complex interplay between structure and function, ultimately advancing our understanding of the intricate workings of biological systems.

2.12 Tools

2.12.1 List of Structural Bioinformatics Tools

2.12.1.1 I-TASSER (<https://zhanggroup.org/I-TASSER>)

Protein structure prediction involves building a three-dimensional model using the protein's amino acid sequence. Protein 3D models are created from amino acid sequences using the bioinformatics method I-TASSER (Iterative Threading ASSEmblY Refinement). The Protein Data Bank's structure templates are located using a technique called fold recognition (or threading). Replica exchange Monte Carlo simulations are used to reconstruct structural parts derived from threading templates into models of the whole structures. Among the best protein structure prediction algorithms, I-TASSER has shown itself to be in the large-scale community-wide CASP trials. Structural matching of target protein models to the known proteins in protein function databases has allowed researchers to get annotations for ligand binding site, gene ontology, and enzyme commission (Yang et al. 2015). To

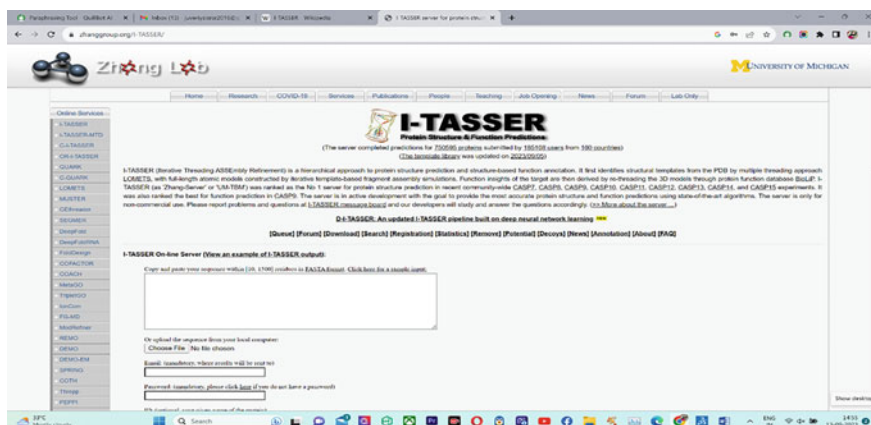


Fig. 2.5 I-TASSER webpage

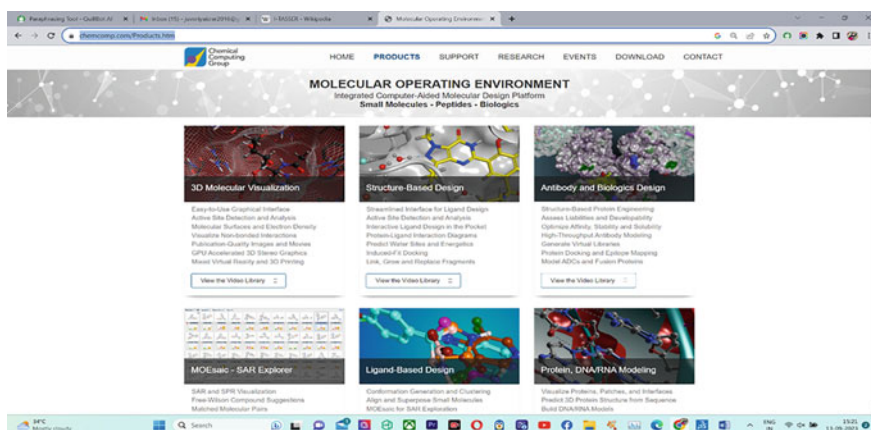


Fig. 2.6 Molecular Operating Environment webpage

help forecast the structures and roles of sequences, the Yang Zhang lab at the University of Michigan in Ann Arbor built a web service. You can download the full version of I-TASSER from the website (Fig. 2.5).

2.12.1.2 Molecular Operating Environment (<https://www.chemcomp.com/Products.htm>)

Modern drug discovery laboratories often use the state-of-the-art Molecular Operating Environment (MOE) platform (Fig. 2.6). Visualization, modeling, simulation, and the creation of new methodologies are only few of the features that are effortlessly included into the whole. MOE allows scientists and researchers to quickly investigate and evaluate molecular structures, predict their behavior, and quicken the pace at which new medicines are developed. In the pharmaceutical,

biotechnology, and academic communities, as well as in the disciplines of biology and medicinal chemistry, the MOE scientific applications are put to good use. MOE is a flexible bioinformatics program that works on a number of different platforms. It works on multiple platforms including macOS, Linux, and Unix (Vilar et al. 2008). MOE's adaptability makes it useful in a number of different areas of bioinformatics. Pharmacophore discovery, structure-based design, fragment-based design, ligand-based design, molecular modeling and simulations, virtual screening, cheminformatics, medicinal chemistry applications, biologics applications, structural biology and bioinformatics, and molecular modeling and simulations are all examples. For the majority of MOE's command, scripting, and application development, Scientific Vector Language (SVL) is the language of choice (Vilar et al. 2008).

2.12.1.3 Structural Bioinformatics Library (<https://sbl.inria.fr>)

SBL, also known as the Structural Bioinformatics Library (Fig. 2.7), is a comprehensive software package that encompasses a wide range of end-user applications and advanced algorithms. Developed specifically for the field of bioinformatics, SBL offers a multitude of tools and resources to aid researchers in their exploration and analysis of structural biology data. With a focus on structural bioinformatics, SBL provides users with a diverse set of applications designed to facilitate various tasks related to the analysis and biological interpretation.

2.12.1.4 BALLView (<https://ball-project.org/ballview/>)

The Biochemical Algorithms Library, or BALL, is a sophisticated and all-encompassing C++ class framework that contains a broad variety of algorithms and data structures that have been developed especially for the purposes of

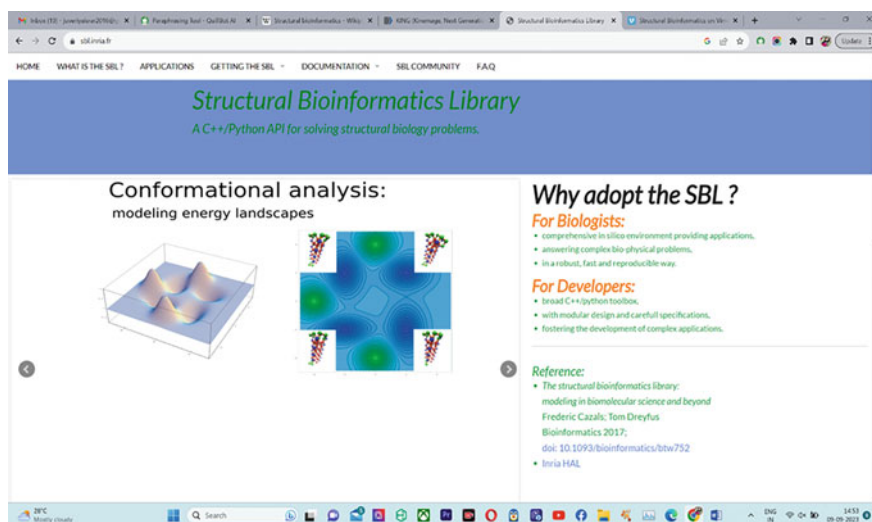


Fig. 2.7 Structural Bioinformatics Library webpage

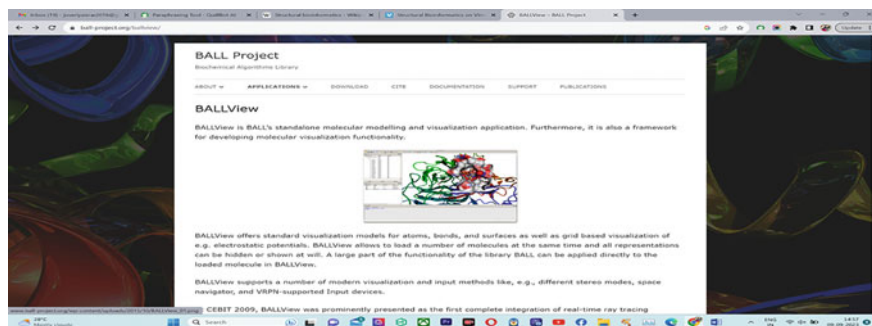


Fig. 2.8 BALLView webpage

molecular modeling and computational structural bioinformatics (Fig. 2.8). This powerful library also offers a Python interface, enabling seamless integration with the Python programming language. Additionally, BALL provides a visually appealing and user-friendly molecule viewer called BALLView, which serves as a graphical user interface for the library (Hildebrandt et al. 2010). With its extensive capabilities and versatile features, BALL is an indispensable tool for researchers and scientists in the field of bioinformatics.

