Simulation Modelling Practice and Theory 121 (2022) 102640

Contents lists available at ScienceDirect

Simulation Modelling Practice and Theory

journal homepage: www.elsevier.com/locate/simpat

Graph theory-based simulation tools for protein structure networks

Konstantinos F. Kantelis [a], Vassilios Asteriou [a], Aliki Papadimitriou-Tsantarliotou [b], Anthi Petrou [b], Lefteris Angelis [a], Petros Nicopolitidis [a], Georgios Papadimitriou [a,*], Ioannis S. Vizirianakis [b,c]

[a] Informatics Department, Aristotle University of Thessaloniki, Thessaloniki, Greece
[b] School of Pharmacy, Aristotle University of Thessaloniki, Thessaloniki, Greece
[c] Department of Life & Health Sciences, School of Sciences and Engineering, University of Nicosia, Nicosia, Cyprus

ARTICLE INFO

MSC:
68
92

Keywords:
Biological networks
Graph theory
Visualization
Simulation

ABSTRACT

Analysis of interactions in biological systems is at the core of almost every biological study. Being a principal tool, graph theory as a mathematical formalism, aims to model relationships and interactions between objects, assisting in the analysis of biological interaction networks at various scales. Over the course of the past few decades, many applications of graph theory to biology have been proposed and developed. In this work, a review of openly available online, graph theory-based tools with biological applications is presented. The selection of tools was based on a structured online search, with a focus on tools that could be utilized to analyze residue interaction and protein–protein interaction networks. After exploring the main features of each tool and with the aim of identifying their use cases in current research, selected papers recent to the extent as possible, referencing the tools, were reviewed with the aim of giving prominence to the effectiveness of graph theory. The latter is of crucial importance upon designing new drug molecules capable of recognizing and modulating the function of proteins (receptors) involved in disease pathophysiology. The knowledge extracted from these tools can be further exploited into cheminformatics approaches and computer-aided new drug discovery techniques. The selected work and specific examples of successful applications were analyzed, in order to highlight how each tool can be integrated in the workflow of different types of studies and to assess to what extent graph theory-based tools can be a valuable ally in the analysis of residue and protein interactions. Moreover, therapeutic development efforts provide further valuable insights and feedback in a plethora of scientific fields like biology, pharmacology and medicine and assist in the further development of graph-theory-based simulation tools for protein structure networks.

1. Introduction

Proteins are the most versatile macromolecules of the living systems and play a vital role in almost every aspect of biological functions in organisms, since they build-up the main structural and functional elements of the cell. The building blocks of proteins are amino acids that being linked via covalent-peptide bonds and forming polypeptide chains. Undoubtedly, the sequence of amino acids and their interactions in their three-dimensional configuration are crucial for the proteins' biological function. Thus, deciphering the protein structure (PS) (Fig. 1) contributes to determine the protein's dynamic, particularly their catalytic-active center, and further to understand how any acquired amino acid modifications (mutations) can influence their physiological function. [1]

* Corresponding author.
E-mail address: gp@csd.auth.gr (G. Papadimitriou).

https://doi.org/10.1016/j.simpat.2022.102640
Received 23 May 2022; Received in revised form 25 July 2022; Accepted 26 July 2022
Available online 3 August 2022
1569-190X/

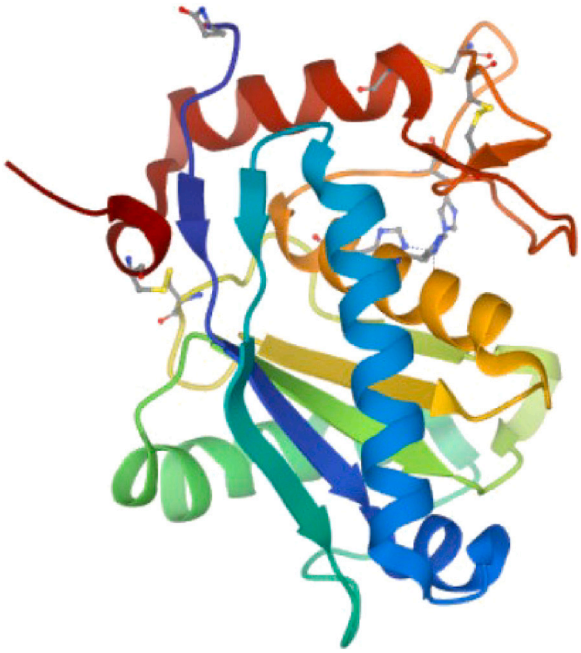


Fig. 1. Protein structure [2].

Based on molecular analysis data, it is well known that proteins in a great percentage, around 80%, do not act individually but interact with each other, as well as with other macromolecules (DNA, RNA) that mediate in metabolism and signal transduction pathways [3]. Therefore, the thorough understanding and precise comprehension of protein interactions is fundamental to clarify their roles within the cellular compartment. Besides that, the term "protein–protein interaction, (PPI)" (Fig. 2), includes a variety of biochemical events, such as transient and stable complexes as well as physical and functional interactions [4].

Since the cell functionality is governed by the protein molecules, mutations that affect their structure and consequently their function, are directly related to various pathophysiological phenotypes [6]. It is imperative to better understand protein structural molecular interactions given the fact that individual organisms respond differently in disease susceptibility and progression. Molecular heterogeneity in populations is fundamental in the new research era of precision medicine and pharmacogenomics, a sector that leads the modern medical and pharmaceutical research efforts, by implementing the translation of molecular knowledge in the clinical setting. Furthermore, the aforementioned empowers our capacity to predict the risk and the progression of illnesses as well as the drug therapy outcome. As a matter of fact, the precise elucidation of protein structure and protein interactions in a way to predict their function, contribute to further decipher the fundamental pathophysiological mechanisms and to enrich the pipeline of therapeutic interventions. Indeed, nowadays, the pharmacological design and development of new drugs that target the mutated gene or gene product (mutated protein) is reachable and applicable in the pharmaceutical environment [5]. It is thus expected that also the most effective treatment in complex and multifactorial diseases that show heterogeneity, with the typical example of cancer and cardiovascular disorders, will come to fruition [7,8]. Unquestionably, the molecular basis of diseases can be more precisely clarified and the therapeutic repertoire can be enormously enriched through building protein structures as well as molecular interaction networks which can be further used as a tool for disease prevention, diagnosis, and monitoring in addition to new drug treatment applications.

In the past decades, the only way to determine protein structure and relevant molecular interactions was through laboratory techniques such as mass spectrometry, NMR, affinity chromatography and X-ray crystallography [9,10]. Most of these methods are time consuming, expensive and require a lot of expertise. However, the development of computational biology, renders possible the prediction and mapping of PS and PPI networks that depend on their physical or functional relationship through the use of mathematical models [11]. More specifically, a PS network unravels the dynamic interactions between the residues of a protein molecule [12] while a PPI network reveals information about the way proteins interact with each other in order to enable any cell function. It is recently proposed however, that these interactions may be also efficiently predicted and physically elucidated based on deposited and available experimental data [10].

All computational approaches use data on proteins and their interactions found in various public databases [13],[14]. The main features used as inputs for these approaches are structure-based features as well as network topology-based features. Computational approaches for predicting protein interactions include the graphic illustration of such networks [15],[16]. An undirected, unburdened graph G, according to graph theory, is defined as a set of nodes V and a set of edges E, where the latter is the set of unordered pairs v1, v2, where v1, v2 $\in$ V when nodes v1 and v2 connected by an edge.

Fig. 2. Protein–protein interaction [5].

Graph theory has many applications in various fields including operational research, genetics, linguistics, and architecture [17]. In the field of protein science, graph theory has been used for the design of Protein Structure Networks [18]. These networks provide information about protein function [19] by investigating the effect of mutations at the amino acid level, in the three-dimensional configuration [20], and representing pathways of protein–protein interactions [21].

In most PS networks, amino-acid residues are identified as nodes and non-covalent interactions between them are considered as edges, based on the distance between C–C or side-chain atoms [22]. The extent of side-chain interactions can be further determined based on the number of atom–atom contacts, which considers the orientation of the interacting residues [23]. Side chain interactions play a key role in protein stability, folding, and thus protein function. Hydrophobic side chains on the surface play an important role in the interaction between proteins as well as proteins with DNA. However, identifying side chain interactions in experimental ways is very difficult, so researchers prefer mathematical methods.

Diverse problems such as the identification of clusters of interacting residues at the active-site, residues important for folding and stability, stabilization of protein–protein interfaces and the level of global connectivity have been addressed by the analyses of PSNs [24]. Further, Protein Structure Networks are also used in detecting problems such as allosteric communications within proteins [25].

In PS networks, protein residues are the “nodes” and their interactions are the “edges” while in PPI networks, proteins are the “nodes” and their interactions are the “edges”. Based on the existing bibliography, each network is formed as a system of differential and stochastic equations [26]. To date, many network simulation tools have been developed, focusing on various protein characteristics. However, many of these tools offer an analysis of a single structural property and are often available in software packages written in different programming languages while others are only available after purchase [9]. Therefore, users need to download and install a lot of tools and consult various documents. Moreover, protein interaction networks may sometimes be deficient [27] and prone to errors [28]. It is therefore important to enrich the tools that are available to the general public and can be easily utilized by each user. Particularly, toolboxes that bring accumulatively a set of appropriate tools in one website and thus reducing the time of analysis, are quite valuable. Contributing to offer a greater picture of the available tools, the present work aims to present some easy-to-use and directly accessible tools through which researchers can quickly and efficiently construct PS and PPI networks and can further analyze them for biological, pharmacological and even clinical exploitation.

As mentioned above, the aim of this review is to assemble some of the most useful and current state-of-the-art tools for the construction and analysis of protein structure and PPI interaction networks. For this purpose, a number of works were reviewed having in mind that the selected articles should have utilized a tool able of analyzing at least a PS or PPI network based mainly on a graph-related overview. The common characteristic was the use of graph based analysis. The selection was based on the most recent and commonly used tools. Precisely, the selected literature was published from 2018 onwards. Therefore, consequently pursues the analysis of each tool’s characteristics and potentials while also their practical application is presented. The tools selected

are PROSNEX, NAPS, CMVIEW, WebPSN, Cytoscape, RING 2.0, GraProStr, STRING, PTGL. Hopefully, this review will help readers, from a broad spectrum of beginners to experts, to understand how each tool works in order to select the one that will be more efficient for their work. To the best of our knowledge this is the first work to present a selection of the most recent works that utilize tools that take advantage of the network theory.

The remainder of this work is organized as follows. In Section 2, the different tools were presented along with the corresponding works focusing on the use of the tools in order to establish papers' results. In Section 3 the application of graph theory in drug design is analyzed along with a sum up table with the characteristics of each tool and one more with their field of application. Furthermore, a discussion for the common and unique characteristics of each tool is also provided in this section. Finally, the work is concluded in Section 4.

2. Network parameters for biological networks

Modularity analysis of Protein Contact Network (PCN) or Residue Interaction Network (RIN) provide essential information on the road to predict salient structural pathways in biomolecular systems like domains and structural repeats in proteins. Network centralities analysis can offer a valuable insight for a plethora of protein characteristics pertaining to the protein structures and interactions, including long-range connections and allosteric behavior. Mathematical parameters like centrality, betweenness and closeness, provide a quantitative base framework for network analysis. More specifically, closeness reflects the inverse distance of all shortest paths to a particular residue, betweenness represents the number of shortest bridges incorporating the specific residue, and measurement of centrality reflects highly connected nodes like hubs. As a result, a residue with a high closeness value indicates the network proximity to neighboring residues while a high betweenness value signifies a high degree of engagement in shortest path interactions.

Protein structure graphs (PSGs) define amino acids as the nodes and the non-covalent interactions among them as the links. Network topology of PSGs depends mainly on the cut-off of the interaction strength between the residues that are used in the constructed graph. While there are numerous ways to present the different bio entities in the form of a graph, in almost all these morphisms, graphs can capture various types of associations between the entities. In the following sections, the most common tools for graph-related representation of biological entities are presented, so as to reveal the strength of networks via graph theory and its application to biological systems.

2.1. Prosnex

2.1.1. Characteristics

Prosnex is a web service for construction and analysis of Protein Structure Networks (PSNs) alongside amino acid flexibility, sequence conservation and annotation features. In order to better understand the role played by each residue in shaping protein function, the tool offers the construction of several types of unweighted and weighted networks. Selecting between four different methods for edge assignment between nodes, the tool offers various configurations. Following network construction, average degree, path length, network density, clustering coefficient as well as node (residue)-centric metrics (node degrees, betweenness-centrality and closeness-centrality) are calculated and reported. In addition, the shortest path between two selected nodes and k-cliques are calculated as well. The following two features give Prosnex the appropriate novelty as a Protein Structure Network tool: (1) comparison of multiple single residue metrics from network analysis as well as other additional information such as sequence conservation scores and Uniprot annotations, and (2) usage of dynamic cross-correlations between pairs of amino acids from Network meta-analysis (NMA) in the weighted PSNs.

2.1.2. Referenced work

In [29] the authors studied the Leucine rich repeats (LRRs) with 20–30 residues that form a super helix arrangement and tried to identify the reason why among 11 classes recognized, only 3 classes adopt an α-helix. Using the Prosnex and the average node degree at each position in the Variable Segment (VS) consensus, the study revealed that all conserved hydrophobic residues show higher node degree than the overall average one. Residue at position 8 shows a very high node degree, revealing that this is a hub node. Consequently, this position in the VS part is occupied by conserved aliphatic residue, adopting an α-helix.

2.2. NAPS

2.2.1. Characteristics

NAPS is an online tool aiming at the analysis and interactive visualization of protein contact networks (Fig. 3). The tool constructs five different network representations from a PDB-formatted file. These networks can have different types of nodes, interaction strengths and edge weights, resulting in various types of representations, offering insights pertaining to protein analysis. Concerning the network visualizations, the tool is capable of providing four different visualization options including 3D structure view, 3D network view, 2D contact maps and distance matrix view. Using these options, one can study molecular interactions of active site residues and ligand binding residues.

The centrality analysis that the tool offers is based on the following measurements owing to identifying the most central or significant nodes in the network. Degree, closeness, betweenness, clustering coefficient, eigenvector centrality, eccentricity, average nearest neighbor degree, strength and edge betweenness are the centrality measures that the tool can calculate. Graph spectral analysis is a valuable tool for protein structure analysis regarding cluster identification, structural domains and repeats analysis. Along the same lines, shortest path and k-clique analysis helps identify allosteric phenomenon and protein folding.

