



# Capabilities of bioinformatics tools for optimizing physicochemical features of proteins used in Nano biosensors: A short overview of the tools related to bioinformatics

Behzad Foroutan<sup>a,b,\*</sup>, Amir Reza Abbasian Najafabadi<sup>c</sup>

<sup>a</sup> Tropical and Communicable Diseases Research Center, Iranshahr University of Medical Sciences, Iranshahr, Iran

<sup>b</sup> Department of Pharmacology, School of Medicine, Iranshahr University of Medical Sciences, Iranshahr, Iran

<sup>c</sup> Central Administration of Semnan University, Campus 1, Semnan, Iran

## ARTICLE INFO

### Keywords:

Bioinformatics tools  
Protein optimization technology  
Nano biosensors

## ABSTRACT

Protein-protein ligand is one of the most detection methods used in Nano biosensors. Based on the advantage of specific docking between two special 3D structures, they have become a potent candidate in bioanalysis and Nanodiagnostic tools. These tools lease users to do a simple, fast, cost-effective, sensitive, and specific detection of molecular biomarkers in real samples. Recent advantages of using protein-protein ligand Nano-biosensors application is remarkable due to its special docking that refers to each protein unique 3D conformation. However, it challenges different problems such as low rate of docking and hard process for fixation on the basic layer. These challenges make developers to optimize the structure and functions of proteins. The process has different Nano scale calculation that could be done with algorithms and solutions are available as bioinformatics tools. This article aimed to have a short overview of the abilities of bioinformatics tools for modeling and optimization of physicochemical features of proteins in Nano scale.

## 1. Introduction

Proteins are complex molecules found in all living organisms. Most proteins consist of linear polymers built from series of up to 20 different L- $\alpha$ -amino acids. Each cell has a unique type and different amount of each protein that makes identity and function of it. Proteins are identity keys for the ecotype of microorganisms and a substantial part to the identity of active metabolism in each living cell. All of the proteins in each cell named proteome which could provide a special characteristic of the cell. Each protein acts in different functions and plays a considerable role in the cell, such as cell signaling and the pathogenesis process of disease. Proteins are ideal materials for nanofabrication of rigid composition because of their unique 3D structure and specific answer to the existence of a specific cell. Determination of protein 3D conformation plays an important role in studying a particular disease live agent or proteins that use as a toxin. The molecular diagnostics based on the analyses of proteins docking, have offered a highly sensitive and quantitative method for the detection of infectious diseases and pathogens such as SARS-CoV-2. Detection of proteins based on docking method creates an ultrasensitive procedure that becomes a new area of research

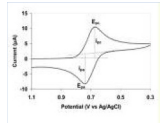
for developing Nano biosensors technology. (see [Table 1](#), [Fig. 1](#))

Recently, nanotechnology has created an integrated concept through biology, electronics and physics that branch a new field of science, the Nanomedicine. It is a combination of technology of diagnostic materials and devices, molecular imaging, drug delivery systems and regenerative medicine [2]. Remarkably, Nanomedicine enables *in vitro* and *in vivo* non-invasive diagnosis and targeted therapy by novel discoveries in sensing, processing and operating processes [3]. Currently, Imaging tools based on Nanotechnology have been medically applied as non-invasive methods of diagnosis [4–6]. The categories of Nano-diagnostic technologies, in addition to protein based Nano biosensors, include DNA-based Nano biosensors, Nano particle-based immunoassays, Nano scale visualization, Nano particulate biolabels, biobarcode assays, biochips, microarrays and combination of multiple diagnostics technologies [3]. Now, many methods to investigate protein-protein interactions are exist and each one has its own strengths and weaknesses, especially with regard to the sensitivity and specificity of its approach. A high sensitivity means that many of the interactions that occur in reality detected by the screen while a high specificity indicates most of the interactions detected by the screen are occurring in reality.

\* Corresponding author. Tropical and Communicable Diseases Research Center, Iranshahr University of Medical Sciences, Iranshahr, Iran.