The BALL software has undergone a remarkable transformation, transitioning from a proprietary commercial product to a freely available open-source solution. With this change, the software's license has also been updated; the GNU Lesser General Public License (LGPL) is now in effect. BALLView, an exceptional software tool, proudly operates under the esteemed GNU General Public License (GPL) license. This license, renowned for its commitment to promoting freedom and collaboration, ensures that BALLView remains accessible to all, fostering a vibrant community of bioinformatics enthusiasts. With its powerful features and user-friendly interface, BALLView empowers researchers to visualize and analyze complex biological data with unparalleled precision and efficiency. By embracing the GPL license, BALLView exemplifies the spirit of open-source bioinformatics, enabling scientists.

The widely utilized bioinformatics software tools, BALL and BALLView, have been successfully adapted and made compatible with various operating systems including Linux, macOS, Solaris, and Windows. This extensive porting effort ensures that researchers and scientists across different computational environments can seamlessly leverage the functionalities and capabilities offered by these powerful tools (Nickels et al. 2013).

BALLView is an advanced molecular visualization tool that has been meticulously crafted by the esteemed BALL project team. This cutting-edge software is implemented in C++ and leverages the power of Qt and OpenGL, while employing the remarkable real-time ray tracer RTFact as its rendering back-ends. BALLView is a cutting-edge software tool that provides advanced capabilities for three-dimensional and stereoscopic visualization in various modes. It seamlessly

integrates with the powerful algorithms of the BALL library, allowing users to leverage its functionalities through an intuitive graphical user interface. With BALLView, researchers and bioinformaticians can effortlessly explore complex molecular structures and gain valuable insights into their data.

Acclaimed research groups from the illustrious Saarland University, Mainz University, and University of Tübingen have painstakingly developed and diligently maintained the BALL project, a ground-breaking endeavor in the field of bioinformatics. In the fields of learning and discovery, the library and the viewer are vital resources that make the gathering and processing of knowledge possible. Users now have easy access to this powerful bioinformatics tool thanks to the Debian project's incorporation of BALL packages into its repository.

2.12.1.5 PyMOL (<https://pymol.org/2/>)

Warren Lyford DeLano's cutting-edge molecular visualization system, PyMOL, uses open source technology and his proprietary expertise. PyMOL helps scientists visualize and analyze complex molecular structures with its advanced features and user-friendly interface. PyMOL helps users understand biomolecules, leading to bioinformatics breakthroughs. Pioneering private software company DeLano Scientific LLC commercialized this innovative technology. Developing cutting-edge tools with broad accessibility for scientific and educational communities, DeLano Scientific LLC helped advance this advancement (Yuan et al. 2017). Schrödinger, Inc., a bioinformatics giant, commercializes this technology. The permissive software license was removed. Later software is distributed under a custom license instead of the Python license. This custom license grants extensive use, redistribution, and modification rights while transferring copyright ownership to Schrodinger, LLC. Note that some source code is no longer available. PyMOL, a powerful bioinformatics tool, creates visually appealing three-dimensional representations of chemical compounds and complex biological macromolecules like proteins (Rosignoli and Paiardini 2022). The primary author claims that PyMOL, a popular software tool, gained popularity in bioinformatics by 2009. Nearly 25% of 3D protein structure images in scientific literature were created using PyMOL. PyMOL is a popular structural biology model visualization tool (Fig. 2.9). It stands out among the few open-source software options in this domain. The "Py" prefix indicates that the software is written in Python. PyMOL, a versatile molecular visualization software, uses GLEW and Free GLUT. PyMOL performs well in solving complex Poisson-Boltzmann equations using the Adaptive Poisson Boltzmann Solver. No user text is provided. PyMOL, a powerful bioinformatics tool, uses Tk for its GUI widgets. Schrödinger provided macOS Aqua binaries. With version 2.0, PyMOL switched to PyQt, ensuring a consistent experience across platforms. No user text is provided.

2.12.1.6 Visual Molecular Dynamics (<https://www.ks.uiuc.edu/Research/vmd/>)

Visual Molecular Dynamics (VMD) is a cutting-edge bioinformatics software application that serves as a powerful tool for molecular modeling and visualization (Fig. 2.10). Developed specifically for the purpose of analyzing complex molecular

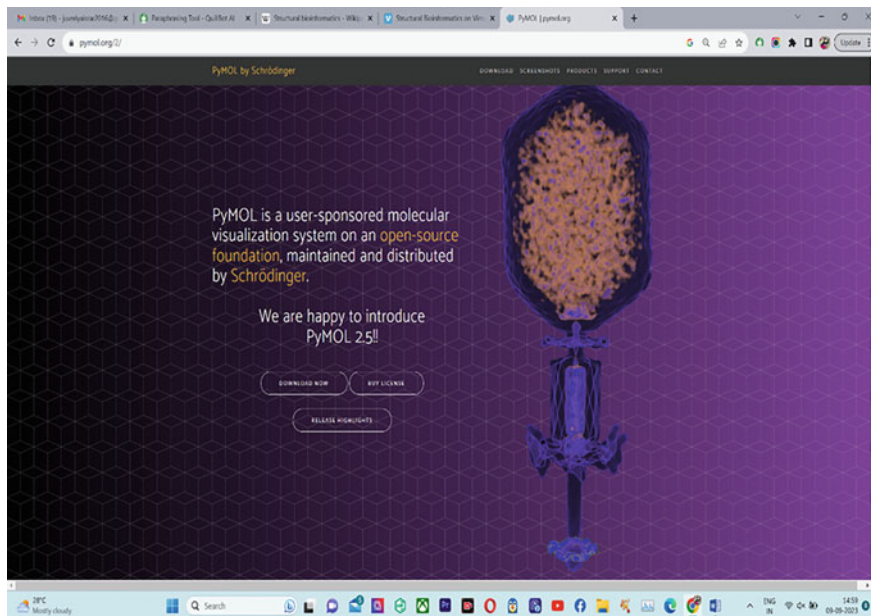


Fig. 2.9 PyMoL webpage

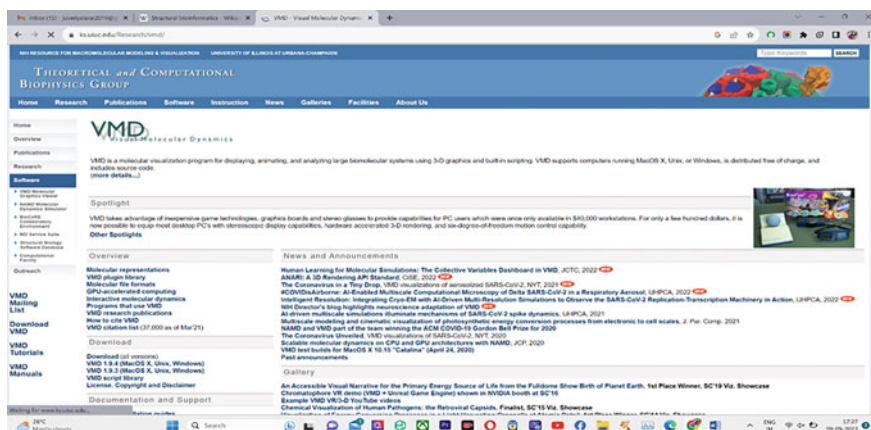


Fig. 2.10 Visual Molecular Dynamics webpage

systems, VMD enables researchers to gain valuable insights into the intricate structures and dynamic behaviors of biomolecules (Mackoy et al. 2021). By employing advanced computational algorithms and sophisticated graphical rendering techniques, VMD empowers scientists to explore and manipulate molecular data with unparalleled precision and clarity. With its user-friendly interface and extensive range of features, VMD has become an indispensable resource in the field of bio