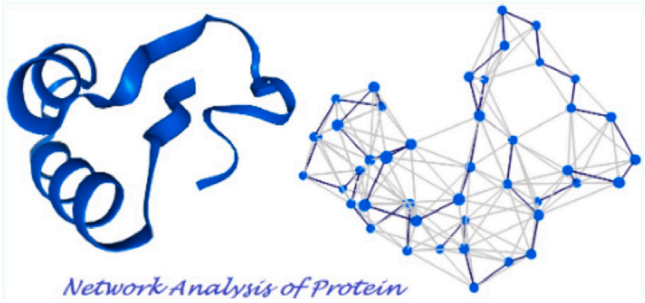


Fig. 3. NAPS.

2.2.2. Referenced work

Studies of protein structure analysis, investigating various aspects of network topology, have revealed the role of topological analysis in identifying residues crucial for protein stability and protein dynamics. These residues also govern enzymatic activity, allosteric regulation, signal transduction and protein folding kinetics.

In [30] the authors performed comparative in silico studies to analyze the structure of the spike glycoprotein of SARS-CoV-2 and its underlying mechanism to bind the host cell receptor. Utilizing a plethora of tools, the authors studied the structural and functional details of the interactions between the spike glycoprotein of SARS-CoV-2 and its cognate cellular receptor ACE2 in order to find the hot-points for designing potential drugs. These drugs will inhibit the interaction between the spike glycoprotein and ACE2 and block the virus–cell interaction.

The interaction pattern between the spike glycoprotein and ACE2 in the aforementioned work has been evaluated for both SARS-CoV and SARS-CoV-2 using NAPS. The reason for this comparison was to find the source of such a dramatically high infection rate of SARS-CoV-2 in comparison to SARS-CoV (more than 3 times higher). The results showed that the spike is attached to the cognate receptor through two regions at the beginning and end of the receptor-binding motif. Using NAPS, the authors were able after the optical inspection, to identify critical amino acids that participated in the spike-ACE2 interaction. These interactions were derived from the Molecular Dynamics (MD) simulation, performed for the analysis of noncovalent interactions between the spike and ACE2 receptor.

The authors in [31] showed that conformational stability of pro-apoptotic BAX is dictated by discrete residues of the protein core and identified a nexus of interactions, involving four residues of the BAX core α5 helix that are individually essential to maintaining the structure and latency of monomeric BAX and are collectively required for dimeric assembly. To investigate how the allosteric nature of BAX conformational regulation could be mediated by the α5 core, an unbiased computational approach was used based on an analysis of BAX as a network of interactions between each of its amino acid residues. Centrality, closeness and betweenness were the three parameters that were investigated and resulted in the identification of the amino acids 113–116 of the BAX α5 core as a nexus of circumferential interaction.

More specifically, using NAPS, the authors identified amino acids with a high closeness value indicating the network proximity to neighboring residues and those with a high betweenness value, signifying a high degree of engagement in shortest path interactions. Specifically, the F114 amino acid of α5 demonstrated the highest closeness value among all BAX residues and the 113–116 quartet of residues was the cluster with the highest scoring of consecutive residues for both closeness and betweenness. These results implicated that core a5 residues 113–116 act as a potential control point for the conformational regulation of BAX.

In another work [32], the authors identified allosteric modulators (SANC190 and SANC651) against P. falciparum Hsp70-1 and Hsp70-x, affecting the conformational dynamics of the proteins, delicately balanced by the endogenous ligands with the aim of combating malaria. Coarse-grained dynamic residue network analysis (CG-DRN) was employed on ligand-free and ligand-bound PfHsp70-x and PfHsp70-1 trajectories as a verification step for all-atom simulations. The edges in the residue interaction networks were weighted based on the defined interaction strength and dynamic residue correlations couplings.

In the above work, the option of NAPS to analyze trajectories was exploited. By using 1000 frames of CABS-CG trajectories, the aforementioned tool computed the average single network using Cα representation of trajectories. So as to achieve this, Cα of amino acid residue is considered as a node and an edge is drawn if two nodes share an edge for 60% of the simulation. Using the constructed protein structure network, the residue-based betweenness centrality parameters were computed with the NAPS server. High BC values correspond to highly connected nodes/residues, thus, to critical connection hubs that could regulate allosteric transitions in proteins.

In [33] the authors performed structural analyses of Rab11 and identified eight representative structures, two of which are novel binding sites that may allosterically regulate Rab11. In order to analyze the signaling between the newly identified binding sites and the active site in Rab11, the residue interaction networks of the representative structures of Rab11 were generated and examined with NAPS. The authors investigated the shortest paths from the point of contacts (i.e. contacting residues) made by the top scoring ligands at site 1 and site 2, to the residues that are in contact with the active site. The results showed that the paths include fewer residues and their lengths are minimal as compared to that of the shortest paths from other residues of site 1 and site 2 to the active site.

Highly correlated central residues that lie on shortest paths can play key roles in allosteric formations. What is more, residues in the shortest paths have been experimentally validated to be important in allosteric interactions. Betweenness centrality values of the residues involved in the shortest paths were also under examination. Derived from the study of residues and the related one with NAPS, residues involved in shortest paths geometrical metrics were found to have high centrality values, while the residues connected to them, in the shortest paths, were also found to have high betweenness centrality.

In another work, [34], the authors propose a new computational method based on an eXtreme Gradient Boosting (XGBoost) algorithm which can effectively predict hot spot residues in protein–RNA interfaces utilizing an optimal set of properties. The study is based on calculations of 156 sequences, structures, exposures and network features for a set of 47 protein–RNA complexes. The network features that were calculated were degree, closeness, eigenvector centrality, betweenness, clustering coefficient, average nearest neighbor degree and eccentricity. Analysis of the results showed that residue interaction network is crucial for protein behavior predictions.

The authors used the NAPS tool to calculate the aforementioned topological features. According to the proposed six optimal features scheme, two network features were selected (closeness and eccentricity) having the highest F-score. The results validate that these network-related indexes are vital for discriminating hot spots and non-hot spots residues at protein–RNA complexes.

In a work for designing rational drugs [35], the authors have investigated the phosphorylation-induced structural dynamics of the With-No-Lysine (WNK) kinase bound to an inhibitor via atomistic MD simulations. Structural network analysis via NAPS have revealed that the phosphorylation causes structural rearrangements and shortens the signaling path between the αC-helix and catalytic loop, making the binding pocket more suitable for accommodating the ligand. The last trajectory of the simulations was used for the representative structure in NAPS.

Network analysis of protein structures had revealed how the aC-helix and catalytic loop connect in uWNK and pWNK and how the connections affect the different interactions. Two network parameters were analyzed further, the shortest distance path and the betweenness centrality. The former was analyzed by considering the salt-bridge interaction between Glu268 (αC-helix) and Arg348 (at catalytic loop) and used as a terminal residue for calculation of the shortest path in NAPS. The latter was utilized so as to validate that the ligand interacts with the residues of the pre-organized pocket and the phosphorylation causes the stabilization of the binding pocket through the induced fit mechanism.

Contributing to the works that utilized the features of NAPS, in [36] the authors make use of machine learning approaches, including tree based methods, random forest and gradient boosted tree (GBT) classifiers along with deep convolutional neural networks (CNN) for prediction of cancer driver mutations in the genomic datasets. These machine learning predictions were leveraged in molecular simulations along with protein stability and network-based analysis of cancer mutations in the protein kinase genes in order to obtain insights about molecular signatures of driver mutations and enhance the interpretability of cancer-specific classification models.

For the protein structure network analysis realized in this work, residues were treated as network nodes and inter-residue edges represented residue interactions. Following the authors hypothesis that cancer mutations may preferentially target the essential mediating residues with high centrality — playing an important role in activity and signaling of protein kinase genes, betweenness of residues was under investigation. After appropriate normalization, the aforementioned metric was used as a mean to identify whether positions of deleterious mutations would overlap with the global mediating nodes in the interaction networks.

2.3. CMView

2.3.1. Characteristics

Contact maps are a valuable visualization tool aiming to identify structural features such as domain architecture, secondary structures and contact clusters (a contact map is a two dimensional representation of residue–residue contacts in a 3D protein structure). Targeting to this research, CMView is a tool that analyzes and displays protein contact maps (Fig. 4). The tool can handle different types of file formats (PDB, CASP TS, CASP RR, native CSV) as long as the contact threshold and the type are specified. The distance map, contact density and triangle inequality relations (“the neighborhoods”) are available as colored overlays in the contact map window of the application. The latter offers a complete set of different groups of contacts and various export formats to select.

A key feature of the CMView is the pairwise structural comparison of two proteins or conformations. The tool can compare two different structures and identify the common and the unique characteristics of each conformation. As a means to achieve this, the program implements three different alignment methods and offers refinement options by electing contacts and recalculating the superposition based on this selection. The aforementioned characteristics contribute to a more intuitive approach to explore the relationships between contact patterns and protein structure.

2.3.2. Referenced work

In [37] the authors evaluated the conformational changes of the Diphtheria Toxin (DT) structure at three different temperature levels using molecular dynamic simulations. Under this study, the tertiary structure of the DT was compared at different temperatures, using the contact map feature of the CMView tool. The purpose of this work was to determine a suitable temperature level for the production process of the diphtheria vaccine due to the need for protection of conformational B-cell epitopes.

More specifically, the contact maps were used to investigate the effect of temperature on the tertiary structure of the DT for three different temperatures. Chain A was utilized and the inspection of the produced contact map revealed the structures that

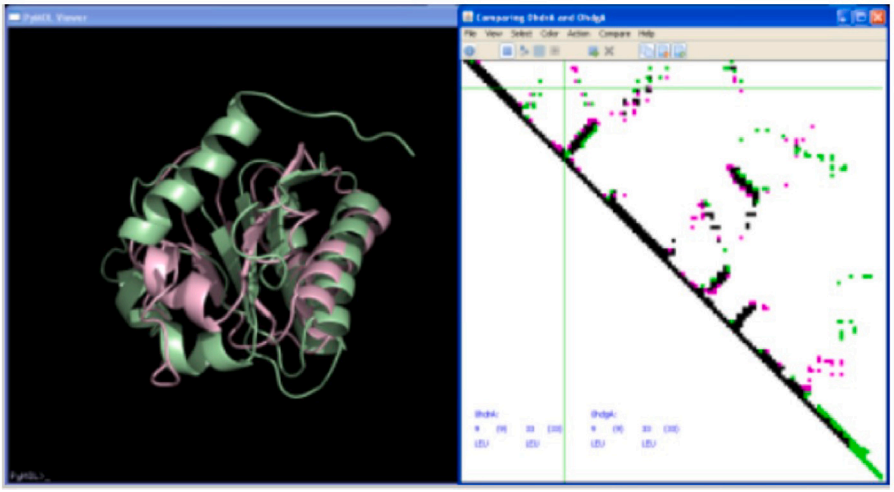


Fig. 4. CMVIEW.

were conserved at specific temperatures. Further experiments and CMView results aimed to identify the critical temperature with reference to the deformation of some valuable formations.

In another work where the authors exploited the features of the CMView, [38], they studied the process of protein denaturation in supercritical CO_2 using molecular dynamics simulations. They proposed a new molecular mechanism which states that the inappropriate interactions of polar side chain of lysine residues with the non-polar nature of supercritical CO_2 enforce these exposed residues to return from the solvent bulk to the protein exterior. Additionally, these residues form new non-native hydrogen bonds with the acidic residues such as glutamic acid.

The CMView tool in this case was used for contact map analysis. The CI2 was under investigation as a member of the potato inhibitor I family of serine protease inhibitors. As it has one domain and contains a-helix, b-sheet and loop structural elements, it is a model protein for structural biology research. Visual inspection of the protein contact maps of the native state in the crystal form and after simulation in $scCO_2$ under specific conditions revealed that lysine residues in these unstable regions of the protein have an important roles in destabilization of CI2 in $scCO_2$. In [39] the authors studied 5 different single nucleotide polymorphisms (SNPs) in the protein kinase R for homology modeling and the generated models were investigated with regard to their secondary structure, residue fluctuations and eIF-2α binding. The work revealed structural changes in all mutants, yielding a more stable structure at the cost of reduced flexibility and less conformational freedom compared to the native protein.

In this work, seven different computational tools were used in order to uncover potential effects of SNPs on the protein's structure, ligand binding and function of the human protein kinase R. CMView was utilized to perform a contact map analysis and to visualize changes between the interacting residues. By comparing the contact maps of native and mutant structures, conformational differences were revealed. Exploiting the capabilities of the tool, the authors were able to visualize the structural differences of the five SNPs and spot the main areas of differentiation.

In the work of [40] the binding of nickel by immune proteins was investigated. This research was focused on Trastuzumab and Pertuzumab IgE variants as serendipitous investigation models shedding light on Ni-NTA interactions independent of Her2 binding being due to glutamine stretches. Using distance based maps between Fab domains of Pertuzumab and Trastuzumab the authors studied the varying changes in the heavy and light chain pairing. Optical investigation of the maps helps identify the areas where the most noticeable changes take place and particularly the contribution of the V- and C- regions of the IgE variants in the structural formation of Ni-binding Q-stretches (continuous clusters of at least 3 glutamines).

In another work, [41] the authors studied the CAZy glycoside hydrolase family GH13 catalyzing the hydrolysis of polysaccharides such as glycogen and starch. Based on this transglycosylation process and the identification of structural determinants associated with GH13 family reaction specificity, a new computational approach for decoding the determinant structural composition defining the aforementioned reaction specificity was proposed. Using MD simulation, this approach is capable of identifying residues outside of the active center, affecting the reaction specificity.

Contact maps and alignments were produced with CMView using a cut-off distance of 5 Å between all atoms of the 14 members of the CAZy family GH13 and were aligned against the contact map of the Thermotoga maritima 1,4-α-glucanotransferase TmGTase (PDB: 1LWJ). An R program has been developed and used for analysis of the frequency of all amino acid pairs for all contacts. The contact map of TmGTase was transformed then into a 3D network, whose nodes were the α-carbon of each residue and the edges of the residue contacts. Further processing of this network was based on the betweenness centrality (BC) of a node using a specific function.

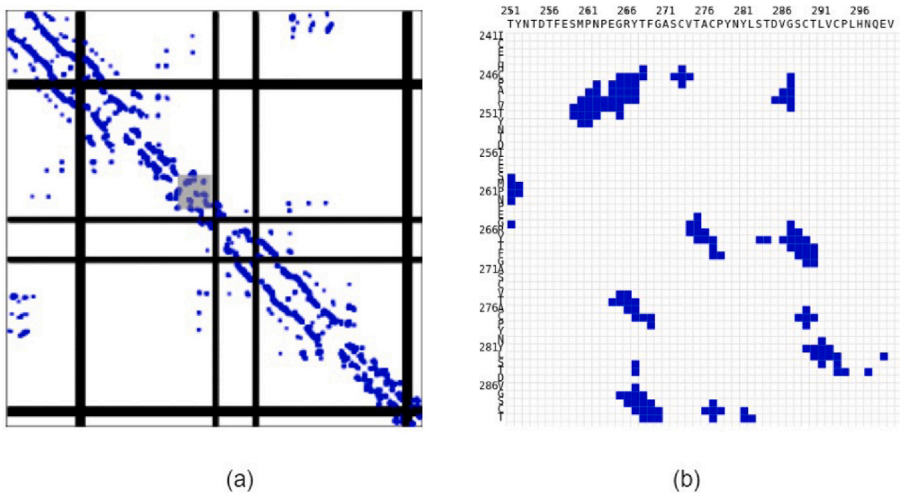


Fig. 5. Contact map of the human HER2 protein (PDB code 1N8Z). (a) CMweb generated contact map, (b) detail of (a).