E-mail address: [behzad\\_foroutan@hotmail.com](mailto:behzad_foroutan@hotmail.com) (B. Foroutan).

**Table 1**  
Components of a Nano biosensor (Modified from Ref. [1]).

Analyte	Receptor	Transducer	Signal output
Enzyme DNA/ RNA	Enzyme Antibody	Electrochemical Optical	
Antigen Protein	Nucleic acid Cell	Thermal Piezoelectric	

Protein-based Nano biosensors have been increasingly used in medical diagnostic for continuous monitoring of human health against pathogens and protein-based toxins [7–9] and to their applications in the field of food analysis [10], bioterrorism [11], and environment [11–13]. Proteins as biological molecules act directly in metabolism processes which are one of the unique identities of each cell that could be defined by their unique reaction to their substrate. One specific method for detection of a protein as a toxin or live agent identity is their reaction to a substrate or antibody, this reaction calls docking and creates change in weight, conformation, and physicochemical features that could be measured by different technologies from Nano biosensors. Using extracted natural proteins from organisms, ending with change/s in conformation and environmental elements decreasing in rate of docking and unstable fixing due to protein structure, are the causes of problems. These problems make developers to use synthetic optimized proteins that should be designed and optimized in details by bioinformatical tools available online. This short overview engages on the ability of bioinformatics tools for optimization and calculation of physicochemical features of proteins which might be a candidate for Nano biosensors.

## 2. 2D structure prediction

Bioinformatics tools support to predict secondary structures of proteins based on the sequencing of amino acid chain. These tools used mathematical algorithms to predict available 2D structure like Alpha-helix and Beta-sheet. Subsequently, an algorithm optimized model according to the data base of structures' studied would be made. The PSIPRED is a protein analysis workbench that provides many available analysis tools into a single web based framework [14–16]. It is a comprehensive tool for prediction of secondary structures with access to GenTHREADER for protein folds' recognition and MEMSAT-2 trans-membrane topology prediction. Other useful tool is COILS [17,18] that provide a service in prediction of coiled coil region. COILS is a program that compares a sequence to a database of known parallel two-stranded coiled-coils and derives a similarity score. By comparing this score to the distribution of scores in globular and coiled-coil proteins, calculates the probability that the sequence will adopt a coiled-coil conformation [17].

## 3. 3D structural features

These tools include services for searching structural motifs, biochemical features found in protein structures and functional sub structures such as binding sites. Computed Atlas of Surface Topography of proteins (CASTp) is one of top index used software [19]. It is a web server that provides online services for locating, delineating and measuring these geometric and topological properties of protein structures. It has been widely used since its inception in 2003 for locates and measures concave surface regions on 3D protein structures. This tool should be used to study surface features, binding sites and functional