VMD serves as a prominent bioinformatics software application primarily designed for the visualization and comprehensive analysis of molecular dynamics simulation outcomes. The bioinformatics field encompasses a wide range of tools and techniques that facilitate the analysis and manipulation of various types of data. This includes the handling of volumetric data, such as three-dimensional structures, as well as the processing of sequence data, such as DNA or protein sequences. Additionally, bioinformatics tools are designed to work with arbitrary graphics objects, enabling researchers to visualize and interpret complex biological data in a meaningful way. It is common practice in the bioinformatics profession to export molecular sceneries to external rendering programs like POV-Ray, Render Man, Tachyon, Virtual Reality Modeling Language (VRML), and many more. VMD, a versatile molecular visualization software, empowers users to execute personalized Tcl and Python scripts seamlessly. This capability is facilitated by the inclusion of embedded Tcl and Python interpreters within the VMD framework. VMD may be used on a variety of platforms, including Unix, macOS, and Windows, making it a highly flexible tool for molecular visualization. Through a distribution-specific license, VMD, a potent piece of molecular visualization software, is made available to users who are not profiting from its use. This license grants users the freedom to utilize the program and make modifications to its source code, all without incurring any charges.

2.12.1.7 KiNG (<http://kinemage.biochem.duke.edu/software/king/>)

Structural biology uses fast, flexible, and customized visualization software to understand biological macromolecules' complex structure and dynamic function. Researchers can better understand complex molecular dynamics and behaviors using bioinformatics visualizations. These software applications must display three-dimensional annotations of model errors or significant interaction sites alongside the structural depiction to be effective.

The Java-based, modular, and extensible scientific visualization tool KiNG (Kinemage, Next Generation) focuses on macromolecular visualization. KiNG, a versatile molecular visualization software, is similar to PyMOL, SwissPdbViewer, Chimera, RasMol, and JMol (Fig. 2.11). These programs provide the means for real-time manipulation and exploration of molecular structures in three dimensions. KiNG's dynamic 3D rotation, translation, cropping, and zooming aids in comprehending molecular depth perception and spatial interactions (Chen et al. 2009). KiNG's molecule-agnostic kinemage graphics format stands out with its versatile color palette, advanced depth cueing, and extensive tools and features. This distinguishes KiNG from other bioinformatics software.

KiNG is state-of-the-art software that builds on the past three decades of progress in molecular graphics, especially in the area of protein ribbon diagrams. It proudly stands on the shoulders of Mage, the pioneering kinemage graphics program that inspired it. Mage, a front-line bioinformatics tool, was carefully designed to help create stunning and accurate molecular illustrations. Its flexible functionality lets researchers and educators easily add captivating visuals to journal articles and classroom materials. Mage's innovative features and user-friendly interface enable



Fig. 2.11 KiNG webpage

users to visually communicate complex molecular concepts, improving scientific communication. Mage quickly became essential to the lab’s research program due to its adaptability. Reimagining kinemage functionality from scratch in the KiNG framework produced a modern user interface that resembles its predecessor. It has also simplified the data structure, making maintenance and expansion easier. KiNG, an essential part of the lab’s research, leads their scientific efforts.

Mage and KiNG’s bioinformatics collaborations have led to different development paths. The simultaneous development of two kinemage viewers has improved both software applications. As mentioned in the “Results and Discussion” section, the KiNG and Mage software platforms initially integrated high-dimensional visualization techniques for specific purposes. We created a robust, versatile, and enhanced functionality through collaboration and synergistic integration of diverse implementations.

Bioinformatics becomes more versatile and adaptable by decoupling molecular information, such as PDB files, from its visual representation using KiNG and Mage. In bioinformatics, “7” provides no context or information to rewrite. Secondary annotations such as helix axis, local validation outliers, and interface contact dots can be seamlessly incorporated into main structural data such as models, ribbons, electron density, and NMR using the proposed method. Additionally, this strategy allows for fully non-molecular visualizations in the same computational tool.

Bioinformatics enthusiasts can find many examples and format documentation at <http://kinemage.biochem.duke.edu>. The “Materials and Methods” section can help interested parties understand the topic. Kinemage is a versatile and efficient plain text format for bioinformatics manual editing and program generation.

KiNG’s adaptability is especially useful when the kinemage format is rigid. The runtime-loaded Java plug-in modules enhance existing features. The flexible graphics engine can be used in a new computational framework. To enable high-dimensional analysis, this study uses protein reconstruction plug-ins and molecular visualization tools while improving the core software. KiNG, a novel bioinformatics tool, can quickly develop and integrate new modules, increasing its flexibility.

2.12.1.8 STRIDE (Algorithm) (<https://webclu.bio.wzw.tum.de/cgi-bin/stride/stridecgi.py>)

Structural Identification (STRIDE), a sophisticated bioinformatics tool, uses atomic coordinates from cutting-edge methods like X-ray crystallography, protein NMR, and other protein structure determination methods (Fig. 2.12). STRIDE accurately assigns secondary structure elements to proteins using a robust algorithm, revealing their structural organization. STRIDE's ability to decipher a protein's intricate atom arrangement helps us understand protein structure-function relationships and facilitate bioinformatics analyses (Matarazzo and Pakzad 2014). In bioinformatics, dihedral angle potentials in hydrogen bond criteria improve the DSSP algorithm in the STRIDE framework. This method uses more complicated secondary structure definition criteria than the popular DSSSP algorithm. The STRIDE energy function includes a distance-dependent 8–6 potential for the hydrogen-bond term, which is inspired by the work of Lennard-Jones. The optimal planarity of the hydrogen bond geometry is captured by integrating two angular dependency components. Like DSSP, this method relies on empirical examinations of solved structures from the Protein Data Bank that have had their secondary structure elements visually assigned, and then uses statistical likelihood factors to identify these elements. One of the first and most used bioinformatics tools is the Dictionary of Secondary Structure of Proteins (DSSP). DSSP remains the most popular structural assignment method despite its age. However, the original definition of STRIDE, another popular method, claimed to outperform DSSP in at least 70% of structural assignments. In the DSSP method, shorter secondary structures are often assigned than by expert crystallographers. The STRIDE algorithm has been improved to address this issue. Minor local structural variations near secondary structure element termini cause this discrepancy. To address this issue, STRIDE has been improved to predict secondary structures more accurately. The bioinformatics-popular STRIDE and DSSP algorithms agree 95.4% of the time. A sliding-window technique reduces single-terminal residue assignment discrepancies, achieving this agreement. No user text is provided. STRIDE and DSSP may underestimate secondary structure elements like pi helices.