2.4. CMweb

2.4.1. Characteristics

CMweb is an online protein contact map calculation tool. In CMweb, five contact prediction methods are implemented: mutual information based, statistical coupling analysis, explicit likelihood of subset covariation, observed minus expected squared and one of the Goebels methods. Using PDB identifiers as queries, users can generate and inspect the contact map with visualizations similar to that of Fig. 5. They can also map and visualize the contacts onto the 3d structure, retrieve multiple sequence alignment results as well as various statistics on the contact map, and compare to other RR contact prediction methods by uploading comparison maps. Contact Maps can be shown in either binary or in continuous formats, the former indicating just the presence/absence of contacts and the latter being a distance map.

2.4.2. Referenced work

The purpose of the study in [42] was to reconstruct a high resolution three dimensional structure of PI3K-SH3 amyloid fibrils using cryogenic electron microscopy. To realize this, the authors produced purified wild type and mutant PI3K-SH3 and then processed it in highly acidic conditions of pH 2.5 in order for fibrils to form. The fibrils were scanned with Atomic Force Microscopy to obtain images, the elongation rate was measured using Quartz Crystal Microbalance and finally the cryo-EM image was acquired and processed in 3D refinement software which yielded the 3.4 angstrom resolution fibril structure. With this data the final model was built.

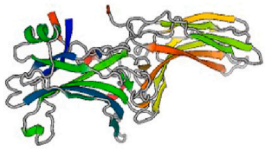
The authors discussed the structure of the fibrils in detail and pointed out the differences between the native conformation of PI3K-SH3 versus the fibril. CMweb was used to compare the contact maps of the native conformation of PI3K-SH3 and the amyloid fibril. The comparison showed no common contacts highlighting the conformational difference. Comparison of the wild type fibrils to the mutant ones, which had one point mutations of some isoleucine residues with alanine, showed variation of the elongation rate.

Another work that used CMweb is [43]. The aim of this research paper was the study of the Plasmodium Falciparum Chloroquine Resistance Transporter, a vacuolar transporter whose function can act antagonistically in antimalarial drug-based therapies. Initially, an evolutionary study and a homology modeling study were carried out. Through homology modeling, conformations of the PfCRT were predicted based on models of other homologous transporter molecules in various conformations. The PfCRT conformations were then compared with the evolutionary analysis. The results suggested that the proposed functional sites of the PfCRT (transmembrane helices, cavity site) have some strongly conserved amino acids. Furthermore, the binding site and the cavity were suggested as the targets of most drug resistance related mutations. Contact map calculations on CMweb were used in the context of this study, in concert with other computational tools, in order to investigate and validate models produced from the homology modeling step.

2.5. WebPSN

2.5.1. Characteristics

The protein structure network analysis tool WebPSN, computes network features (e.g. nodes, hubs, links, communities etc.) and shortest connection pathways on a single high resolution structure while information on system dynamics (i.e. cross-correlation of atomic motions) is supplied from Network Model-Normal Mode Analysis (ENMNMA)-based strategy (Fig. 6). The latest version of the program has the ability to compute the differences in nodes, links and signaling pathways between two structures (i.e. PSN



Links | Hubs | Communities | Global MetaPath | Filtered MetaPath

Color Legend

Show 10 entries Search:

Index	Node1	Node2	Force	Hub1?	Hub2?	Community
1	I.2 (A.11)	I.4 (A.13)	3.80333	No	No	0
2	I.4 (A.13)	II_III.3 (A.29)	18.8833	No	No	0
3	I.5 (A.14)	II.4 (A.25)	8.36	No	Yes	0
4	I.6 (A.15)	X.2	5.05667	No	No	0

Fig. 6. WEBPSN.

difference) and (b) infer links, hubs, communities and metapaths from consensus networks computed on a number of structures, a similar characteristic to the CMView tool. A number of network features including hubs or node communities and shortest paths are also available. The server provides a global metapath, made up by the most recurrent links in the pool of filtered paths and gives the user the option to filter those paths related to the desired residues.

The tool offers also an interactive table listed all pathways with a number of associated indices including path length, mean square distance fluctuations between all node pairs in a path (MSDFp), an index of path stiffness, hub content, fraction of user-defined relevant amino acid nodes, correlation score, average similarity score between the path and each path in the cluster and the number of the cluster the path belongs to. Additionally, distribution of: path length, MSDFp index, hub content, correlation score, node frequency in paths and frequency of four amino acid fragments in the predicted paths are also covered from the features of the tool. PSN analysis can also be carried out on two structures to infer differences/commonalities between links, hubs, or metapaths (PSN difference), or on a set of homologous/analogous proteins/nucleic acids to infer consensus links, hubs, communities, or metapaths (consensus PSN). 3D maps, data tables and plots are available to complete the usefulness of the tool.

2.5.2. Referenced work

In [44] the authors studied the implications of protein sequence divergence on the structure, dynamics and function of homologous yeast and human SF3b spliceosomal subcomplexes. Among the findings, the one that stands out is that SF3b6 potentially acts as an allosteric regulator of SF3b1 for BPS selection and might play a role in alternative splicing and overall suggesting that the human SF3b may have evolved sophisticated mechanisms to fine tune its molecular function. Results highlight inherently variable features within the homologous SF3b structures. According to this research, such variations are primarily contributed by differences in the relative orientation of human SF3b4 vis-a-vis its yeast homologue Hsh49 as well as in the interaction patterns of SF3b3–SF3b5 and SF3b3–SF3b1 interfaces.

In order to complete the study for long-range residue–residue interfaces that can facilitate allostery, the authors examine whether SF3b6 binding causes an allosteric effect for SF3b1 considered in isolation and in complex with SF3b6 using webPSN. The tool integrates graph-based protein structure networks and elastic network model-based NMA, to combine inherent structural connections and their dynamic properties. Based on cross-correlation between two residue motions and an overall network structure analysis, webPSN identified a metapath validating the allosteric hypothesis (Metapath provides information about the most recurrent long-range connections present within a protein).

In [45] the authors studied through molecular dynamics, the dynamics of diffuse large B-cell lymphoma (DLBCL)-linked N11Y mutation to facilitate resistance against Bcl- 2-competitive inhibitors and “BH3 mimetics”. Utilizing appropriate tools for in depth analysis of residue interaction network, dynamic cross-correlation and free energy landscape, the authors revealed that the mutation modifies the conformations of key residues, thereby altering the shape of the hydrophobic groove. Utilizing trajectories from MD simulations, the authors studied the structural pathway by computing the shortest path and the one of the most connected pathways between the pairs of residues (non-linked residues) between the point of mutation (N11 or Y11) and the hydrophobic groove-forming residues. Exploiting webPSN server processing results, they identified a specific residue mutation to play possibly a key role to the signalings established within the protein.

In [46] the authors have attempted to study a structural comparison of SARS-CoV-1 and SARS-CoV-2 M pro in apo and inhibitor bound states using protein structure network (PSN) based approach at contacts level. Furthermore, they have shown how inhibitor binding perturbs the PSG and the structural pathways in M pros and identified critical residue pairs. (NAPS server was also used in order to examine and analyze three centrality measurement parameters for the aforementioned critical residue pairs.)

Hubs, community and structural formations analysis were analyzed using a mixed PSN ENM-NMA approach implemented in WebPSN. As MPro structures do not show much difference, an analysis of the contact points either generated or lost due to change of few amino acids, was conducted with the webPSN on the road to provide an insight into restructuring of modules within the Mpro. As network parameters for both proteins did not have any significant difference, using the tool, community structure analysis of PSN for various residues was conducted and revealed several rearrangements in hub residues and signaling pathway. The latter probe subtle conformational changes associated through-out the structure of the two proteases under investigation.

In another work [47], the role of conformational dynamics in the structural interconnection and long-range allosteric regulation of germline PTEN variants associated with autism spectrum disorder (ASD) or cancer was demonstrated. Using network proximity analysis performed on the human interactome, molecular simulations, and residue-interaction networks, this work showed that the PTEN interactome shares significant overlap with the ASD and cancer interactomes.

Utilizing the protein-structure network (PSN) analysis, the authors sought to deduce the effects that mutations have on the native PTEN fingerprint and to identify key hub residues that govern allosteric regulation within the long-range signaling pathway. Results from the global metapath analysis indicated that cancer-associated variants are distinct in that they possess a global metapath signaling pathway that propagates across the entire protein structure. This global path extracted from webPSN comprises critical inter-domain hub nodes that govern allosteric formation. Specific pathways of structural connections, where the cancer-associated global metapath distinctly spreads across the inter-domain interface were revealed from the analysis performed using the webPSN, whereas the ASD-associated global metapath is predominantly restricted to the phosphatase domain.

In [48] the authors tried to understand the mechanism of possible inhibition by the phytochemicals to combat SARS-CoV-2 with their capability of targeting multiple proteins. Via MD and network centrality analysis the efficiency of phytochemicals from Withania somnifera, to bind to a total of six SARS-CoV-2 targets, have been shown. Network communities increase while hubs and links decrease as a result of the phytochemical binding.

WebPSN v2.0 was used to build protein structure networks for the six SARS-CoV-2 free protein targets and in complex with the phytochemicals. Results from metapath analysis revealed the residues important in allosteric formations with the phytochemicals. Moving a step forward, the global metapath differences revealed that the metapath of free and phytochemical-bound form differed considerably in their network connections and resulted in global network rewiring. The authors utilizing the Webpsn to analyze the aforementioned network change, found that modifications in residue interaction and rewiring of residue network related to functionally important residues, can cause SARS-CoV-2 target inhibition.

In [49] the authors present a comparative analysis of the potential theoretical inhibitory effect of both avermectins (avermectin-B1a [HB1a] + avermectin-B1b [HB1b]) on biomolecules associated with COVID-19. Moreover, the results suggest that ivermectin through its homologs has a multiobjective behavior. This behavior was mainly influenced by the hydrophobicity of each homolog and of the binding pockets.

Using the webPSN tool, the contact stability between HB1a/HB1b and their target receptors were investigated. The force constant relative to the initial frame of each homolog in complex with IMPα1, IMPβ1, Helicase and Mpro was studied, using two alternative versions of ENM to calculate the cross-correlation of the motion of the Cα atoms and for the pairwise interactions between the atoms of Cα. Results from the webPSN analysis suggest that compound HB1a binds more strongly to the two importins studied, while HB1b binds more strongly to viral proteins (Helicase and Mpro).

2.6. RING 2.0

2.6.1. Characteristics

RING-2.0 is another tool featuring the calculation of both intra and inter chain interactions as well as contacts involving 'hetero atoms' (i.e. ligands, DNA/RNA, cofactors, metal ions and solvent molecules) (Fig. 7). The program generates the residue interaction network following a two step process: firstly it identifies specific interaction types and secondly it filters those interactions based on the user provided specifications. A PDB identifier or any structural PDB file can be loaded and the exported files include the network in GraphML (XML) and text format. Moreover, the processed PDB structure with hydrogen atoms and the vector image (SVG) of the graph for further analysis using other specific tools like PyMol and Cytoscape are also some additional features of the tool.

RING offers edge rendering of interactions by type, distance, orientation and node identifier filtering. Likewise, it offers a way to customize the node and edge view transparently as for normal atom selections. The latest version of the tool is able to find both intra- and inter-chain interactions and contacts involving hetero atoms. Structural features are reported for each node, which include the secondary structure, vertex degree, experimental uncertainty for X-ray structures, conformational energy preferences determined with FRST and TAP, conservation and cumulative mutual information, calculated from PSI-BLAST profiles. The RING algorithm calculates atomic interactions based on geometrical criteria and takes into consideration different interaction types including Van der Waals interactions, Hydrogen bonds, salt bridges, π–π stacking interactions and π–cation interactions. (Since March 2022 the version 3.0 of RING is released, having optimization of the algorithm for the calculation of interactions.)

2.6.2. Referenced work

In [50], an analysis of the SARS-CoV-2 spike protein based on coarse grained molecular simulation and structural methods is carried out, to characterize residues targeted by antibody escaping mutations that maintain high binding affinity to the host receptor ACE2, with respect to their energetic and allosteric dynamics. After the preprocessing of relevant PDB files, coarse grained molecular simulation was used to analyze the protein's conformational dynamics. Using PCA on the simulation trajectories and Elastic Network Modeling, it was determined that mutations may favor positions that play important roles in functional coordination and allosteric changes, as such positions can inhibit antigen binding while maintaining host receptor binding affinity. Mutational sensitivity analysis and alanine scanning led to the hypothesis that mutations may target residues that are allosterically adaptive and structurally plastic and thus modulate global motions and allosteric formation. Perturbation response scanning, which quantifies allosteric response of residues to residue perturbations, corroborates the view that widespread mutations target allosteric hotspots that regulate recognition of antibodies and link with ACE2.

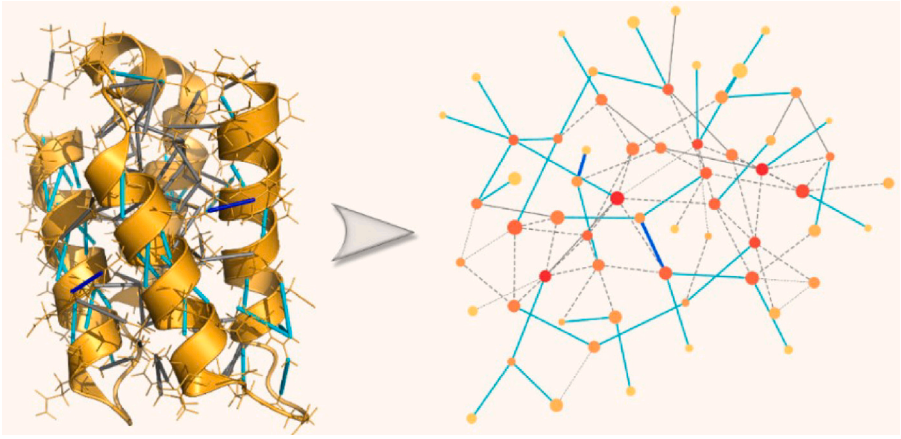


Fig. 7. RING.