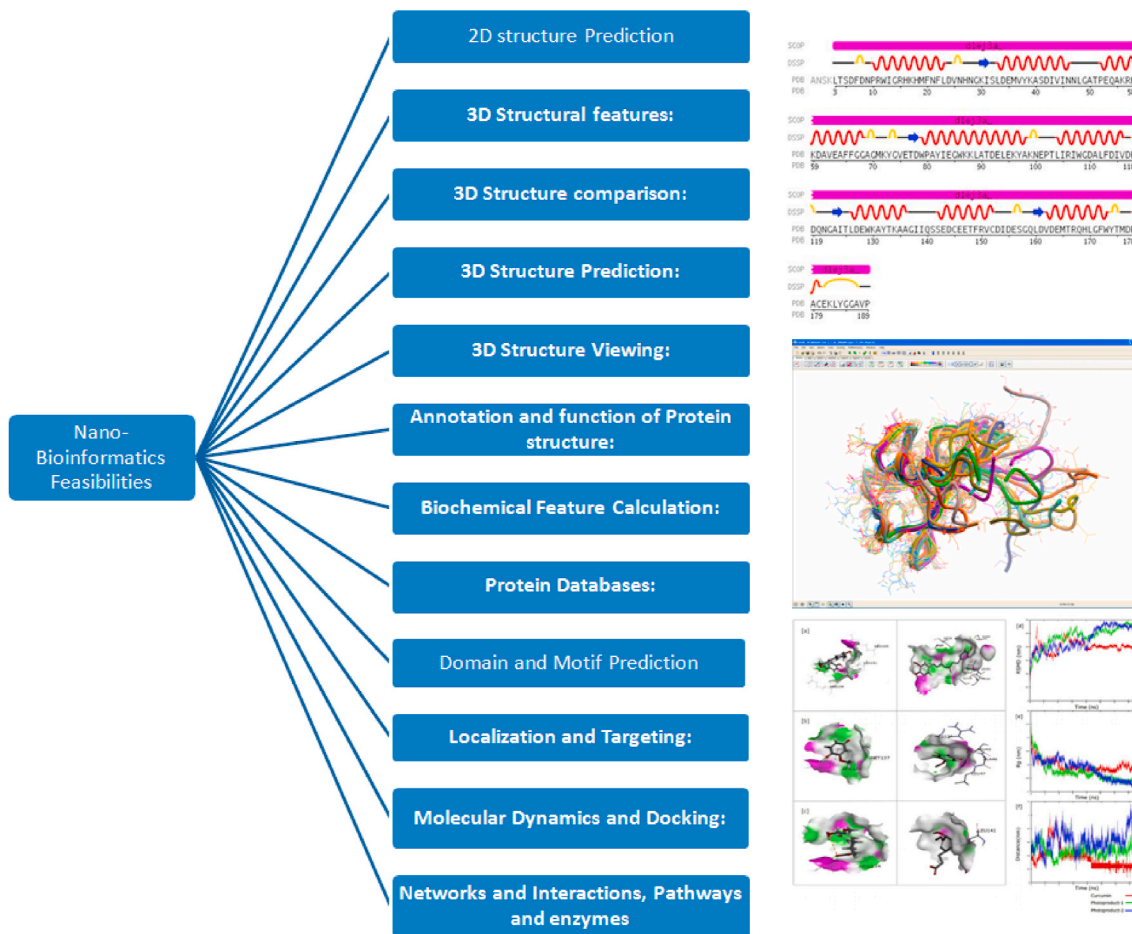


Fig. 1. Schematic illustration of the strategies explained in the text.

regions of proteins. Another tool in this category is an online service from Andrej Sali's Lab. It contains various resources for comparative protein structure modeling and analysis from the Sali's Lab at the University of California at San Francisco [20].

#### 4. 3D structure comparison

This category contains resources for the comparison of sequences at the level of tertiary structures. This includes tools for superimposing structures and structural alignments. One of the best services in this field is SCOP [21]. It provides a structural classification of proteins. Database created by a combination of automated methods and manual inspection and contains a comprehensive ordering of all proteins of known structure, according to evolutionary and structural relationships [22,23]. One more tool which might be used for comparison of 3D structure is ProCKSI [22]. It is a multi-layer protein comparison meta-server that computes structure similarities using various information theory measures. ProCKSI integrates various protein similarity measures through an easy to use interface that allows the comparison of multiple proteins simultaneously. Based on a diverse set of similarity measures, ProCKSI computes a consensus similarity profile for the entire protein set. All results can be clustered, visualized and analyzed for users [24].

#### 5. 3D structure prediction

This category aid in protein 3D structure prediction based on the multiple strategy that each algorithm used to report a prediction. The top index tools used in this area named Swiss Model [25,26]. It is a fully automated protein structure homology-modelling server, accessible via the ExPASy web server, or from the program DeepView (Swiss Pdb-Viewer) [26]. The purpose of this server is to make different algorithms for protein modeling. The other service used as a top index interface is MODELLER [27,28]. It is used for homology or comparative modeling of protein 3D structures [29–31]. The user provides an alignment of a sequence to be modeled with known related structures and MODELLER automatically calculates a model containing all non-hydrogen atoms. MODELLER implements comparative protein structure modeling by satisfaction of spatial restraints [32,33].