Fig. 2.12 STRIDE (algorithm) webpage

2.12.1.9 DSSP (Algorithm) (<https://www.blopig.com/blog/2014/08/dssp/>)

The DSSP algorithm (Fig. 2.13), widely recognized as the gold standard in the field of bioinformatics, serves as the primary tool for the precise determination of secondary structure elements within protein sequences. This system efficiently assigns secondary structure annotations to individual amino acids using atomic-resolution protein coordinates, providing researchers with a wealth of information about the structural organization and functional features of proteins (Sekihara et al. 2016). The acronym, mentioned singularly within the confines of the 1983 publication, delineates the nomenclature assigned to the Pascal software application responsible for executing the algorithm denoted as Define Secondary Structure of Proteins.

2.12.1.10 MolProbity (<http://molprobity.manchester.ac.uk/>)

MolProbity is an essential web-based tool that validates the quality of complicated 3D structures, such as those of proteins, nucleic acids, and complexes. The software provides an exhaustive examination of all atom interactions, which aids in the detection of steric hindrances in molecular structures (Fig. 2.14). In addition, it can calculate and display hydrogen bond and van der Waals interactions between molecules at their interfaces. Polar and non-polar hydrogen atoms must be included and refined thoroughly as part of the aforementioned procedure (Williams et al. 2018). The KiNG viewer presents the results in a number of formats, including numerical scores, lists, downloadable PDB and graphics files, and, most crucially, online interactive 3D kinemage images. The aforementioned service is provided at no cost to users and can be accessed at <http://kinemage.biochem.duke.edu>.

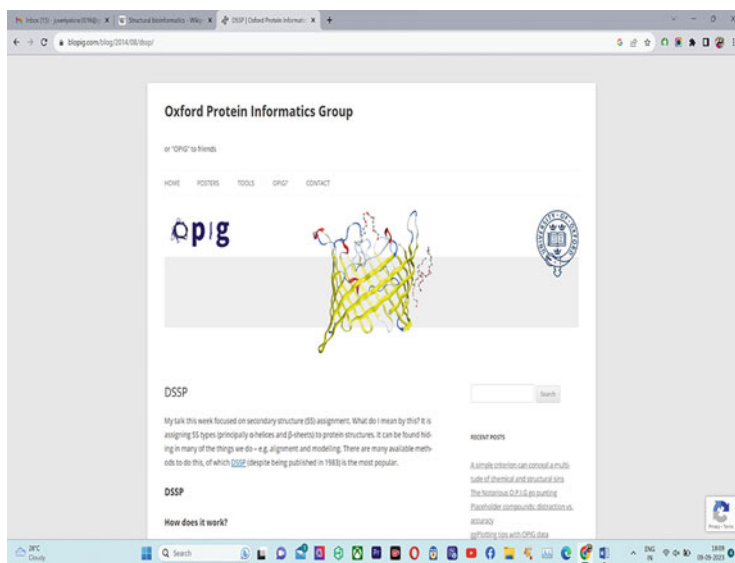


Fig. 2.13 DSSP (algorithm) webpage

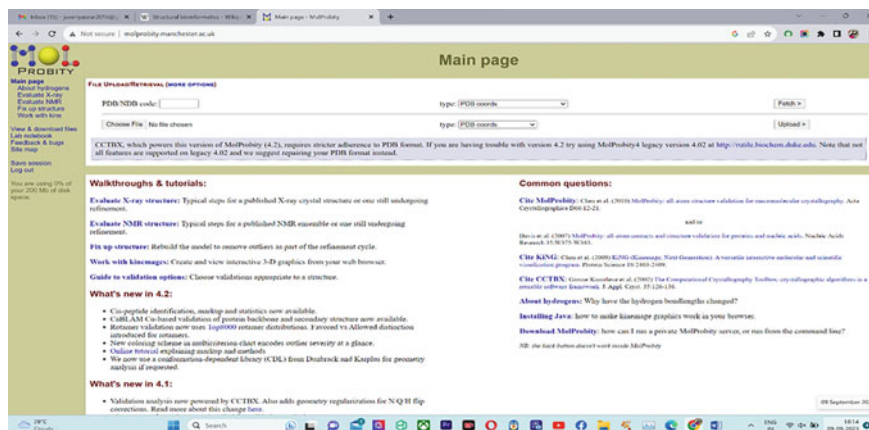


Fig. 2.14 MolProbity (algorithm) webpage

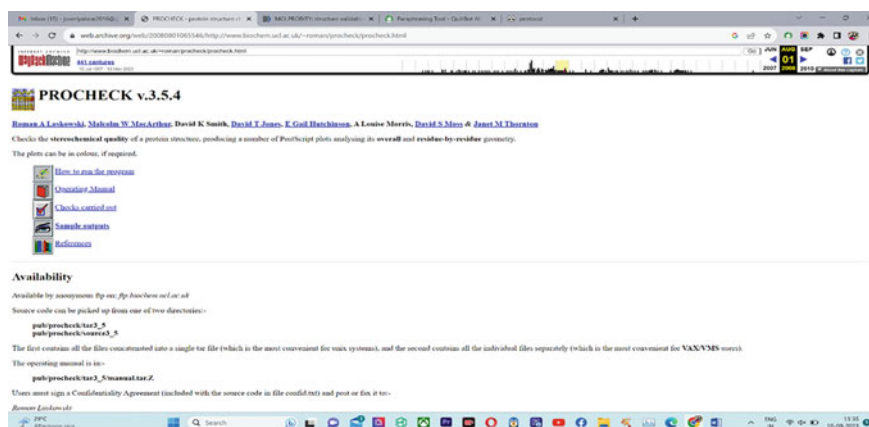


Fig. 2.15 PROCHECK webpage

2.12.1.11 PROCHECK (<https://web.archive.org/web/20080801065546/http://www.biochem.ucl.ac.uk/~roman/procheck/procheck.html>)

PROCHECK is an all-inclusive suite of separate Fortran and C programs that are run in order by a shell script (Fig. 2.15). The initial step of the computational pipeline involves the preprocessing of the input PDB file. This entails the reassignment of specific side-chain atoms in accordance with the established IUPAC naming conventions as outlined by the IUPAC–IUB Commission on Biochemical Nomenclature in 1970. Subsequently, an exhaustive analysis of the protein’s stereo chemical parameters is performed, allowing for a comprehensive comparison against established norms. At the end of the process, the pipeline produces a visually beautiful PostScript output in addition to a meticulously detailed summary of the

protein's structural features, residue by residue. The exclusion of hydrogen atoms and atoms with zero occupancy is a standard practice in bioinformatics analyses. When atoms can take on multiple shapes, only the one with the highest occupancy rate is taken into account (Yao and Cao 2023).

The comprehensive collection of program source codes can be accessed at the esteemed web address <http://www.biochem.ucl.ac.uk/roman/procheck/procheck.html>. The software in question, which has been integrated into the CCP4 suite of programs, was developed as part of the Collaborative Computational Project, Number 4 in 1994. More information about the CCP4 suite can be found at <http://www.dl.ac.uk/CCP/CCP4/main.html>. Additionally, users have the option to directly access and utilize the software through the Biotech Validation Server, which is available at <http://biotech.embl-ebi.ac.uk:8400/>.

2.12.1.12 CheShift (<http://www.cheshift.com/>)

The innovative bioinformatics tool CheShift-2 revolutionizes the computation of $^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$ protein chemical shifts and offers vital insights into protein structure validation (Fig. 2.16). CheShift-2's cutting-edge algorithms and methods help researchers and scientists understand protein chemistry's complicated intricacies. The study analyzes $^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$ chemical shift patterns in connection to the torsional angles (φ , ψ , ω , and χ_1 , χ_2) of 20 amino acids using quantum mechanics simulations (Vila et al. 2009).