Protein structure networks and community analysis were also employed in this study. Specifically, using RING 2.0, interaction network models of the spike protein conformations were built and analyzed. On this geometry-based residue interaction network the edges were weighted with dynamic correlation and coevolutionary residue coupling data from the above processing in order to obtain a representation which is sensitive and accurate in its correlation of long range inter residue dependencies. The network then was subjected to community analysis with a modified version of the Girvan Newman algorithm and the Leaders–Followers algorithm, as well as analysis with a bridgeness metric similar to the Rao Stirling index. The key community anchor residues pointed to by the analysis are also regulatory nodes, that contribute to the stability of the complex and control allosteric signaling between the spike protein and the antibodies.

The focus of [51] was the study the hERG1 potassium ion selective channel at the atomic level in order to obtain insights into the binding modes of hERG blocker compounds by using structural methods and the latest cryo-EM hERG model instead of homologs. Since the hERG channel has high affinity with several compounds, assessing hERG liability during the preclinical stage of drug development is necessary to avoid unintended cardiological side-effects. In this paper, a structural methods and MD simulation based workflow was developed and applied to characterize the binding of specific drugs to hERG. After performing a docking step and obtaining a few poses for each drug, the complexes were embedded in a bilipid membrane followed by solvation in water. The binding free energy was calculated and adaptive biasing force simulation was used to find the power mean force of unbinding, which revealed new minima in the energy landscape.

Three main sites in the hERG channel were identified as highly important to compound binding, close to the central cavity and the selectivity filter, with the main modes of binding being the cation-π, $\pi-\pi$ and hydrogen bonds. Mutation of one of these sites in particular, led to the experimental observation of decreased binding affinity of the mutant hERG compared to wild type hERG. Since the $\pi-\pi$ interactions are known to play a role in both stabilizing the conformation of hERG and in drug binding to aromatic residues, the $\pi-\pi$ interaction network of hERG was computed in RING2.0. It was shown that cisapride, one of the drugs considered in this study, significantly changes the $\pi-\pi$ interaction network close to the critical binding sites.

In [52] the goal was to investigate whether approaches based on fast computational methods and web servers can help in the analysis and impact assessment of missense substitutions in proteins. The study focused on specific proteins and variants for which detailed data can be found in the literature and applied a collection of methods, both sequence and 3D structure based, some not specifically designed for missense mutations analysis, in order to determine how effective such a methodology can be in this type of study. Data were obtained from PDB and PharmVar and preprocessed with Chimera. Amongst the analyses used are sequence based mutation labeling, 3d stability prediction, 3d structural analysis of variants including disease propensity prediction, protein flexibility prediction, residue interaction network based analysis with RING 2.0 and Cytoscape, binding pocket and PPI site prediction, antibody and heparin docking and protein–membrane interaction.

Regarding the role of residue interaction networks, the authors pointed out that the connection affinity of a residue to the rest of the network, as well as whether it participates in a flexible or rigid location of the protein, play an important role in the substitution impact assessment. The authors concluded that using 3d-structure information in this type of study can greatly increase understanding of the effects of residue alteration compared to only using sequence based methods and besides that, it can help narrow the focus of assays. Another finding derived from this work was the fact that paying special attention towards ionic interaction-related substitutions may require special attention in case the data is not provided.

In [53] the objective was to determine the 3d structure of alternansucrase and to specify which structural characteristics of the enzyme give rise to its linkage specificity and alternance properties. The core of the research was the in vitro production of ASR as well as 17 single mutants and two double mutants and the subsequent characterization of the enzymatic reaction products, as well as the crystallization and x-ray diffraction analysis of the enzyme. Manual docking and structural inspection were also used to help with the structural characterization. The analysis pointed to specific sites in the enzyme responsible for alternance. Hydrophobic

core residues and $\pi - \pi$ stacking interactions determined with RING2.0 were identified as stabilizers of the enzyme packing. The calcium binding site, as well as the structure of domain V were found to also be important contributors to structural stability.

This study [47] comprises a molecular dynamics and residue interaction network based analysis of germline PTEN missense variants related to autistic spectrum disorder and cancer, with a focus on characterizing specific molecular features associated with either or both disease phenotypes. A prospective cohort study was the basis for the identification of PTEN missense variants, from which 138 samples were collected, processed and screened arriving at 17 variants. Mutant PTEN models were generated from a wild type PDB model with in silico side chain replacement and subjected to MD simulation with a series of preprocessing steps, preceding the production simulation, whose trajectories were then clustered and used for analysis of conformational changes.

The article also focuses heavily on network based methods. A human interactome network model was obtained by filtering experimental data from several bioinformatics databases. Network proximity of ASD and cancer related genes to PTEN related genes and proteins was computed and found to be statistically significantly lower than that of random nodes in the interactome network. RING2.0 was used to compute residue interaction networks on which statistics such as degree, closeness and betweenness centrality were calculated per residue and used in the analysis. Residues far from the solvent were more well connected in the network and thus considered as playing an important role in long-range signaling. Also, hub residues determined on the basis of centrality measures were found to be preferentially enriched in different domains for each mutation class. Finally, webPSN was used to perform a combined Protein Structure Network and Normal Mode Analysis in search of global metapaths that modulate allosteric behavior. This analysis highlighted the role of the inter domain interface in stabilizing the complex and generally in its dynamics.

2.7. Cytoscape

2.7.1. Characteristics

Cytoscape is a desktop application program focusing on visualization and analysis of network-resembling structures across various bioinformatics applications (Fig. 8). Its main interface allows importing, defining, displaying, browsing and annotating graph data, which are presented to the user both graphically, where various layout tools enable creating comprehensible depictions of the data, and in tabular format, where the user can view and edit the data and create new attributes/columns. Other features include editing the style and annotation layers and also filtering data based on attributes or even graph-based criteria such as degree and topology.

The main strength of Cytoscape, however, is that the above core functionality is also exposed programmatically. In combination with open source licensing, this enables the development of apps (formerly plugins), which are extensions to the main Cytoscape application and a number of which are available for installation online. In this way, apart from a desktop application, Cytoscape is also a platform for the development of any graph related data processing in bioinformatics.

2.7.2. Referenced work

The work of [54] tested Brain MRI markers of brain structure for association with disease markers in other organs. Having as a starting point that changes in brain structure and cognitive decline occur in Chronic Obstructive Pulmonary Disease (COPD) and that they also occur with smoking and coronary artery disease (CAD), this work tried to group markers within organ systems into composite markers by utilizing principal component analysis (PCA). Univariate relationships between brain structure and the disease markers were explored using hierarchical regression and then entered into multivariable regression models.

Using the functionality of cytoscape to layout relationships and having performed Pearson's correlations to investigate univariate relationships between brain structure and markers from other organs, a network diagram was generated representing these relationships. In addition, subnetworks generated showed the specific markers that were correlated with Brain Tissue Volume, White Matter Hyperintense lesion WMH Volume and White Matter Microstructure (retinal markers did not correlate with any brain measures). It should be noted that the aforementioned network representation was based on a rule where all lines indicate significance at $p < 0.05$ while the line thickness indicates strength of correlation (correlation coefficient).

In [55] the authors explored the effects of iron death on diabetic nephropathy by analyzing the iron death score of diabetic nephropathy based on the network and pathway levels. They also applied principal component analysis method in order to get transcription factors and non-coding genes, which interact with the Hub gene.

They used the human protein interaction database (STRING) to analyze the protein interaction of key module genes. Using Cytoscape, the authors draw the Protein–Protein Interactions (PPI) network diagram of candidate key genes (First screening the interactions with a comprehensive score >900 to obtain protein–protein interactions and then screening the genes with protein interactions >6 - the candidate key genes). In addition to this, Cytoscape plotted the results of the principal component analysis method, based on TRRUST V2 database and RAID v2.0 database, in order to find out the main regulatory factors that interact with the Hub gene ($p < 0.05$).

In another work, [56] the crotonylation profiling of fish pathogen, S. agalactiae by combining affinity enrichment with LC MS/MS was studied. The crotonylation (Kcr) modification of several selected proteins were further validated by Western blotting. The authors identified 241 Kcr sites from 675 screened proteins. As a result, four crotonylation modified proteins were predicted as virulence factors or being part of the quorum sensing system post-translational modifications on bacteria.

Using the Cytoscape tool and in order to identify the cellular processes regulated through crotonylation in S. agalactiae, a total of 193 proteins were drawn. Visual inspection of the output revealed the highly interconnected proteins for 5 different translation processes that may were regulated by crotonylation in S. agalactiae. The results of PPI networks suggest that lysine crotonylation probably plays a role in regulating post translational modification of protein in S. agalactiae, which contributes to cooperation and coordination of metabolic pathways.

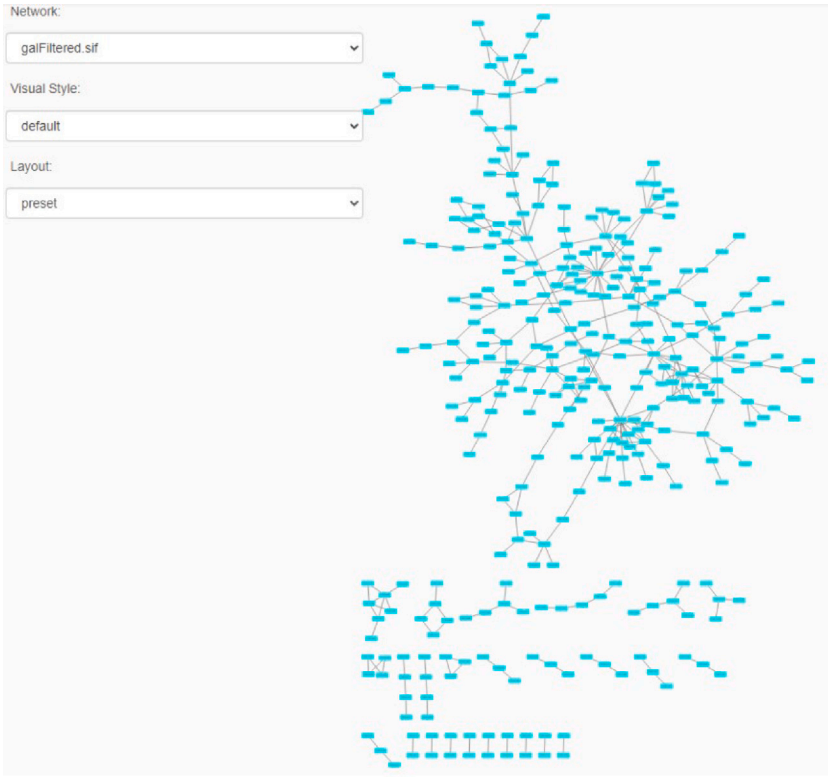


Fig. 8. CYTOSCAPE.

In [57] the authors carried out viral metagenomics studies following in-situ mitomycin C induction experiments in PMMs, in order to unveil the mechanisms by which lysogenic viruses influence their host-microbial community. Compositional changes of viral communities at two different sites were analyzed at the genomic and gene levels, with the aid of the Cytoscape tool.

The authors cluster and visualize closely related viral genomes (vOTUs) from natural and MitC-induced Porcelana viral communities, to build a gene-sharing network contained 2698 nodes (455 vOTUs and 2243 reference genomes), which formed 168 clusters that can be considered viral genera. After the visual inspection from the output of the Cytoscape processing, the authors identified all the Lysogenic vOTUs that were associated with clusters and those that were unclassified (12 and 11 respectively) from the complete Protein-sharing network of Porcelana vOTUs.

In [58] the authors aimed to identify and analyze differentially expressed genes (DEGs) between dermatomyositis and healthy control samples. For this purpose, two microarray skeletal muscle datasets were retrieved from the Gene Expression Omnibus (GEO) database and subsequently analyzed statistically for DEGs. The 63 identified genes were subjected to GO and KEGG functional enrichment analysis to obtain related biological pathways and GO terms. Furthermore, the STRING database was queried with the DEGs and the resulting PPI was imported in Cytoscape and analyzed, and the top 10 hub genes were determined. The 10 hub genes were then tested against different GEO datasets, to determine if their expression levels can help diagnosis of dermatomyositis. ROC-AUC analysis indicated that the top 10 hub genes may exhibit good diagnostic performance, however the authors noted that more experimental evidence is needed to support this conclusion.

In [59], the authors explore the literature-based hypothesized connection of CD147, DCR3 and IL33 with the Kawasaki Disease through the PI3K-AKT pathway. The core of the study was an ELISA laboratory measurement of the above factors in children serum from different stages of IVIG treatment and comparison to healthy control serum. The measurement demonstrated a statistically significant drop in these factors' levels during the treatment, however they were still higher than those of the healthy control group. Additionally, the authors used GeneMANIA and Cytoscape to explore the PPI network around the factors as well as the predicted function of each gene.

In [60] the aim of the authors was to explore the genomic and epigenomic signatures in hepatoblastoma by probing for aberrantly methylated genes. The study consists of two large sections, the first being a bioinformatics-based identification and screening workflow and the second an experimental validation. Starting with three GEO datasets, and after preprocessing, the authors screened statistically for differentially expressed genes and differentially methylated genes, whose intersection was selected. The selected genes were then submitted to STRING and METASCAPE in order to run functional enrichment analyses and determine relevant

pathways, diseases and tissues. Afterwards, the PPI network from STRING was imported into Cytoscape and analyzed for hub genes, whose function was then predicted using GENEMANIA. Using CIBERSORT and statistical analysis, the authors established a correlation between Tumor Infiltrating Immune Cells and the hub genes. To validate their findings, the authors tested the expression levels of four proteins in hepatoblastoma versus nearby normal tissue using western blot and immunohistochemistry. Finally, it was found that transfecting hepatoblastoma cells with anti-NOTUM, NOTUM being one of the validated proteins, decreased tumor volume and suppressed migration and invasion of tumor cells.

The objective of the study in [61] was to investigate and identify miRNAs and the role they might be playing in atrial fibrillation with the use of weighted gene coexpression network analysis. Two GEO datasets served as the source data for this study, containing both AF and non-AF atrial tissue sample data which, after preprocessing, were analyzed for differentially expressed genes and miRNAs (DEGs and DEMs). Using METASCAPE, functional enrichment in GO terms and KEGG pathways was carried out. Using the WGCNA package, the DEM co-expression network was built and, using hierarchical clustering, functional modules were estimated. By processing these modules and the module eigengenes statistically, the key modules were screened. Based on the key modules, hub miRNAs were identified and verified against 3 miRNA target prediction databases. Finally, the miRNA-gene interaction network was defined in Cytoscape using both the PPI network defined from DEGs through METASCAPE as well as the upstream and downstream miRNA-gene relations.