#### 6. 3D structure viewing

One of the major challenges in bioinformatics is to show all calculation, data, and features in a simple real model. Bioinformatics based on the Nano scale modelers and 3D structure calculators provide a service for 3D structure viewing which is using for visualizing of 3D structures. Swiss-Pdb viewer is an Excellent tool for annotating, comparing, coloring, and mutating 3D structures [34]. Other service that provides a different method of 3D structure viewing is MolProbity [35]. It is a structure validation web service for diagnosing problems in 3D models of proteins, nucleic acids or complexes. It adds and optimizes H atoms (correcting 180° flipped Asn/Gln/His side chains) and then calculates global and local validation for all-atom contacts (steric clashes, H-bonds, and vdW), covalent geometry and conformation (Ramachandran and rotamers for protein, ribose puckers and suite conformers for RNA). Results are displayed online as 3D graphics and sortable charts [36,37]. The web-native Mol\* Viewer assists 3D visualization and streaming of macromolecular coordinate and experimental data, together with capabilities for displaying structure quality, functional, or biological context annotations. High-performance graphics and data management allows users to concurrently visualize up to hundreds of (superimposed) protein structures, stream molecular dynamics simulation trajectories, render cell-level models, or display huge I/HM structures [38].

#### 7. Annotation and function of protein structure

One of the major data requires in Nanoscale modeling and

optimization is annotating and function prediction, localization and classification of proteins before doing more experiments. Computational calculation let us to have a satisfied area of calculated data before experiments. The top index tool is SIFT which is a sequence homology-based tool that will predict whether an amino acid substitution will affect protein function [39]. The other tool is InterProScan [40]. It allows users to query using different protein signature recognition methods to look up InterPro annotations for their sequences [41]. These annotations result often include gene ontology terms that let user to associate with their sequence.

#### 8. Biochemical features calculation

This category of bioinformatical software includes a protein identification tools which can give users information on the chemical structures and amino acid properties of peptide sequences. MASCOT [42] is a high index online tool that provides a service about protein identification by peptide mass; excellent documentation; incorporates code from MOWSE but allows more search methods on more sequence databases. Other reliable online tool available open access is EMBOSS [43,44] which provides multiple independent tools. These tools provide different calculation methods for analyzing and reporting biochemical features.

#### 9. Protein databases

Protein databases contain databases of protein sequences, properties, targeting, motifs, domains, structures, and protein families. Top index data bases which is available online and open access is Pfam [45]. Pfam is a database of protein families and domains that uses HMMER3, the latest version of the popular profile hidden Markov model package. Pfam release 24.0 contains 11,912 families. SCOP [46] is another top reference database that provides structural classification of proteins. Database created by a combination of manual inspection and automated methods. Comprehensive ordering of all proteins of known structure according to evolutionary and structural relationships provided.

#### 10. Domain and motif prediction

An important area in Nano scale optimization in proteins is domain and motif prediction which is based on primary data available. It can give information about protein domains and/or predict motifs, domains, and patterns in peptide sequences. Top index online tool used for this aim is Berkeley Phylogenomics Group [47] which provides a series of web servers for phylogenomic analysis: classification of sequences to pre-computed families and subfamilies using the PhyloFacts Phylogenomic Encyclopedia [48], FlowerPower clustering of proteins sharing the same domain architecture [49], MUSCLE multiple sequence alignment [50,51], SATCHMO simultaneous alignment and tree construction [52] and SCI-PHY subfamily identification [53]. Other useful service is COGs [54,55] that provides clusters of orthologous groups represent ancient conserved protein domains; use COGnitor tool to find COGS in sequence of interest [56].

#### 11. Localization and targeting

One of the major challenges in developing optimized Nanoscale modeling is localization and targeting of the protein before and after optimization. These tools related to predicting sub-cellular localization, the presence of *trans*-membrane regions, and/or targeting including the prediction of signal peptides. Top index algorithms with user-friendly interface is PSIPRED [14] which provides a protein analysis workbench unites available analysis tools into a single web based framework. An excellent tool for prediction of secondary structure, with access to GenTHREADER [57] for protein folds recognition and MEMSAT-2 transmembrane topology prediction. Other specialized tool with

specific use in localization is [PSORT.ORG](#) [58] that provides links to the PSORT family of web-based programs for sub-cellular localization prediction, including PSORTb [59] and WoLF PSORT [60], as well as other datasets and resources relevant to localization prediction.