Bioinformatics tool CheShift-2 analyzes PDB protein structures to gain insights. CheShift-2 generates a complete collection of theoretical chemical shift values using advanced algorithms. Researchers use this knowledge to understand protein behavior and structural and functional qualities. CheShift-2 helps bioinformaticists analyze and interpret protein structures quickly and accurately with its correct predictions (Martin et al. 2012). A supplied PDB file and chemical shift values enable three-dimensional protein model display in the software. The 3D protein



Fig. 2.16 CheShift webpage

model's five-color code shows the differences between anticipated and experimental chemical shift values. Using the discrepancies between experimentally measured and projected $^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$ chemical shifts can reveal probable abnormalities in protein structures. A strong bioinformatics method, CheShift-2, uses $^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$ chemical shifts to identify alternate χ_1 and χ_2 side-chain torsional angles. CheShift-2 generates a complete list of probable torsional angles by minimizing chemical shift discrepancies. These insights can improve protein structure quality and accuracy by fixing flaws.

The official website <http://www.cheshift.com> makes CheShift-2, a cutting-edge bioinformatics tool, accessible online. This powerful program can also be smoothly incorporated into PyMOL, a popular molecular visualization platform, via a plugin.

2.12.1.13 3Dmol.js (<https://3dmol.csb.pitt.edu/>)

3Dmol.js uses WebGL (Fig. 2.17), a cutting-edge technology, to create spectacular and interactive molecular images on web platforms. This JavaScript module makes real-time molecular structure exploration easy via hardware acceleration. 3Dmol.js helps bioinformatics researchers understand molecular biology with its powerful features (Rego and Koes 2015). Many different types of molecular data files and presentation formats are supported by the software. These include volumetric data like cube files and simulation data like AMBER or GROMACS data (Shkurti et al. 2016). A rich JavaScript API lets users alter molecular structures with 3Dmol.js, a sophisticated bioinformatics application. Additionally, its embedding API lets molecular views be seamlessly integrated into web pages using a concise div declaration. A hosted viewer API from 3Dmol.js uses URLs to easily retrieve and visualize molecular data. The introduced observer, <http://3dmol.csb.pitt.edu/viewer.html>, has been carefully designed to include all the data needed for molecular visualization in the URL. In bioinformatics, a captivating scenario can be created

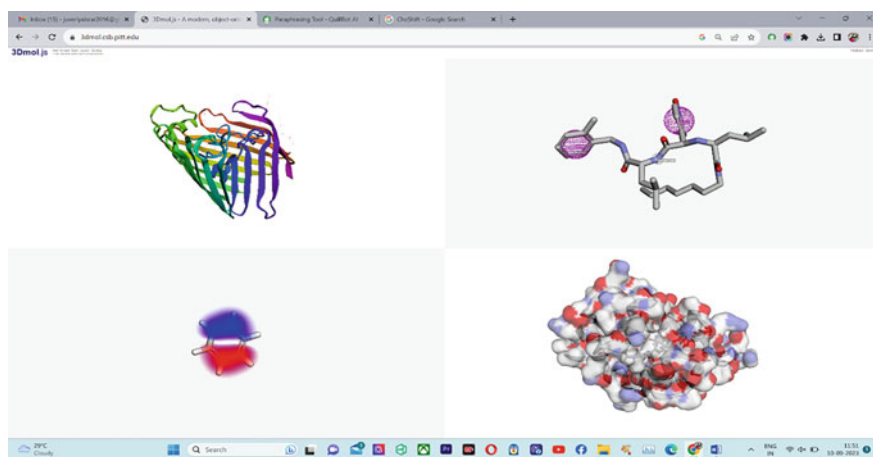


Fig. 2.17 3Dmol.js webpage

by carefully importing molecular data and applying styles using the edit panel. Next, pupils can be given the URL of this carefully designed scene, perhaps as a QR code. This allows students to actively participate in the same scene, making learning lively and participatory. The viewer can load molecular data using a PubChem compound id to ensure a 3D structure. An externally hosted file or PDB identification can likewise be used. By extending the functionality of the 3Dmol.js API and hosted viewer, a dynamic learning environment has been established. In this interactive whiteboard, students react to questions by selecting molecules in three dimensions.

2.12.1.14 PROPKA (<https://web.archive.org/web/20070113065659/http://propka.ki.ku.dk/>)

PROPKA predicts protein ionizable residue pK(a) values by using non-proteinaceous ligands and their ionizable group pK(a) values (Fig. 2.18) (Saoudi et al. 2011). PROPKA 2.0 extensively uses 1.0's empirical criteria for ligand functional groups. Due to its speed, PROPKA can calculate the pK(a) values of all ionizable groups in seconds for most proteins. Several protein-ligand complexes are explored, comparing PROPKA 2.0 predictions to experimental results. This complex contains trypsin, thrombin, three pepsins, HIV-1 protease, chymotrypsin, xylanase, hydroxynitrile lyase, and dihydrofolate reductase. Four of the 14 trypsin-thrombin ligand complexes have considerable protonation state changes ($lnI > 0.5$). PROPKA 2.0 and Klebe's PEOE method show a 0.4-unit protonation shift at pH 6.5 and 7.0 when plasmin II, cathepsin D, and endothiapepsin bind to pepstatin. PROPKA 2.0 data shows that ligand binding alters structure effect proton uptake/release and residues away from the binding site. The residues' surroundings and hydrogen bonding network have changed, generating these alterations. PROPKA 2.0 can quickly and correctly forecast the protonation states of critical residues and ligand functional groups at a protein's binding or active site

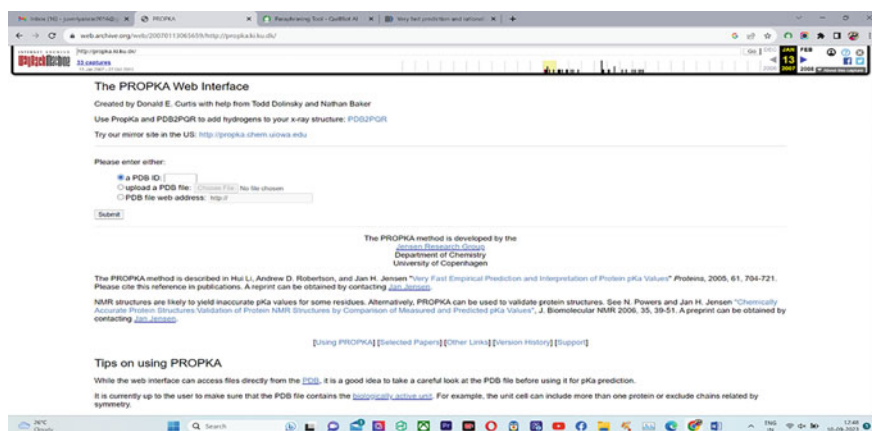


Fig. 2.18 PROPKA webpage

by describing protein-ligand interactions that modify titratable groups' pK(a) values (Saoudi et al. 2011).

2.12.1.15 CARA (<http://cara.nmr.ch/doku.php>)

CARA (Computer Aided Resonance Assignment) (Fig. 2.19), a cutting-edge bioinformatics tool for structural biology resonance assignment. CARA transforms nuclear magnetic resonance (NMR) data analysis and interpretation with its powerful algorithms and user-friendly interface. Visit our website to learn about CARA's remarkable features and join the expanding community of scientists who use it for structural elucidation. Advanced bioinformatics software CARA is powerful. Its main use is NMR spectrum analysis and computer-aided resonance assignment (Bosso et al. 2017). CARA is helpful in molecular and structural biology due to its concentration on bio macromolecules. Researchers and scientists use NMR spectroscopy to understand biomacromolecular structures, and its powerful capabilities and specific features make it indispensable. Structure determination is aided by dedicated software for identifying backbones, assigning side chains, and integrating peaks. These devoted tools simplify biomolecule structure elucidation steps.