In [62] the authors focus on understanding the connection between the physiological mechanisms of drought resistance and the proteomic profile of drought resistant maize species by comparing the drought stress response of tolerant and sensitive maize species. For the purpose of this study, an in-vivo experiment was carried out using both a drought resistant and a drought sensitive maize variety under both water deficit, as well as normal water control conditions. Leaf samples were collected and processed analytically to determine the proteomic profiles of the two varieties. Differentially expressed proteins were identified and analyzed using GO and KEGG pathway enrichment analysis while the PPI network was built using STRING and Cytoscape, in which four groups of proteins were identified according to function.

2.8. GraProStr

2.8.1. Characteristics

GraProStr is an online tool for graph theoretic analysis of three dimensional protein representations. The tool uses an "Interaction Strength" metric between monomers, in order to construct the network. The interaction between the monomers can be computed on the basis of Ca distances, Cb distances, between side chains or between the protein and a ligand. The user can specify a PDB code or upload a PDB file as input, which is then processed according to the specified cutoff parameter. The tool output includes the graph adjacency matrix, as well as several graph theoretic features of the network such as cliques, hubs, clusters and communities. Cliques are defined as a group of n nodes that are connected to each other (community is a collection of cliques that share n-1 nodes between them). Cliques indicate a kind of aggregation between entities while the presence of communities indicate a set of nodes which have dense connections within the group and sparse connections outside the group. The network is also available for export in the CFinder format for visualization.

2.8.2. Referenced work

The goal of [63] was to study subtle conformational changes of protein side chains when the backbone remains mostly the same, using graph theoretic network measures focused on cliques and communities. In order to explore this approach, three PDB structure datasets were compiled, two of serine proteases with different levels of resolution and one of cationic trypsin. The protein structure network was computed with GraProStr, using the interaction strength of noncovalent side chain interactions, quantified by the number of atoms closer than a proximity cutoff. CFinder was utilized to compute cliques and communities. Comparative measures between different structures considered were unique cliques, unique edges of unique cliques and unique residues of unique cliques (URUC) in one structure compared to another. Based on the geometric proximity of each residue to the catalytic triad, the network was divided in the near, mid and far regions. Finally, URUC measures were compared to degree connectivity and clustering coefficient calculations. The analysis showed that even small conformational changes, as measured by RMSD, can actually indicate important but subtle side chain interaction changes, visible at even distant sites. The same approach can aid in the comparison of side chain features of proteins with similar backbone topology.

The aim of [64] is the study of how different inserts in different protein isoforms lead to functional differences, and specifically what is the effect of the Loop3 insert in TdT. The TdT-short isoform's structure was retrieved from PDB, while the long isoform model and its heteroatoms were generated manually using multiple software approaches and screening the results based on scoring. After preprocessing, such as energy minimization and equilibration, a series of MD simulations were run at different temperatures in GROMACS. Analysis included trajectory PCA, RMSD and RMSF calculations, salt bridge and secondary structure identification with STRIDE and VMD, residue network analysis in GraProStr for cluster and clique analysis and measurement of active site cavity dimensions with PoreWalker. The analysis indicated global changes in the conformation due to the presence of Loop3, increased flexibility and destabilization and gave clues as to why the geometry of the long isoform induces a lower polymerization rate. The contribution of network analysis concluded that several side chain interactions observable in the short isoform were disrupted in the long one, decreasing stability.

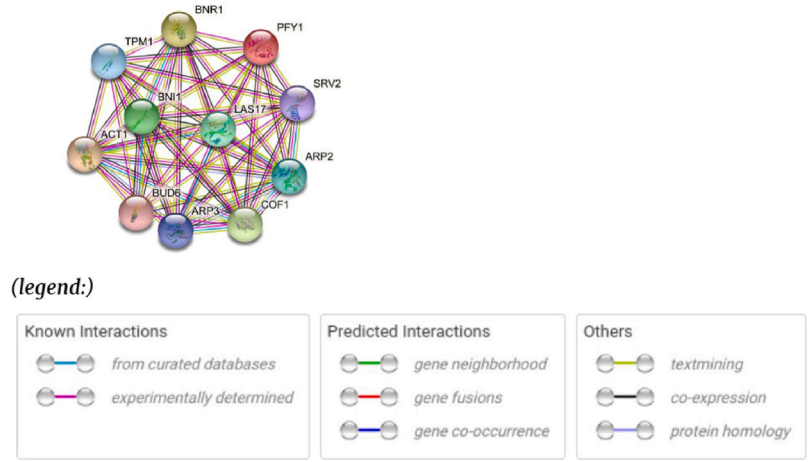


Fig. 9. STRING.

2.9. String

2.9.1. Characteristics

STRING is an online protein–protein interaction database (Fig. 9). The main feature of the tool is how it combines knowledge from a broad variety of publicly available information sources in a single comprehensive data set. Among such information sources, from which protein association data are drawn, are genomic context techniques, co-expression data, text-mining for co-citation based associations, experimental evidence and other databases. Queries to the STRING database consist of one or more proteins or amino-acid sequences. The database response consists of an interaction network including the specified proteins as well as the most strongly associated with them. The user can browse the various proteins and the prediction or evidence-based associations and follow links to data sources in order to inspect details regarding each item. Fig. 9 shows the response of STRING to a query about Saccharomyces cervisiae profilin, which consists of proteins functionally linked to the query. STRING can also run interaction enrichment analyses, through which it is roughly determined if the associations in a given set of proteins are more significant statistically than those of a random set of proteins, and also functional enrichment analysis, which determine what functions (gene ontology terms, pathways, keywords) these proteins may statistically play a role in. K-means and Markov clustering are available for initial network analysis. Query results can be exported to a variety of formats including image, table and XML formats. STRING is also available as a Cytoscape app.

2.9.2. Referenced work

In [65], the authors investigated possible pathways that may be involved in Guizhi/Wullingsan treatment of nephrotic syndrome in humans, with the goal of exploring plausible mechanisms of action. The study was carried out in three steps: target identification, network pharmacology and docking study. The target identification step involved the analysis of the interaction network between active compounds and their targets, which was done by identifying active components, matching them to targets in biological databases and selecting candidate targets. The next step is network pharmacology, in which various interaction networks were analyzed. Here, the compound-target network was built in Cytoscape, the PPI network was queried on STRING, and the overall network of AC-Targets-Pathways was visualized. Gene ontology and KEGG enrichment analysis were also performed. In the docking step, the affinity of Guizhi active compounds when bound to the top 5 predicted targets was found to be good. Overall, this study proposes a theory for the mechanism of action of Guizhi when treating NS, and thus forms a basis for further experimentation.

In [66], hypoxia is investigated as a prognosis for glioma. The main objective is the development of a risk model based on hypoxia related gene expression levels, with the intention of use in glioma prognosis. Sample data from glioma tissue versus normal tissue were retrieved from online databases and hypoxia associated genes were selected from the literature. The PPI network of those genes was retrieved from STRING and subsequently analyzed in Cytoscape. Using network analysis, the top 10 most important hub genes were screened for subsequent regression analysis. Using the LASSO-Cox regression method, a risk model was built using 7 of the 10 hub genes, which was then subjected to Kaplan–Meier analysis to estimate the probability of survival and verified with AUC calculations from ROC curves. Finally, Cox regression was used to find that the hub genes may be independent prognostic factors. Moreover, the model was verified by inspecting the difference in the hub gene expression levels in glioma versus normal tissue and gene set enrichment analysis was used to identify important pathways of these genes.

The goal of [67] was the investigation of differences in the genomic expression profiles between IVF-ET versus naturally conceived placentals, that could form a basis of understanding the differential placental function due to IVF-ET. The study involved the collection of tissue from 3 IVF-ET and 3 naturally conceived placentals after delivery which satisfied specific selection criteria. The tissue was subjected to sequencing and the results were processed statistically to determine differentially expressed genes and

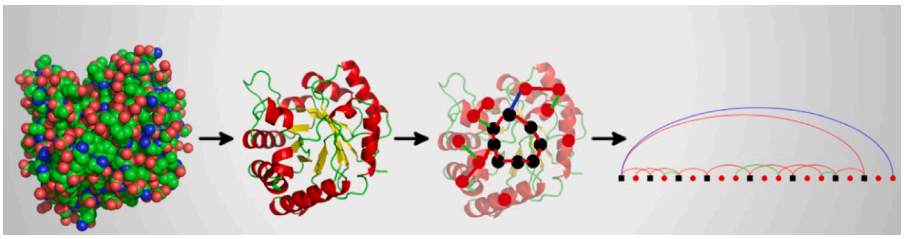


Fig. 10. PTGL.

micro RNA. Using functional enrichment, upstream transcription factors were determined to aid in the analysis of the gene, RNA and protein interaction network. The miRNA–mRNA network was built on Cytoscape by screening 3 interaction databases yielding insight on the difference between over and under-expressed miRNAs. DEG enrichment analysis revealed several connections with GO terms and KEGG pathways. Using STRING and Cytoscape, the PPI network was defined and analyzed for hub genes which were also used for KEGG enrichment.

In [68], the authors aim to study the effects of the phthalates DBP and DEHP on the enzymes involved in the glucocorticoid biosynthesis pathway in the adrenal gland, as well as on pro-inflammatory cytokines in the serum of rats. In order to conduct this study, a laboratory experiment with rats was carried out, at the end of which serum and adrenal gland tissue samples were collected. Histopathological changes were confirmed by inspection of the samples. The adrenal glands were subjected to quantitative analysis for mRNA expression of enzymes of steroidogenesis. Corticosterone and cytokine measurements were performed on the serum. Top 50 queries were submitted to STRING and the results were enriched over the GO and KEGG databases, so as to obtain other candidate mechanisms and organs which involve the same enzymes that could be potentially affected by phthalates.

2.10. PTGL

2.10.1. Characteristics

A tool for examining the secondary structure topology of a protein is PTGL (Fig. 10). This is an online database based on 3D structures of proteins, as provided by the Protein Data Bank. The basic function of PTGL is the analysis of a submitted PDB file or id, which consists of identifying secondary structure elements (SSEs) and their contacts, based on the contacts of individual residues within each SSE. The result is a graph whose vertices are SSEs, annotated as alpha or beta structures and whose edges are SSE contacts, annotated as parallel, antiparallel or mixed. Once a protein SSE graph is formed, it can be visualized on the protein with JMol or it can be exported as an image or in a graph data format. PTGL also supports viewing each connected component individually and representing it in one of four linear notations. Furthermore, these representations are searchable, thus facilitating the identification of proteins that share similar secondary structure level topologies.

2.10.2. Referenced work

In [69] a database system for secondary structure based protein queries and annotation was described. The system, called PASSTA, aims to be a homology database that receives amino acid sequences as queries and, based on the secondary structure elements and topology, attempts to find proteins of homologous structure. PASSTA relied on the integration of three secondary source databases: PTGL, PDBfinder II and SCOP. The three sources are integrated using a dedicated two pass annotation algorithm, where the first one is a filtering pass and the second one is the actual annotation. The article also comments on the results of two queries.

3. Graph theory for computer-aided drug design

The application of graph-theory-based simulation tools for protein structure networks is relevant upon attempting to develop drug candidate molecules capable to recognize and molecularly interact with proteins implicated in disease pathophysiology. Drug discovery is a long-lasting and costing process that needs around ten to fifteen years for a drug to reach the market [70]. Drug discovery begins with the identification of the appropriate drug target, hit to lead discovery, optimization of lead molecules and finishes with the preclinical and clinical studies [71]. However, in the most cases the success rate of drugs through clinical trials is only 13% and usually due to lack of optimal pharmacokinetic properties and toxicity [72].

Nowadays, computer-aided drug discovery (CADD) techniques are widely used in preliminary studies by researchers and have aided to speed up the drug discovery and development process diminishing the costs and failures in the clinical trials [73]. Rational drug design is a vital part of CADD which help understanding of the binding affinity and molecular interactions between a target protein and a ligand and has been simplified by the development of supercomputers, parallel processing, and advanced programs and algorithms. Additionally, the current improvements in machine learning methods have significantly supported the analyzing of pharmaceutical-related big data in the drug discovery process [74].

In order to identify new inhibitors from chemical databases, different methods can be used, including pharmacophore modeling, quantitative structure–activity relationship (QSAR), molecular docking, quantum mechanics and statistical methods. For the

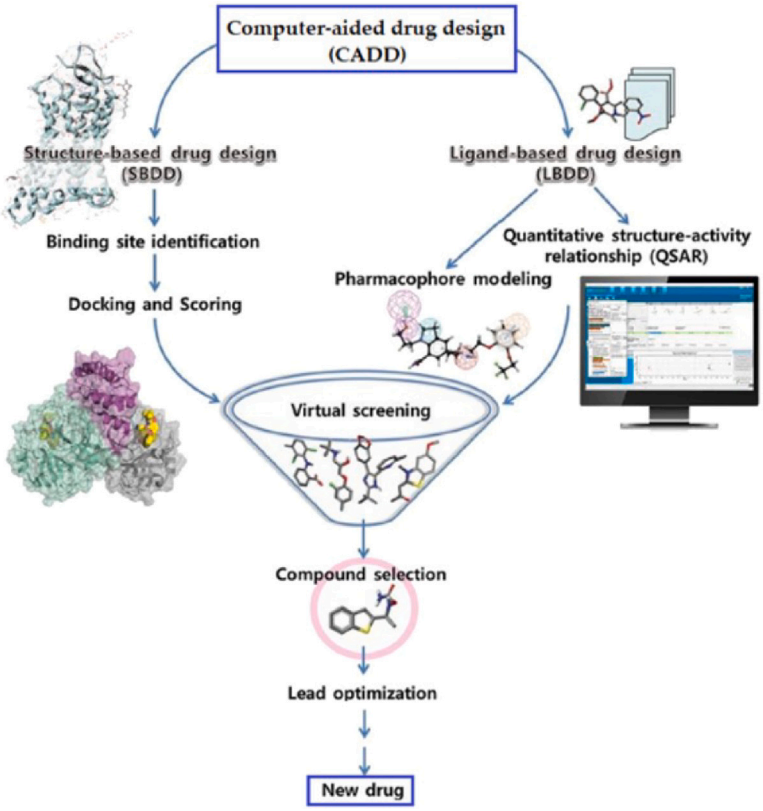


Fig. 11. Depiction of the drug discovery process through the computer-aided drug design and the exploitation of protein structure simulation tools.

identification of lead molecules to categories of computer-aided drug discovery, can be used two drug design approaches, the structure-based and ligand-based. The structure-based drug design approach depends on the three-dimensional structure of the receptor to understand the interactions between the receptor–ligand, while the ligand based-drug design depends on the interaction of known ligands with the receptor [75] (Fig. 11). Computer-aided drug design has been used successfully over the years [76] and can play a critical role in discovery new drug candidates against coronavirus disease 2019 (COVID-19).