## 12. Molecular dynamics and docking

The most important aim of protein optimization in Nano biosensors is process of docking between two proteins. Calculation of molecular dynamics and docking need a Nano scale view and data about structure and atomic view of each atom reaction in bonding to other proteins. This category provides resources for molecular dynamics including tools that can predict the movements of structures and/or conformational changes. One of top index online tools available is oGNMs [61,62] that calculates the equilibrium dynamics of any structure submitted in Protein Data Bank (PDB) format, using the Gaussian Network Model (GNM). Other useful service is ClusPro [63] which is a tool for automatically computing the docking of two protein structures supplied by the user (or as PDB IDs). The result set is a ranked list of putative complexes, ordered by clustering properties.

## 13. Networks and interactions, pathways and enzymes

Complexity of life is in atomic scale and complexity of reactions that create a network of interactions. All of these interactions create pathways that controlled and run by series of enzymes. Knowing details about these may help to detect a special protein that is simple and easy for detection by Nano biosensors mechanisms. This area of bioinformatics includes tools and resources for enzymes, metabolic and proteomic pathways and networks. Many of these resources contain dynamic pathway diagrams and protein-protein interactions. Top index online tool is Kyoto Encyclopedia of Genes and Genomes (KEGG) [64–66] that has pathway maps, molecular catalogs, genome maps, and gene catalogs that capture knowledge about interactions in terms of information pathways. KEGG comprises several databases, including BRITe (protein-protein interactions) [67], PATHWAY (interaction networks for cellular processes) [68], and LIGAND (chemical compounds and chemical reactions) [69,70]. KEGG Atlas is a new tool for the global analysis of metabolic pathways. Cytoscape [71] is another online and offline tool that is a visualization platform for use with molecular interaction networks. Interaction data could be integrated with other state data such as gene expression profiles. The comments to Cytoscape includes lists of interaction pairs, and tab/space delimited files containing mRNA expression profiles [72]. The nodes of the interaction networks could be filtered by such variables as GO annotations and number of interactions [73].

## 14. Conclusion

New research and development in Nano biosensors based on protein detection methods, needs a large-scale data analysis and structural optimizations that aimed for better docking and more stability in detection. All biochemical and biophysical factors affect atomic behaviors in the docking of protein-protein interactions. These atomic, dynamic, and static behaviors would be calculated by different algorithms. The calculations and modeling could be categorized in a branch of fields that all become the science of bioinformatics. Absolutely from one primary data, branch of result based on calculations with different algorithms extracted, so it seems that it's better to precede optimization with different tools which use the previous database for optimizing protein conformation based on the previous understandings and experiences. Bioinformatics tools help us to have an atomic view in modeling and understanding biological molecule behavior *in vivo* and *in vitro*. Thus in association with Nano sciences, bioinformatics tools could develop a highly specific technology that can detect toxic proteins or be used as a diagnostic device for detecting bioterrorism agents or other proteins. In

addition, it could able us to analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs.

## Additional information

This overview is not a part of MSc thesis.