2.12.1.16 Docking Server (<https://www.dockingserver.com/web>)

Docking Server handles ligand and protein setup and molecular docking with a web-based, easy-to-use interface (Fig. 2.20). Docking Server's user-friendly interface lets researchers from all biochemistry fields calculate and evaluate docking results, but advanced users can set ligand and protein parameters and docking calculations (Yu et al. 2016). The app can dock and analyze single ligands and dock ligand libraries to target proteins at fast speed. Parameters for optimizing ligand shape, minimizing energy, calculating charges, docking molecules, and representing protein-ligand complexes are all precisely calculated using Docking Server's computational chemistry program. Thus, Docking Server combines a number of popular in silico chemistry products into a single, all-encompassing web service, allowing for extremely fast and accurate docking computations.

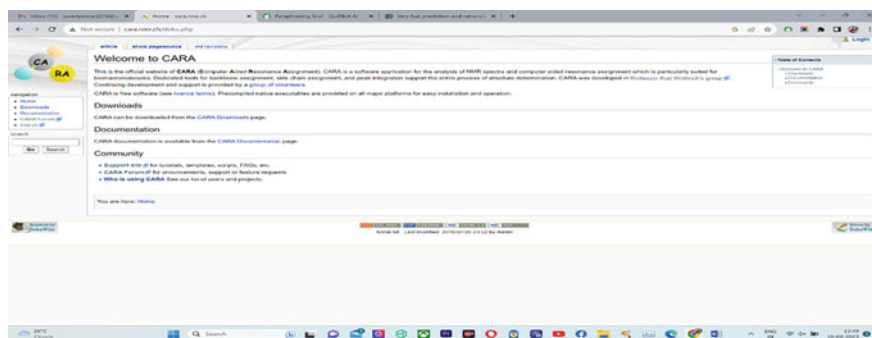


Fig. 2.19 CARA webpage

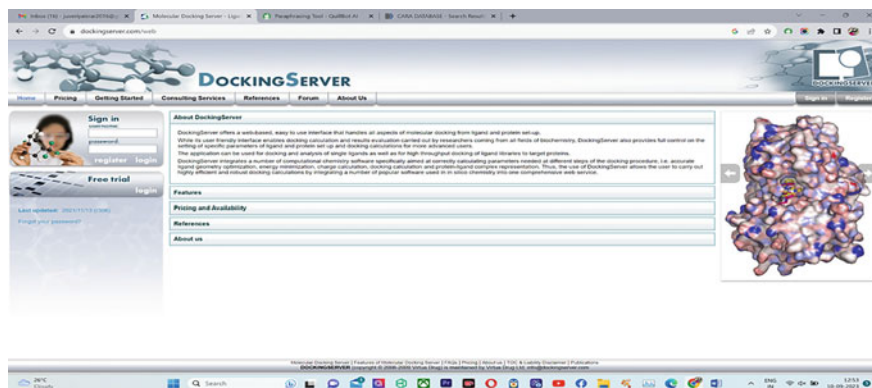


Fig. 2.20 Docking server webpage

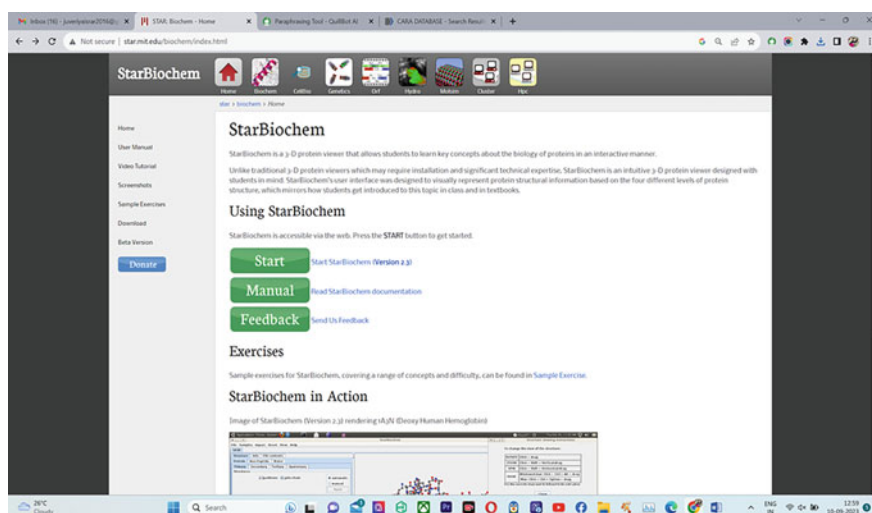


Fig. 2.21 StarBiochem webpage

2.12.1.17 StarBiochem (<http://star.mit.edu/biochem/>)

Star Biochem (Fig. 2.21), a forefront bioinformatics tool, helps researchers understand protein fundamentals. With three-dimensional visualization, this cutting-edge protein viewer lets students interact and immerse themselves in biological topics. Star Biochem helps students grasp protein structures and functions with its intuitive UI and powerful capabilities.

The creative and user-friendly Star Biochem 3-D protein viewer serves students. Star Biochem unlike ordinary viewers requires no complex setups or technical expertise. Its user-friendly interface makes protein structure visualization easy. The Star Biochem user interface was carefully designed to present protein structural data

in accordance with classroom and textbook pedagogy. It smoothly unifies the four levels of protein structure, making it easy to understand.

2.12.1.18 SPADE (Structural Proteomics Application Development Environment) (<https://sites.google.com/view/spade>)

SPADE visualizes and studies molecular structures on multiple platforms. SPADE's various functionalities aid protein structure research by academics and bioinformaticians (Fig. 2.22). This advanced tool enables users quickly explore, alter, and analyze structural data to comprehend complex biological systems. SPADE shows protein three-dimensional architecture and dynamic behavior to help bioinformaticists address molecular challenges (Manjasetty et al. 2012). I enjoy Structure Prediction and Design Engine, a protein engineering innovation. Its innovative design may help structural biology algorithm developers.

Bioinformatics tool SPADE has a simple UI and apps. The powerful evolutionary calculating visualization tool Sequence Pad is one. Researchers can precisely visualize protein-protein interactions with SequencePad (Li et al. 2017). SPADE helps researchers understand complex molecular interactions. RAVE is a sophisticated bioinformatics tool for chemically probing projected structure models for experimental validation. Researchers can swiftly and methodically test computational predictions with RAVE's numerous functions. RAVE studies molecular structures at unprecedented depth utilizing cutting-edge algorithms and data processing. RAVE uses experimental data and computational simulations to understand complex biological systems and uncover new treatment targets faster. SPADE is a powerful computer platform for Molnir, a genetic algorithm-driven hybrid protein structure modeling tool. Programmers in bioinformatics use many computational tools. These technologies improve accuracy and efficiency through numerous means. Calculators assess biomolecular solvent and surface accessibility. Multi-feature dynamic programming easily aligns and compares large biological