Moreover, with the availability of complete genome sequence of SARS-CoV-2 and X-ray structures of the viral proteins, computer-aided drug design can be an important tool for the identification of novel SARS-CoV-2 agents. Structure-based drug design uses different methods including molecular docking, structure-based virtual screening and MD. These methods have been used by many pharmaceutical industries and researchers and aided in the development of many commercial drugs, such as amprenavir, a protease inhibitor of the human immunodeficiency virus using molecular docking and MD simulations [77], norfloxacin a topoisomerase II and IV inhibitor [78], antituberculosis drug, isoniazid, discovered through structure-based virtual screening and pharmacophore modeling [79] and flurbiprofen, targeting cyclooxygenase-2, a nonsteroidal anti-inflammatory drug against rheumatoid arthritis, osteoarthritis, discovered through molecular docking [80]. SBDD hollow a series of steps including of the target structure preparation, identification of the ligand binding site, preparation of compounds library, molecular docking, molecular dynamic simulation and calculation of the binding free energy.

Molecular docking is a computational method that studies the interactions between a target receptor and a ligand at a molecular level and ranks the ligands using various scoring functions [81]. There are a wide variety of molecular docking programs that can be used such as AutoDock [82], AutoDock Vina [83], GOLD [4], FlexX [84], GLIDE [85], etc. Molecular docking can be divided into two categories: flexible-ligand search docking and flexible-protein docking. With the absence of approved drugs and affective vaccines for COVID-19 and the accessibility in the complete genome sequence of SARS-CoV-2 [86] and structural elucidation of its proteins, the research for novel antiviral agents against COVID-19 disease has been rapidly pursued.

Researchers are swiftly working on identifying and designing inhibitors against all possible viral key protein targets of SARS-CoV-2 such as the structural proteins spike, envelope, membrane, and nucleocapsid and nonstructural proteins the main protease which is also known as 3C-like protease 3CLpro, papain-like protease, RNA-dependent RNA polymerase, nsp16 2-O-methyltransferase, nsp15 endoribonuclease and nsp13 helicase. The structures of all these proteins can be used for structure-based virtual screening for identification of specific inhibitors of the target proteins. A strong example is the study of [86]. Khan et al. using molecular docking and molecular dynamic simulation studies, target chymotrypsin-like protease (3CLpro). The results showed that three FDA

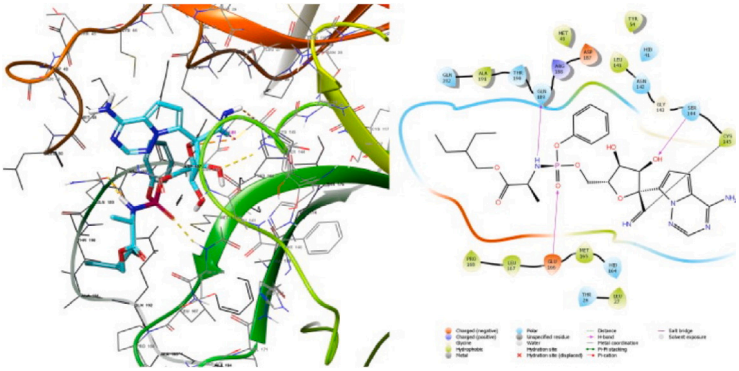


Fig. 12. Covalent docking analysis of remdesivir inside the SARS-CoV-2 Mpro.

approved drugs remdesivir, saquinavir, and darunavir and two natural compounds flavone and coumarin derivatives could act as promising inhibitors of chymotrypsin-like protease enzyme [86].

Moreover, Al-Khafaji et al. performed a covalent docking screening procedure joined with the molecular dynamic simulation studies to identify molecules that can form a covalent bond with residue Cys145 within the binding pocket of SARS-CoV-2 main protease and identified saquinavir, ritonavir and remdesivir FDA approved drugs as the most promising inhibitors of SARS-CoV-2 Mpro [87]. Finally, Ibrahim et al. combined molecular docking and molecular dynamic approaches to explore the potentialities of eighteen repurposed drugs in clinical development against SARS-CoV-2 Mpro. The results revealed that TMC-310911 and ritonavir could be promising drugs for the treatment of COVID-19 [88].

In addition, Petrou et al. [89], combined the traditional medicinal chemistry, structural biology, and computational chemistry. They design a series of new compounds that combine in their structure the minimum pharmacophores required to inhibit the main protease of SARS-CoV-2 Mpro (Fig. 12), using molecular docking studies and the followed experimental data revealed that among the fifteen compounds chosen, five compounds showed inhibitory activity with IC50 within the range of 0.01–34.4 uM. It is noteworthy to mention that these data provide evidence on the potential antiviral activity of these compounds against the main protease of SARS-CoV-2, to serve as potential candidates for COVID-19 therapeutics.

3.1. Discussion

With regards to Tables 1 and 2, every tool characteristic is analyzed so as to offer a better readership while a categorization of their main applications is done via the selected references. The different network constructions refers to the provided option to construct various network representations based on different forms of edges while the various global properties refers to the number of different parameters displayed, related to the characteristics of the constructed network, like the number of nodes and edges. Characteristics like 3D view, network view, contact map, distance matrix and sub-network view are all related to the different options for visualizing the structure of the loaded protein and most of the time along with the three dimensional view. It is interesting enough that the presented tools may have different licences as shown in Table 1, but the main characteristic remains the same: to help in the production of new knowledge with ease of distribution.

Furthermore, node centrality analysis stands for some of the most common node-based centrality measures like degree and betweenness while edge centrality combines the edge betweenness value with the optical inspection of a contact map and the corresponding 3d view of the specific PCN. The shortest path analysis shows (visually and numerically) the minimum number of nodes that needs to be traversed in order to reach one node from another. In addition, the K-clique analysis characteristic shows a sub-network of k nodes with all the k nodes being connected to each other. The graph spectral analysis provides spectral analysis of adjacency and Laplacian matrices. Besides that, analysis of the protein complex refers to the characteristic of a tool to perform analysis on a protein complex with up to four chains or a complex of 2–4 interacting proteins. The next characteristic is the RNA network analysis feature, which offers the equivalent analysis of a protein to a RNA structure. By the same token, Molecular Dynamics analysis provides an option to analyze MD trajectory data for a number of properties like Cross Correlation Map, centrality analysis and shortest path analysis.

Most of the tools have some support for data exports in a variety of formats. Clustering refers to the capability of a tool to group network nodes into clusters using graph-specific clustering algorithms. Some tools offer the option of processing protein–protein interaction networks. Graphs of secondary structure elements are referred to the visualization of formations like alpha helices and beta sheets.

A closer inspection of table A sheds light on the functional characteristics of each tool. As the most common characteristic between them, stands their potential to define with various ways the notion of network edge and network node. As a direct consequence of this, the majority of the tools also offer information about various global network characteristics. The filtering capability, the feature that gives the user the option to select nodes based on a common characteristic, is not present in the contact map analysis tools like the CMView and Cmweb while from the other tools Graprostr does not give this option. From the tools that

Table 1
Features of each presented tool.

	Prosnex	Naps	Cmview	Webpsn	RING 2.0	Cytoscape	Cmweb	GraProStr	STRING	PTGL
Features supported by each tool										
Various options in network construction	•	•		•		•			•	•
Network view + filtering	•	•	•	•	•	•			•	
Various global properties	•	•	•	•			•2			•2
3d molecule view	•	•	•				•			
Contact map/distance matrix	•	•	•	•						
Differenct PDB analysis - correlation (structure/trajectory)			•	•						
MetaPath analysis				•						
Node centrality analysis	•	•			•	•				
Edge centrality		•			•	•1				
Shortest path analysis	•	•		•		•1				
k-clique analysis	•	•		•		•1		•		
Graph spectral analysis		•				•1				
Analysis of Protein complexes A and B		•		•		•1				
Entropy/energy		•			•					
RNA Network analysis		•								
Network Analysis of Molecular Dynamics Data/Trajectory		•		•						
Protein–protein interaction networks		•				•1			•	
Graph clustering		•				•1		•	•	
Graphs of protein secondary structure elements	•	•			•					•
Licence Agreement	No licence	No licence	GPL v3	No licence	No licence	LGPL v2.1	No licence	No licence	Web Interface-no licence PSICQUIC-Apache 2.0 Db - CC BY 4.0	Artistic License v2.0

1. Cytoscape features marked with a (•1) require extra plugins.
2. Some tools (•2) offer 3D molecule views with older versions of Jmol, which require deprecated Java plugins on the user's browser. is absent from modern web browsers.
3. All tools support at least one export option.

Table 2
Scientific works associated to the presented tools.

Tools	Physiologies/Pathophysiologies	Virus, bacterial and other diseases	Others (ecology)
PROSNEX	[29]		
NAPS	[31,33–36]	[30,32]	
CMVIEW	[39]; [90]; [40,41]	[37]	
CMWeb			
WebPSN	[44,45,47]	[46,48,49]	
Ring 2.0	[47,53]; [91]; [51]	[50]	
Cytoscape	[54,55,58–61]	[56,57]	[62]
GraProStr	[92]; [64]		
STRING	[65–68]		
PTGL	[40,41,69]		

offer the 3D view option, it should be mentioned that CMweb and PTGL offers this through older versions of Jmol, via the Jmol applet which is not supported anymore from the major web browsers.

Contact map analysis along with distance matrix is a feature offered from NAPS and Prosnex, being the only non contact map analysis tools having this option. Cmweb and CMView that offer this analysis model are contact map analysis specialized tools and

they offer a limited number of additional features related to network graph analysis. On the other hand, analysis between two pdb files is a feature of CMView along with the webPSN. The latter, except of the option to analyze different protein complexes, it offers a unique feature named metapath analysis that filters the shortest path that contains only residues with a correlation ≥0.7 with at least one of the two path extremities (i.e. the first and last amino acids in the path). These filtered paths were used to build the global meta path, which is made of the most recurrent links, i.e. the links that are present in a number of paths ≥10% of the number of paths in which the most recurrent link is present. According to the developers of the tool, such metapath represents a coarse/global picture of the structural interconnection state in the considered system.

Node, edge, shortest path and k-clique analysis are offered from the NAPS tool being the one that provides a plethora of information, based mainly in the protein structure graph analysis. Moreover, graph spectral analysis is a unique feature of NAPS in addition to the RNA analysis one. What should be mentioned is that from the other tools that offer the aforementioned information, Cytoscape needs additional plugins in order to fully report those details. Likewise, RING also provides a part of this information but for additional extracts Cytoscape integration is needed.

WebPSN and NAPS also offer analysis of Molecular Dynamics trajectory data while String and NAPS provide information about protein–protein interaction networks. (As stated above, Cytoscape needs additional plugins so as to export this information). Among all the tools, graph clustering is provided by GraProStr, String and NAPS and differs from k-clique analysis as the latter specifies a group of nodes that are connected to each other while in a cluster all nodes can be reached from all other nodes of it by a path inside this cluster.

Pertaining to some unique characteristics of the tools presented, RING 2.0 provides the Shannon's entropy of a given residue while STRING offers a built-in functional enrichment analysis feature, which statistically assesses categories, such as biological functions or pathways from relevant databases, in which a set of genes is overrepresented. Finally, networks generated with PTGL define nodes as secondary structure elements such as alpha helices and beta sheets which is also a feature of NAPS.

The main role these tools can play in biological studies is to provide an additional perspective in the analysis of interactions between residues of a protein or between proteins themselves. The way this happens is by allowing the researcher to manipulate a more abstract representation of these interactions. This representation can capture essential structural information and filter out a lot of noise. The analysis of this graph-based representation then takes place using the techniques of graph theory. In this phase, the objective is to study and characterize structural features such as neighborhoods of interrelated residues, or proteins, or central proteins and genes that play a crucial role in a pathway. At the end, this method of analysis may be somewhat crude, however the aim is not to establish evidence but to aid in the conceptualization of relations and interactions at the molecular level, corroborate high level characterization of processes and provide guidance for further investigation. As these tools do not demand specialized personnel to operate them, they offer also the flexibility and ease to be used as an additional tool for scientific research in biological entities.

Based on principle roles of proteins in biological function, their interactions determine molecular and cellular mechanisms, which control healthy and diseased states in organisms. Network analysis and graph analysis of protein interaction networks facilitate the understanding of pathogenic and physiologic mechanisms that trigger the onset and progression of diseases like cancer and autoimmune disorders. By integrated known genetic modifiers of prostate cancer with expression dynamics and protein interaction networks Ergun [93] team were lead to the development of methods which were suitable enough to reveal molecular network differences between aggressive and non-aggressive prostate cancers. Cancer proteins particularly tend to interact with more number of proteins and this is reflected to the network centralities and therefore they tend to participate in graphs as backbone of the proteome rather than in a network brunches. In another example of network analysis used in real life scenarios, the use of PPI networks was effective enough to prioritize positional candidate disease genes for Type 1 Diabetes [94].

In an effort to categorize the way the tools, presented in this work, are utilized, we could discern the following categories:

- Residue networks only
  The study of residue interaction networks can provide insight on several aspects of the conformational and functional characteristics of a protein molecule. Centrality analysis, where graph algorithms are used to quantify the importance of graph nodes and edges, can help identify hub residues that are most important from a structural perspective, for example in the number of interactions in which they are involved, or in the number of signaling pathways which they mediate, or for participating in important groups of nodes such as cliques or communities. Such a structural importance may be linked to other characteristics such as a role as a conformational stabilizer or an allosteric interaction regulator.
- Residue networks and contact maps
  Comparative analysis of residue interaction networks and contact maps can provide input as to which interactions are preserved and which are not, across conformational changes. Such information can help reveal the structural and functional role of residues.
- Protein–protein interaction networks
  The graph-theoretic study of protein–protein interaction networks is another analysis technique which can clarify the functional aspect of protein sets. In this case, the graph representation is compiled from a multitude of sources into a database which can be then explored as part of a research workflow, to identify protein interaction links based on previous research work and other factors. There are two main uses of a protein–protein interaction database. The first is to determine, given a set of proteins, what other proteins they interact with. The second is to find, given a set of proteins, what biological pathways they collectively play a role in, using enrichment analysis. Applying centrality analysis to protein interaction networks can provide a designation of which proteins or genes are hubs. This can provide a guide for example in compound target selection and screening.

4. Conclusion

Presentation of experimental datasets, visualization and interactive navigation of them, along with computation of network centralities and formations that are difficult to be detected from observation data, due to various types of noise, were realized through the applied graph theory which has pushed the envelope of biological scientific research. Under this rationale, a plethora of tools were developed in order to utilize graph theory to the advantage of biology. Multifarious interactions among biological entities are processed via a mathematical formalism, generating different types of data, the processing of which leads to the revelation of biological interactions. This work, based on the latest publications, presents a complete list of tools that are able to extract in a graph-related manner the relations between residues, proteins and biological macromolecules. The work stands by a network perspective and tries to shed light on the biological aspect of graph theory. Future work will be focus on the combination of the presented used of graphs in a tool that will combine not only the graph theory but inherent ideas from network theory to extract helpful biological insights.

Data availability

No data was used for the research described in the article.

Acknowledgments

The authors thank Christos Ouzounis from School of Informatics, Aristotle University of Thessaloniki, for provided insight and expertise that assisted the research.

References

[1] M.S. Vijayabaskar, S. Vishveshwara, Interaction energy based protein structure networks, Biophys. J. 99 (11) (2010) 3704–3715, http://dx.doi.org/10.1016/j.bpj.2010.08.079.
[2] H. Watanabe, K. Yamasaki, S. Honda, Tracing primordial protein evolution through structurally guided stepwise segment elongation, J. Biol. Chem. 289 (6) (2014) 3394–3404, http://dx.doi.org/10.1074/jbc.M113.530592.
[3] M.W. Gonzalez, M.G. Kann, F. Lewitter, Chapter 4: Protein interactions and disease, PLoS Comput. Biol. 8 (12) (2012) e1002819, http://dx.doi.org/10.1371/journal.pcbi.1002819.
[4] G. Jones, P. Willett, R.C. Glen, A.R. Leach, R. Taylor, Development and validation of a genetic algorithm for flexible docking, J. Mol. Biol. 267 (3) (1997) 727–748, http://dx.doi.org/10.1006/jmbi.1996.0897.
[5] Z. Banday, G. Ashraf, Protein-protein interactions as potential targets of drug designing, 2018, pp. 1–15.
[6] S. Jaeger, P. Aloy, From protein interaction networks to novel therapeutic strategies, IUBMB Life 64 (6) (2012) 529–537, http://dx.doi.org/10.1002/iub.1040.
[7] G. Astras, C.I. Papagiannopoulos, K.A. Kyritsis, C. Markitani, I.S. Vizirianakis, Pharmacogenomic testing to guide personalized cancer medicine decisions in private oncology practice: A case study, Front. Oncol. 10 (2020) 394.
[8] N. Mittas, F. Chatzopoulou, K.A. Kyritsis, C.I. Papagiannopoulos, N.F. Theodoroula, A.S. Papazoglou, E. Karagiannidis, G. Sofidis, D.V. Moysidis, N. Stalikas, A. Papa, D. Chatzidimitriou, G. Sianos, L. Angelis, I.S. Vizirianakis, A risk-stratification machine learning framework for the prediction of coronary artery disease severity: Insights from the GESS trial, Front. Cardiovasc. Med. 8 (2021) 812182, http://dx.doi.org/10.3389/fcvm.2021.812182.
[9] G. Agapito, P.H. Guzzi, M. Cannataro, Visualization of protein interaction networks: problems and solutions, BMC Bioinformatics 14 (S1) (2013) 1, http://dx.doi.org/10.1186/1471-2105-14-S1-S1.
[10] M. Koutrouli, E. Karatzas, D. Paez-Espino, G.A. Pavlopoulos, A guide to conquer the biological network era using graph theory, Front. Bioeng. Biotechnol. 8 (2020) 179, http://dx.doi.org/10.3389/fbioe.2020.00034.
[11] A.-L. Barabási, Z.N. Oltvai, Network biology: understanding the cell functional organization, Nature Rev. Genet. 5 (2) (2004) 101–113, http://dx.doi.org/10.1038/nrg1272.
[12] J. Salamanca Viloria, M.F. Allega, M. Lambrughi, E. Papaleo, An optimal distance cutoff for contact-based protein structure networks using side-chain centers of mass, Sci. Rep. 7 (1) (2017) 2838, http://dx.doi.org/10.1038/s41598-017-01498-6.
[13] D. Barker, M. Pagel, Predicting functional gene links from phylogenetic-statistical analyses of whole genomes, PLoS Comput. Biol. 1 (1) (2005) e3, http://dx.doi.org/10.1371/journal.pcbi.0010003.
[14] A. Chatr-aryamontri, A. Ceol, L.M. Palazzi, G. Nardelli, M.V. Schneider, L. Castagnoli, G. Cesareni, MINT: the molecular INTeraction database, Nucleic Acids Res. 35 (Database issue) (2007) D572–D574, http://dx.doi.org/10.1093/nar/gkl950.
[15] A. Dandekar, R. Baker, M.A. Vannice, Characterization of activated carbon, graphitized carbon fibers and synthetic diamond powder using TPD and DRIFTS, Carbon 36 (12) (1998) 1821–1831, http://dx.doi.org/10.1016/S0008-6223(98)00154-7.
[16] E.V. Koonin, Genome sequences: Genome sequence of a model prokaryote, Curr. Biol. 7 (10) (1997) R656–R659, http://dx.doi.org/10.1016/s0960-9822(06)00328-9.
[17] T.D. Wilson, Fifty years of information behavior research, Bull. Amer. Soc. Inf. Sci. Technol. 36 (3) (2010) 27–34, http://dx.doi.org/10.1002/bult.2010.1720360308.
[18] M. Bhattacharyya, S. Ghosh, S. Vishveshwara, Protein structure and function: Looking through the network of side-chain interactions, Curr. Protein Pept. Sci. 17 (1) (2016) 4–25, http://dx.doi.org/10.2174/1389203716666150923105727.
[19] J.C.H. Sim, S.M. White, P.J. Lockhart, ARID1B-mediated disorders: Mutations and possible mechanisms, Intractable Rare Dis. Res. 4 (1) (2015) 17–23, http://dx.doi.org/10.5582/irdr.2014.01021.
[20] M.E. Parente, A. Ochoa Andrade, G. Ares, F. Russo, Jimnez-Kairuz, Bioadhesive hydrogels for cosmetic applications, Int. J. Cosmet. Sci. 37 (5) (2015) 511–518, http://dx.doi.org/10.1111/ics.12227.
[21] S.-Y. Huang, Search strategies and evaluation in protein-protein docking: principles, advances and challenges, Drug Discov. Today 19 (8) (2014) 1081–1096, http://dx.doi.org/10.1016/j.drudis.2014.02.005.
[22] G. Bagler, S. Sinha, Network properties of protein structures, Physica A 346 (1–2) (2005) 27–33, http://dx.doi.org/10.1016/j.physa.2004.08.046.
[23] S. Vishveshwara, A. Ghosh, P. Hansia, Intra and inter-molecular communications through protein structure network, Curr. Protein Pept. Sci. 10 (2) (2009) 146–160, http://dx.doi.org/10.2174/138920309787847590.

[24] K.V. Brinda, S. Vishveshwara, A network representation of protein structures: implications for protein stability, Biophys. J. 89 (6) (2005) 4159–4170, http://dx.doi.org/10.1529/biophysj.105.064485.
[25] A. Ghosh, S. Vishveshwara, Variations in clique and community patterns in protein structures during allosteric communication: investigation of dynamically equilibrated structures of methionyl tRNA synthetase complexes, Biochemistry 47 (44) (2008) 11398–11407, http://dx.doi.org/10.1021/bi8007559.
[26] A. Gilman, A.P. Arkin, GENETIC CODE: Representations and dynamical models of genetic components and networks, Annu. Rev. Genom. Hum. Genet. 3 (1) (2002) 341–369, http://dx.doi.org/10.1146/annurev.genom.3.030502.111004.
[27] H. Yu, P. Braun, M.A. Yıldırım, I. Lemmens, K. Venkatesan, J. Sahalie, T. Hirozane-Kishikawa, F. Gebreab, N. Li, N. Simonis, T. Hao, J.-F. Rual, A. Dricot, A. Vazquez, R.R. Murray, C. Simon, L. Tardivo, S. Tam, N. Svrzikapa, C. Fan, A.-S.d. Smet, A. Motyl, M.E. Hudson, J. Park, X. Xin, M.E. Cusick, T. Moore, C. Boone, M. Snyder, F.P. Roth, A.-L. Barabási, J. Tavernier, D.E. Hill, M. Vidal, High-quality binary protein interaction map of the yeast interactome network, Science 322 (5898) (2008) 104–110, http://dx.doi.org/10.1126/science.1158684.
[28] J.d. Las Rivas, C. Fontanillo, F. Lewitter, Fran Lewitter, Protein-protein interactions essentials: Key concepts to building and analyzing interactome networks, PLoS Comput. Biol. 6 (6) (2010) e1000807, http://dx.doi.org/10.1371/journal.pcbi.1000807.
[29] D. Batkhishig, P. Enkhbayar, R.H. Kretsinger, N. Matsushima, A crucial residue in the hydrophobic core of the solenoid structure of leucine rich repeats, Biochim. Biophys. Acta Proteins Proteom. 1869 (6) (2021) 140631, http://dx.doi.org/10.1016/j.bbapap.2021.140631.
[30] F. Jafary, S. Jafari, M.R. Ganjalikhany, In silico investigation of critical binding pattern in SARS-CoV-2 spike protein with angiotensin-converting enzyme 2, Sci. Rep. 11 (1) (2021) 6927, http://dx.doi.org/10.1038/s41598-021-86380-2.
[31] N.B. Bloch, T.E. Wales, M.S. Prew, H.R. Levy, J.R. Engen, L.D. Walensky, The conformational stability of pro-apoptotic BAX is dictated by discrete residues of the protein core, Nature Commun. 12 (1) (2021) 4932, http://dx.doi.org/10.1038/s41467-021-25200-7.
[32] A. Amusengeri, L. Astl, K. Lobb, G.M. Verkhivker, O. Tastan Bishop, Establishing computational approaches towards identifying malarial allosteric modulators: A case study of plasmodium falciparum hsp70s, Int. J. Mol. Sci. 20 (22) (2019) http://dx.doi.org/10.3390/ijms20225574.
[33] A.P. Kumar, S. Lukman, Allosteric binding sites in rab11 for potential drug candidates, PLoS One 13 (6) (2018) e0198632, http://dx.doi.org/10.1371/journal.pone.0198632.
[34] L. Deng, Y. Sui, J. Zhang, XGBPRH: Prediction of binding hot spots at protein⁻RNA interfaces utilizing extreme gradient boosting, Genes 10 (3) (2019) http://dx.doi.org/10.3390/genes10030242.
[35] N.A. Jonniya, M.F. Sk, P. Kar, Investigating phosphorylation-induced conformational changes in WNK1 kinase by molecular dynamics simulations, ACS Omega 4 (17) (2019) 17404–17416, http://dx.doi.org/10.1021/acsomega.9b02187.
[36] S. Agajanian, O. Oluyemi, G.M. Verkhivker, Integration of random forest classifiers and deep convolutional neural networks for classification and biomolecular modeling of cancer driver mutations, Front. Mol. Biosci. 6 (2019) 44, http://dx.doi.org/10.3389/fmolb.2019.00044.
[37] S. Ghaderi, M.R. Bozorgmehr, M. Ahmadi, S. Tarahomjoo, Identification of conformational B-cell epitopes in diphtheria toxin at varying temperatures using molecular dynamics simulations, Arch. Razi Inst. 75 (4) (2021) 427–437, http://dx.doi.org/10.22092/ari.2019.127251.1377.
[38] H. Monhemi, M.R. Housaindokht, The molecular mechanism of protein denaturation in supercritical CO2: The role of exposed lysine residues is explored, J. Supercrit. Fluids 147 (1) (2019) 222–230, http://dx.doi.org/10.1016/j.supflu.2018.11.004.
[39] A.M. Melzer, N. Palanisamy, Deleterious single nucleotide polymorphisms of protein kinase R identified by the computational approach, Mol. Immunol. 101 (2018) 65–73, http://dx.doi.org/10.1016/j.molimm.2018.05.026.
[40] C.T.-T. Su, W.-H. Lua, J.-J. Poh, W.-L. Ling, J.Y. Yeo, S.K.-E. Gan, Molecular insights of nickel binding to therapeutic antibodies as a possible new antibody superantigen, Front. Immunol. 12 (2021) 676048, http://dx.doi.org/10.3389/fimmu.2021.676048.
[41] R.A. Arreola-Barroso, A. Llopiz, L. Olvera, G. Saab-Rincón, Modulating glycoside hydrolase activity between hydrolysis and transfer reactions using an evolutionary approach, Molecules (Basel, Switzerland) 26 (21) (2021) http://dx.doi.org/10.3390/molecules26216586.
[42] C. Röder, N. Vettore, L.N. Mangels, L. Gremer, R.B.G. Ravelli, D. Willbold, W. Hoyer, A.K. Buell, G.F. Schröder, Atomic structure of PI3-kinase SH3 amyloid fibrils by cryo-electron microscopy, Nature Commun. 10 (1) (2019) http://dx.doi.org/10.1038/s41467-019-11320-8.
[43] R. Coppée, A. Sabbagh, J. Clain, Structural and evolutionary analyses of the Plasmodium falciparum chloroquine resistance transporter, Sci. Rep. 10 (1) (2020) http://dx.doi.org/10.1038/s41598-020-61181-1.
[44] A. Yazhini, S. Sandhya, N. Srinivasan, Rewards of divergence in sequences, 3-D structures and dynamics of yeast and human spliceosome SF3b complexes, Curr. Res. Struct. Biol. 3 (2021) 133–145, http://dx.doi.org/10.1016/j.crstbi.2021.05.003.
[45] K. Singh, J.M. Briggs, Impact of lymphoma-linked Asn11Tyr point mutation on the interaction between Bcl-2 and a BH3 mimetic: Insights from molecular dynamics simulation, Chem. Biol. Drug Des. 95 (4) (2020) 435–450, http://dx.doi.org/10.1111/cbdd.13653.
[46] S. Lata, M. Akif, Comparative protein structure network analysis on 3CLpro from SARS-CoV-1 and SARS-CoV-2, Proteins 89 (9) (2021) 1216–1225, http://dx.doi.org/10.1002/prot.26143.
[47] I.N. Smith, S. Thacker, M. Seyfi, F. Cheng, C. Eng, Conformational dynamics and allosteric regulation landscapes of germline PTEN mutations associated with autism compared to those associated with cancer, Am. J. Hum. Genet. 104 (5) (2019) 861–878, http://dx.doi.org/10.1016/j.ajhg.2019.03.009.
[48] P.K. Parida, D. Paul, D. Chakravorty, Free energy landscapes and residue network analysis for Six SARS-CoV-2 targets in complex with plausible phytochemical inhibitors from withania somnifera: 1 μ S molecular dynamics simulations, 2021, http://dx.doi.org/10.26434/chemrxiv.14499306.v1.
[49] L. González-Paz, M.L. Hurtado-León, C. Lossada, F.V. Fernández-Materán, J. Vera-Villalobos, M. Loroño, J.L. Paz, L. Jeffreys, Y.J. Alvarado, Comparative study of the interaction of ivermectin with proteins of interest associated with SARS-CoV-2: A computational and biophysical approach, Biophys. Chem. 278 (2021) 106677, http://dx.doi.org/10.1016/j.bpc.2021.106677.
[50] G.M. Verkhivker, S. Agajanian, D.Y. Oztas, G. Gupta, Comparative perturbation-based modeling of the SARS-CoV-2 spike protein binding with host receptor and neutralizing antibodies: Structurally adaptable allosteric communication hotspots define spike sites targeted by global circulating mutations, Biochemistry 60 (19) (2021) 1459–1484, http://dx.doi.org/10.1021/acs.biochem.1c00139.
[51] S. Kalyaanamoorthy, S.M. Lamothe, X. Hou, T.C. Moon, H.T. Kurata, M. Houghton, K.H. Barakat, A structure-based computational workflow to predict liability and binding modes of small molecules to hERG, Sci. Rep. 10 (1) (2020) http://dx.doi.org/10.1038/s41598-020-72889-5.
[52] A. Gyulkhandanyan, A.R. Rezaie, L. Roumenina, N. Lagarde, V. Fremeaux-Bacchi, M.A. Miteva, B.O. Villoutreix, Analysis of protein missense alterations by combining sequence- and structure-based methods, Mol. Genet. Genom. Med. 8 (4) (2020) http://dx.doi.org/10.1002/mgg3.1166.
[53] M. Molina, C. Moulis, N. Monties, S. Pizzut-Serin, D. Guieysse, S. Morel, G. Cioci, M. Remaud-Siméon, Deciphering an undecided enzyme: Investigations of the structural determinants involved in the linkage specificity of alternansucrase, ACS Catal. 9 (3) (2019) 2222–2237, http://dx.doi.org/10.1021/acscatal.8b04510.
[54] C.A. Spilling, M.-P.K. Dhillon, D.R. Burrage, S. Ruickbie, E.H. Baker, T.R. Barrick, P.W. Jones, Factors affecting brain structure in smoking-related diseases: Chronic obstructive pulmonary disease (COPD) and coronary artery disease, PLoS One 16 (11) (2021) e0259375, http://dx.doi.org/10.1371/journal.pone.0259375.
[55] Y. Hu, S. Liu, W. Liu, Z. Zhang, Y. Liu, D. Sun, M. Zhang, J. Fang, Bioinformatics analysis of genes related to iron death in diabetic nephropathy through network and pathway levels based approaches, PLoS One 16 (11) (2021) e0259436, http://dx.doi.org/10.1371/journal.pone.0259436.
[56] X. Chen, B. Fan, C. Fan, Z. Wang, E. Wangkahart, Y. Huang, Y. Huang, J. Jian, B. Wang, First comprehensive proteome analysis of lysine crotonylation in streptococcus agalactiae, a pathogen causing meningoencephalitis in teleosts, Proteome Sci. 19 (1) (2021) 14, http://dx.doi.org/10.1186/s12953-021-00182-y.