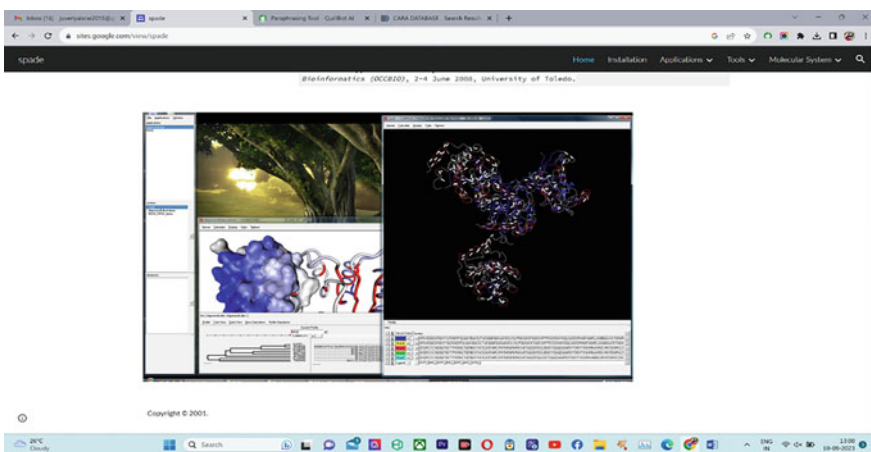


Fig. 2.22 SPADE webpage

sequences. Hydrogen bond calculators let programmers measure hydrogen atom–molecular interactions. Bioinformatics programmers can clearly comprehend biological systems with these essential components. Using algorithms, bioinformatics interprets biological data, including molecular structures. Unconstrained multiple structure alignment that supports distantly linked structures is a big development in this field. Different structures are aligned by this innovative method. It may freely align structurally diverse molecules, showing conserved areas and functional patterns in many biological units. The algorithm’s remote structure handling. The massive biological data set awaits discovery.

2.13 Conclusion, Future prospective and challenges

Structural bioinformatics uses computer methods to study biological molecules’ complex arrangements and functions, making it dynamic and progressive. The study of structural bioinformatics has led to several important findings and exciting prospects. Computational and experimental methods have revealed biomolecules’ three-dimensional structures, improving our understanding of their functions and interactions. These advances have enabled new drug discovery, protein engineering, and molecular design methods. Machine learning algorithms and artificial intelligence have changed protein structure prediction, making it highly accurate. As structural bioinformatics advances, biological complexity can be unraveled.

Numerous bioinformatics triumphs have marked successes enabled over the past decade, structural bioinformatics has advanced greatly. These advances have helped predict protein structures, understand complex protein–protein interactions, and produce new drugs. The tremendous growth in processing capability, smart algorithms, and growing data have made these astonishing discoveries possible.

Several bioinformatics fields are set for significant advances. These include improving protein structure prognostication, using machine learning to analyze large datasets, and combining structural and functional data to better understand complex biological systems.

While structural bioinformatics has made significant progress, it still faces several obstacles in understanding biomolecular structures. In bioinformatics, protein structures are difficult to predict, especially for large, complex proteins. Researchers in bioinformatics must constantly focus on many issues. Integrating data from several sources is a difficult task. Harmonizing information allows a holistic perspective of biological phenomena. To improve molecular simulations, more accurate force fields are needed. Simulating molecular activity and interactions with these force fields illuminates their complex dynamics and functions. Big data presents another challenge: efficient algorithms. The exponential rise of datasets requires new computational methods that can quickly and effectively analyze massive amounts of data. The search for more efficient algorithms aims to reveal patterns and insights in these massive datasets, improving our understanding of biological systems. In

conclusion, bioinformatics must integrate varied data sources, optimize molecular simulations with precise force fields, and build efficient algorithms to evaluate enormous information. These issues motivate bioinformatics research and innovation.

Structured bioinformatics offers several opportunities for biological researchers to make significant contributions. Combining computational and experimental methods helps scientists understand biological macromolecules' complex architecture and dynamic activity. This synergistic strategy has great potential for uncovering their mechanisms of action, enabling the discovery and design of novel pharmacological drugs and therapeutic interventions for a variety of diseases.

In conclusion, structural bioinformatics has advanced biological molecule investigation to new heights, with good prospects for future growth. Bioinformatics researchers must constantly develop new methods and algorithms to overcome several hurdles. Despite these challenges, structural bioinformatics offers many opportunities for researchers to make biological science advances.

References

- Alder BJ, Wainwright TE (1959) Studies in molecular dynamics. I. General method. *J Chem Phys* 31(2):459–466. Bibcode:1959JChPh, 31.459A. ISSN 0021-9606. <https://doi.org/10.1063/1.1730376>
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. *Nucleic Acids Res* 28(1):235–242. <https://doi.org/10.1093/nar/28.1.235>. PMID: 10592235; PMCID: PMC102472
- Bosso A, Pirone L, Gaglione R, Pane K, Del Gatto A, Zaccaro L, Di Gaetano S, Diana D, Fattorusso R, Pedone E, Cafaro V (2017) A new cryptic host defense peptide identified in human 11-hydroxysteroid dehydrogenase-1 β -like: from *in silico* identification to experimental evidence. *Biochim Biophys Acta Gen Subj* 1861(9):2342–2353
- Chen VB, Davis IW, Richardson DC (2009) KING (Kinemage, next generation): a versatile interactive molecular and scientific visualization program. *Protein Sci* 18(11):2403–2409
- Chothia C, Lesk AM (1986) The relation between the divergence of sequence and structure in proteins. *EMBO J* 5(4):823–826. <https://doi.org/10.1002/j.1460-2075.1986.tb04288.x>. PMC 1166865. PMID 3709526
- Costa LS, Mariano DC, Rocha RE, Kraml J, Silveira CH, Liedl KR et al (2019) Molecular dynamics gives new insights into the glucose tolerance and inhibition mechanisms on β -glucosidases. *Molecules* 24(18):3215. <https://doi.org/10.3390/molecules24183215>. PMC 6766793. PMID 31487855
- da Silveira CH, Pires DEV, Minardi RC, Ribeiro C, Veloso CJM, Lopes JCD, Meira W, Neshich G, Ramos CHI, Habesch R, Santoro MM (2009) Protein cutoff scanning: a comparative analysis of cutoff dependent and cutoff free methods for prospecting contacts in proteins. *Proteins Struct Funct Bioinform* 74(3):727–743. <https://doi.org/10.1002/prot.22187>
- Dhasmana A, Raza S, Jahan R, Lohani M, Arif JM (2019) Chapter 19—high-throughput virtual screening (HTVS) of natural compounds and exploration of their biomolecular mechanisms: an *in silico* approach. In: Ahmad Khan MS, Ahmad I, Chattopadhyay D (eds.) *New look to phytomedicine*. Academic, pp 523–548. <https://doi.org/10.1016/b978-0-12-814619-4.00020-3>. isbn:978-0-12-814619-4. S2CID 69534557
- Gauthier J, Vincent AT, Charette SJ, Derome N (2019) A brief history of bioinformatics. *Brief Bioinformatics* 20(6):1981–1996