[57] S. Guajardo-Leiva, F. Santos, O. Salgado, C. Regeard, L. Quillet, B. Díez, Unveiling ecological and genetic novelty within lytic and lysogenic viral communities of hot spring phototrophic microbial mats, Microbiol. Spectr. 9 (3) (2021) e0069421, http://dx.doi.org/10.1128/Spectrum.00694-21.
[58] L. Xiao, W. Xiao, S. Lin, Q. Zhao, Ten genes are considered as potential biomarkers for the diagnosis of dermatomyositis, PLoS One 16 (11) (2021) e0260511, http://dx.doi.org/10.1371/journal.pone.0260511.
[59] Y. Qi, J. Xu, Z. Lin, Y. Tao, F. Zheng, Y. Wang, Y. Sun, S. Fu, W. Wang, C. Xie, Y. Zhang, F. Gong, The network of pro-inflammatory factors CD147, dcr3, and IL33 in the development of kawasaki disease, J. Inflamm. Res. Volume 14 (2021) 6043–6053, http://dx.doi.org/10.2147/JIR.S338763.
[60] Y. Zhang, T. Zhang, Q. Yin, H. Luo, Development and validation of genomic and epigenomic signatures associated with tumor immune microenvironment in hepatoblastoma, BMC Cancer 21 (1) (2021) http://dx.doi.org/10.1186/s12885-021-08893-3.
[61] Q. Qu, J.-Y. Sun, Z.-Y. Zhang, Y. Su, S.-S. Li, F. Li, R.-X. Wang, Hub microRNAs and genes in the development of atrial fibrillation identified by weighted gene co-expression network analysis, BMC Med. Genom. 14 (1) (2021) http://dx.doi.org/10.1186/s12920-021-01124-5.
[62] H. Li, M. Yang, C. Zhao, Y. Wang, R. Zhang, Physiological and proteomic analyses revealed the response mechanisms of two different drought-resistant maize varieties, BMC Plant Biol. 21 (1) (2021) http://dx.doi.org/10.1186/s12870-021-03295-w.
[63] A. Sukhwal, M. Bhattacharyya, S. Vishveshwara, Network approach for capturing ligand-induced subtle global changes in protein structures, Acta Crystallogr. D Biol. Crystallogr. 67 (5) (2011) 429–439, http://dx.doi.org/10.1107/S0907444911007062.
[64] E. Mutt, R. Sowdhamini, L. Menéndez-Arias, Molecular dynamics simulations and structural analysis to decipher functional impact of a twenty residue insert in the ternary complex of mus musculus TdT isoform, PLoS One 11 (6) (2016) e0157286, http://dx.doi.org/10.1371/journal.pone.0157286.
[65] D. He, Q. Li, G. Du, J. Sun, G. Meng, S. Chen, M.F. Nisar, Research on the mechanism of guizhi to treat nephrotic syndrome based on network pharmacology and molecular docking technology, BioMed Res. Int. 2021 (2021) 1–13, http://dx.doi.org/10.1155/2021/8141075.
[66] F. Gao, Z. Wang, J. Gu, X. Zhang, H. Wang, A hypoxia-associated prognostic gene signature risk model and prognosis predictors in gliomas, Front. Oncol. 11 (2021) http://dx.doi.org/10.3389/fonc.2021.726794.
[67] S. Yang, W. Zheng, C. Yang, R. Zu, S. Ran, H. Wu, M. Mu, S. Sun, N. Zhang, R.F. Thorne, Y. Guan, Integrated analysis of hub genes and micrornas in human placental tissues from in vitro fertilization-embryo transfer, Front. Endocrinol. 12 (2021) http://dx.doi.org/10.3389/fendo.2021.774997.
[68] S. Ahmad, S. Sharma, M.A. Afjal, H. Habib, J. Akhter, P. Goswami, S. Parvez, M. Akhtar, S. Raisuddin, mRNA expression and protein-protein interaction (PPI) network analysis of adrenal steroidogenesis in response to exposure to phthalates in rats, Environ. Toxicol. Pharmacol. 89 (2022) 103780, http://dx.doi.org/10.1016/j.etap.2021.103780.
[69] C. Bannert, J. Stoye (Eds.), Protein Annotation by Secondary Structure Based Alignments (PASSTA), Springer Berlin Heidelberg, 2005, http://dx.doi.org/10.1007/11560500_8.
[70] J.A. DiMasi, H.G. Grabowski, R.W. Hansen, Innovation in the pharmaceutical industry: New estimates of R&D costs, J. Health Econ. 47 (2016) 20–33, http://dx.doi.org/10.1016/j.jhealeco.2016.01.012.
[71] D. Vohora, G.e. Singh, Pharmaceutical medicine and translational clinical research, 2017.
[72] T. Hou, X. Xu, Recent development and application of virtual screening in drug discovery: an overview, Curr. Pharm. Des. (ISSN: 1381-6128) 10 (9) (2004) 1011–1033, http://dx.doi.org/10.2174/1381612043452721.
[73] W. Yu, A.D. MacKerell, Computer-aided drug design methods, Methods Mol. Biol. (Clifton, N.J.) 1520 (2017) 85–106, http://dx.doi.org/10.1007/978-1-4939-6634-9_5.
[74] W. Duch, K. Swaminathan, J. Meller, Artificial intelligence approaches for rational drug design and discovery, Curr. Pharm. Des. 13 (14) (2007) 1497–1508, http://dx.doi.org/10.2174/138161207780765954.
[75] S.-Y. Huang, X. Zou, Advances and challenges in protein-ligand docking, Int. J. Mol. Sci. 11 (8) (2010) 3016–3034, http://dx.doi.org/10.3390/ijms11083016.
[76] M.H. Baig, K. Ahmad, S. Roy, J.M. Ashraf, M. Adil, M.H. Siddiqui, S. Khan, M.A. Kamal, I. Provazník, I. Choi, Computer aided drug design: Success and limitations, Curr. Pharm. Des. 22 (5) (2016) 572–581, http://dx.doi.org/10.2174/1381612822666151125000550.
[77] A. Wlodawer, J. Vondrasek, Inhibitors of HIV-1 protease: a major success of structure-assisted drug design, Annu. Rev. Biophys. Biomol. Struct. 27 (1998) 249–284, http://dx.doi.org/10.1146/annurev.biophys.27.1.249.
[78] M. Batool, B. Ahmad, S. Choi, A structure-based drug discovery paradigm, Int. J. Mol. Sci. 20 (11) (2019) http://dx.doi.org/10.3390/ijms20112783.
[79] H. Marrakchi, G. Lanéelle, A.K. Quémard, Inha, a target of the antituberculous drug isoniazid, is involved in a mycobacterial fatty acid elongation system, FAS-II, Microbiology (Reading, England) 146 (Pt 2) (2000) 289–296, http://dx.doi.org/10.1099/00221287-146-2-289.
[80] Z. Miller, K.-S. Kim, D.-M. Lee, V. Kasam, S.E. Baek, K.H. Lee, Y.-Y. Zhang, L. Ao, K. Carmony, N.-R. Lee, S. Zhou, Q. Zhao, Y. Jang, H.-Y. Jeong, C.-G. Zhan, W. Lee, D.-E. Kim, K.B. Kim, Proteasome inhibitors with pyrazole scaffolds from structure-based virtual screening, J. Med. Chem. 58 (4) (2015) 2036–2041, http://dx.doi.org/10.1021/jm501344n.
[81] H.-J. Huang, H.W. Yu, C.-Y. Chen, C.-H. Hsu, H.-Y. Chen, K.-J. Lee, F.-J. Tsai, C.Y.-C. Chen, Current developments of computer-aided drug design, J. Taiwan Inst. Chem. Eng. 41 (6) (2010) 623–635, http://dx.doi.org/10.1016/j.jtice.2010.03.017.
[82] G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A.J. Olson, AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility, J. Comput. Chem. 30 (16) (2009) 2785–2791, http://dx.doi.org/10.1002/jcc.21256.
[83] O. Trott, A.J. Olson, AutoDock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, J. Comput. Chem. 31 (2) (2010) 455–461, http://dx.doi.org/10.1002/jcc.21334.
[84] B. Kramer, M. Rarey, T. Lengauer, Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking, Proteins 37 (2) (1999) 228–241, http://dx.doi.org/10.1002/(sici)1097-0134(19991101)37:2.
[85] R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, M.P. Repasky, E.H. Knoll, M. Shelley, J.K. Perry, D.E. Shaw, P. Francis, P.S. Shenkin, Glide: a new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy, J. Med. Chem. 47 (7) (2004) 1739–1749, http://dx.doi.org/10.1021/jm0306430.
[86] S.A. Khan, K. Zia, S. Ashraf, R. Uddin, Z. Ul-Haq, Identification of chymotrypsin-like protease inhibitors of SARS-CoV-2 via integrated computational approach, J. Biomol. Struct. Dyn. 39 (7) (2021) 2607–2616, http://dx.doi.org/10.1080/07391102.2020.1751298.
[87] K. Al-Khafaji, D. Al-Duhaidahawi, T. Taskin Tok, Using integrated computational approaches to identify safe and rapid treatment for SARS-CoV-2, J. Biomol. Struct. Dyn. 39 (9) (2021) 3387–3395, http://dx.doi.org/10.1080/07391102.2020.1764392.
[88] M.A.A. Ibrahim, A.H.M. Abdelrahman, K.S. Allemailem, A. Almatroudi, M.F. Moustafa, M.-E.F. Hegazy, In silico evaluation of prospective anti-COVID-19 drug candidates as potential SARS-CoV-2 main protease inhibitors, Protein J. 40 (3) (2021) 296–309, http://dx.doi.org/10.1007/s10930-020-09945-6.
[89] A. Petrou, P. Zagaliotis, N.F. Theodoroula, G.A. Mystridis, I.S. Vizirianakis, T.J. Walsh, A. Geronikaki, Thiazole/thiadiazole/benzothiazole based thiazolidin-4-one derivatives as potential inhibitors of main protease of SARS-CoV-2, Molecules (Basel, Switzerland) 27 (7) (2022) http://dx.doi.org/10.3390/molecules27072180.
[90] H. Monhemi, M.R. Housaindokht, The molecular mechanism of protein denaturation in supercritical co2: the role of exposed lysine residues is explored, The Journal of Supercritical Fluids (ISSN: 0896-8446) 147 (2019) 222–230, http://dx.doi.org/10.1016/j.supflu.2018.11.004, https://www.sciencedirect.com/science/article/pii/S0896844618304704.
[91] A. Gyulkhandanyan, A.R. Rezaie, L. Roumenina, N. Lagarde, V. Fremeaux-Bacchi, M.A. Miteva, B.O. Villoutreix, Analysis of protein missense alterations by combining sequence- and structure-based methods, Molecular Genetics & Genomic Medicine 8 (4) (2020) e1166, http://dx.doi.org/10.1002/mgg3.1166, https://onlinelibrary.wiley.com/doi/abs/10.1002/mgg3.1166.

[92] A. Sukhwal, M. Bhattacharyya, S. Vishveshwara, Network approach for capturing ligand-induced subtle global changes in protein structures, Acta crystallographica. Section D, Biological crystallography 67 (Pt 5) (2011) 429–439, http://dx.doi.org/10.1107/S0907444911007062.
[93] A. Ergün, C.A. Lawrence, M.A. Kohanski, T.A. Brennan, J.J. Collins, A network biology approach to prostate cancer, Mol. Syst. Biol. 3 (1) (2007) 82, http://dx.doi.org/10.1038/msb4100125.
[94] S. Gao, X. Wang, Predicting type 1 diabetes candidate genes using human protein-protein interaction networks, J. Comput. Sci. Syst. Biol. 2 (2009).