- Gong S, Worth CL, Cheng TM, Blundell TL (2011) Meet me halfway: when genomics meets structural bioinformatics. *J Cardiovasc Transl Res* 4:281–303
- Gu J, Bourne PE (2011) *Structural bioinformatics*. Wiley. Gu J, Bourne PE (2009-03-16). *Structural bioinformatics*. Wiley. 978-0-470:18105-8
- Hildebrandt A, Dehof AK, Rurainki A, Bertsch A, Schumann M, Toussaint NC, Moll A, Stöckel D, Nickels S, Mueller SC, Lenhof HP (2010) BALL-biochemical algorithms library 1.3. *BMC Bioinformatics* 11(1):1–5
- Ilyin VA, Abyzov A, Leslin CM (2004) Structural alignment of proteins by a novel TOPOFIT method, as a superimposition of common volumes at a topomax point. *Protein Sci* 13(7): 1865–1874. <https://doi.org/10.1110/ps.04672604>. PMC 2279929. PMID 15215530
- Kaczanowski S, Zielenkiewicz P (2010) Why similar protein sequences encode similar three-dimensional structures? *Theor Chem Accounts* 125(3–6):643–650. <https://doi.org/10.1007/s00214-009-0656-3>. issn:1432-881X. S2CID 95593331
- Klebe G (2015) Protein-ligand interactions as the basis for drug action. In: Scapin G, Patel D, Arnold E (eds) *Multifaceted roles of crystallography in modern drug discovery*. NATO science for peace and security series a: chemistry and biology. Springer, Dordrecht, pp 83–92. https://doi.org/10.1007/978-3-642-17907-5_4. isbn:978-3-642-17906-8
- Kocincová L, Jarešová M, Byška J, Parulek J, Hauser H, Kozlíková B (2017) Comparative visualization of protein secondary structures. *BMC Bioinformatics* 18:23. <https://doi.org/10.1186/s12859-016-1449-z>. PMID: 28251875; PMCID: PMC5333176
- Li H, Chang YY, Lee JY, Bahar I, Yang LW (2017) DynOmics: dynamics of structural proteome and beyond. *Nucleic Acids Res* 45(W1):W374–W380
- Mackoy T, Kale B, Papka ME, Wheeler RA (2021) View Sq, a visual molecular dynamics (VMD) module for calculating, analyzing, and visualizing X-ray and neutron structure factors from atomistic simulations. *Comput Phys Commun* 264:107881
- Manjasetty BA, Büssov K, Panjekar S, Turnbull AP (2012) Current methods in structural proteomics and its applications in biological sciences. *3 Biotech* 2:89–113
- Mariano DC, Santos LH, Machado KD, Werhli AV, de Lima LH, de Melo-Minardi RC (2019) A computational method to propose mutations in enzymes based on structural signature variation (SSV). *Int J Mol Sci* 20(2):333. <https://doi.org/10.3390/ijms20020333>. PMC 6359350. PMID 30650542.
- Martin OA, Vila JA, Scheraga HA (2012) Che Shift-2: graphic validation of protein structures. *Bioinformatics* 28(11):1538–1539
- Martins PM, Mayrink VD, de Silveira S, da Silveira CH, de Lima LH, de Melo-Minardi RC (2018) How to compute protein residue contacts more accurately? Proceedings of the 33rd annual ACM symposium on applied computing. Pau: ACM Press, pp 60–67. isbn:978-1-4503-5191-1. S2CID 49562347 <https://doi.org/10.1145/3167132.3167136>
- Matarazzo TJ, Pakzad SN (2014) Modal identification of golden gate bridge using pseudo mobile sensing data with STRIDE. In: *Dynamics of civil structures*, vol. 4: proceedings of the 32nd IMAC, a conference and exposition on structural dynamics. Springer International Publishing, pp 293–298
- Meyers J, Fabian B, Brown N (2021) De novo molecular design and generative models. *Drug Discov Today* 26(11):2707–2715
- Nickels S, Stöckel D, Mueller SC, Lenhof HP, Hildebrandt A, Dehof AK (2013) Presenta BALL—A powerful package for presentations and lessons in structural biology. In: *2013 IEEE symposium on biological data visualization (BioVis) 2013 Oct 13*. IEEE, pp 33–40
- Pagadala NS, Syed K, Tuszynski J (2017) Software for molecular docking: a review. *Biophys Rev* 9:91–102
- Patel B, Singh V, Patel D (2019) *Structural bioinformatics*. In: *Essentials of bioinformatics*, vol I: Understanding bioinformatics: genes to proteins, pp 169–199
- Pires DE, de Melo-Minardi RC, dos Santos MA, da Silveira CH, Santoro MM, Meira W (2011) Cutoff scanning matrix (CSM): structural classification and function prediction by protein

- inter-residue distance patterns. *BMC Genomics* 12 Suppl 4(S4):S12. <https://doi.org/10.1186/1471-2164-12-S4-S12>. PMC 3287581. PMID 22369665
- Rego N, Koes D (2015) 3Dmol.js: molecular visualization with WebGL. *Bioinformatics* 31(8):1322–1324
- Rigden DJ (2009) From protein structure to function with bioinformatics. In: Rigden DJ (ed) Springer, Berlin
- Rosignoli S, Paiardini A (2022) Boosting the full potential of PyMOL with structural biology plugins. *Biomolecules* 12(12):1764
- Saoudi N, Latcu DG, Rinaldi JP, Ricard P (2011) Graphical analysis of pH-dependent properties of proteins predicted using PROPKA. *Bull Acad Natl Med* 192:1029–1041
- Sekihara K, Kawabata Y, Ushio S, Sumiya S, Kawabata S, Adachi Y, Nagarajan SS (2016) Dual signal subspace projection (DSSP): a novel algorithm for removing large interference in bio magnetic measurements. *J Neural Eng* 13(3):036007
- Shi M, Gao J, Zhang MQ (2017) Web3DMol: interactive protein structure visualization based on WebGL. *Nucleic Acids Res* 45(W1):W523–W527. <https://doi.org/10.1093/nar/gkx383>. PMID: 28482028; PMCID: PMC5570197
- Shkurti A, Goni R, Andrio P, Breitmoser E, Bethune I, Orozco M, Laughton CA (2016) pyPcazip: a PCA-based toolkit for compression and analysis of molecular simulation data. *SoftwareX* 1(5):44–50
- Stanfield RL, Wilson IA (1995) Protein-peptide interactions. *Curr Opin Struct Biol* 5(1):103–113. [https://doi.org/10.1016/0959-440X\(95\)80015-S](https://doi.org/10.1016/0959-440X(95)80015-S). PMID: 7773739
- Travers A, Muskhelishvili G (2015) DNA structure and function. *FEBS J* 282(12):2279–2295
- Vila JA, Arnautova YA, Martin OA, Scheraga HA (2009) Quantum-mechanics-derived ¹³C α chemical shift server (Che shift) for protein structure validation. *Proc Natl Acad Sci U S A* 106(40):16972–16977
- Vilar S, Cozza G, Moro S (2008) Medicinal chemistry and the molecular operating environment (MOE): application of QSAR and molecular docking to drug discovery. *Curr Top Med Chem* 8(18):1555–1572
- Webb B, Sali A (2014) Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics* 47(1):5.6.1–32. PMC: 4186674. PMID: 25199792 <https://doi.org/10.1002/0471250953.bi0506s47>
- Wei D, Xu Q, Zhao T, Dai H (2014) Advance in structural bioinformatics. Springer
- Wermuth CG, Villoutreix B, Grisoni S, Olivier A, Rocher JP (2015) Strategies in the search for new lead compounds or original working hypotheses. In: Wermuth CG, Aldous D, Raboisson P, Rognan D (eds) *The practice of medicinal chemistry*. Academic, pp 73–99. <https://doi.org/10.1016/B978-0-12-417205-0.00004-3>. isbn:978-0-12-417205-0
- Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, Verma V, Keedy DA, Hintze BJ, Chen VB, Jain S (2018) MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci* 27(1):293–315
- Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y (2015) The I-TASSER suite: protein structure and function prediction. *Nat Methods* 12(1):7–8
- Yao Z, Cao D (2023) 4.3.9 Procheck module. *PyBioMed Documentation* 10:53
- Yousif RH (2020) Exploring the molecular interactions between neoculin and the human sweet taste receptors through computational approaches. *Sains Malays* 49(3):517–525. <https://doi.org/10.17576/jsm-2020-4903-06>
- Yu J, Vavrusa M, Andreani J, Rey J, Tufféry P, Guerois R (2016) InterEvDock: a docking server to predict the structure of protein–protein interactions using evolutionary information. *Nucleic Acids Res* 44(W1):W542–W549
- Yuan S, Chan HS, Hu Z (2017) Using PyMOL as a platform for computational drug design. *Wiley Interdiscipl Rev Comput Mol Sci* 7(2):e1298