## References

- [1] J. Kim, et al., Wearable biosensors for healthcare monitoring, *Nat. Biotechnol.* 37 (4) (2019) 389–406.
- [2] R. Duncan, R. Gaspar, Nanomedicine(s) under the microscope, *Mol. Pharm.* 8 (6) (2011) 2101–2141.
- [3] K.M. Abu-Salah, et al., DNA-based nanobiosensors as an emerging platform for detection of disease, *Sensors* 15 (6) (2015) 14539–14568.
- [4] B.D. Malhotra, A. Chaubey, S.P. Singh, Prospects of conducting polymers in biosensors, *Anal. Chim. Acta* 578 (1) (2006) 59–74.
- [5] B.D. Malhotra, et al., Nanomaterial-based biosensors for food toxin detection, *Appl. Biochem. Biotechnol.* 174 (3) (2014) 880–896.
- [6] S. Choi, et al., Recent advances in flexible and stretchable bio-electronic devices integrated with nanomaterials, *Adv. Mater.* 28 (22) (2016) 4203–4218.
- [7] R. El-Sayed, et al., Thermostable luciferase from *Luciola cruciate* for imaging of carbon nanotubes and carbon nanotubes carrying doxorubicin using in vivo imaging system, *Nano Lett.* 13 (4) (2013) 1393–1398.
- [8] K.M. Abu-Salah, A.A. Ansari, S.A. Alrokayan, DNA-based applications in nanobiotechnology, *J. Biomed. Biotechnol.* 2010 (2010) 715295.
- [9] K.M. Abu-Salah, et al., Nanomaterials as analytical tools for genosensors, *Sensors* 10 (1) (2010) 963–993.
- [10] Y. Uludag, et al., Lab-on-a-chip based biosensor for the real-time detection of aflatoxin, *Talanta* 160 (2016) 381–388.
- [11] F.M. Burkler Jr., Measures of effectiveness in large-scale bioterrorism events, *Prehospital Disaster Med.* 18 (3) (2003) 258–262.
- [12] F. Mouffouk, et al., Development of a highly sensitive bacteria detection assay using fluorescent pH-responsive polymeric micelles, *Biosens. Bioelectron.* 26 (8) (2011) 3517–3523.
- [13] S. Kroger, S. Piletsky, A.P. Turner, Biosensors for marine pollution research, monitoring and control, *Mar. Pollut. Bull.* 45 (1–12) (2002) 24–34. <http://bioinf.cs.ucl.ac.uk/psipred/>.
- [14] T. Kosciolk, D.T. Jones, De novo structure prediction of globular proteins aided by sequence variation-derived contacts, *PLoS One* 9 (3) (2014) e92197.
- [15] T. Nugent, D. Cozzetto, D.T. Jones, Evaluation of predictions in the CASP10 model refinement category, *Proteins* (82 Suppl) (2014) 98–111, 2.
- [16] A. Lupas, M. Van Dyke, J. Stock, Predicting coiled coils from protein sequences, *Science* 252 (5009) (1991) 1162–1164.
- [17] [https://embnet.vital-it.ch/software/COILS\\_form.html](https://embnet.vital-it.ch/software/COILS_form.html).
- [18] W. Tian, et al., CASTp 3.0: computed atlas of surface topography of proteins, *Nucleic Acids Res.* 46 (W1) (2018) W363–W367. [http://salilab.org/our\\_resources.shtml](http://salilab.org/our_resources.shtml).
- [19] A. Andreeva, et al., The SCOP database in 2020: expanded classification of representative family and superfamily domains of known protein structures, *Nucleic Acids Res.* 48 (D1) (2020) D376–D382.
- [20] A.G. Murzin, et al., SCOP: a structural classification of proteins database for the investigation of sequences and structures, *J. Mol. Biol.* 247 (4) (1995) 536–540. <http://cl.sdsc.edu/ce.html>.
- [21] D. Barthel, et al., ProCKSI: a decision support system for protein (structure) comparison, knowledge, similarity and information, *BMC Bioinf.* 8 (2007) 416. <http://www.expasy.ch/swissmod/SWISS-MODEL.html>.
- [22] A. Waterhouse, et al., SWISS-MODEL: homology modelling of protein structures and complexes, *Nucleic Acids Res.* 46 (W1) (2018) W296–W303. <http://salilab.org/modeller/>.
- [23] A. Fiser, A. Sali, Modeller: generation and refinement of homology-based protein structure models, *Methods Enzymol.* 374 (2003) 461–491.
- [24] B. Webb, A. Sali, Comparative protein structure modeling using MODELLER, *Curr. Protoc. Bioinf.* 47 (2014), 5 6 1–32.
- [25] U. Pieper, et al., ModBase, a database of annotated comparative protein structure models and associated resources, *Nucleic Acids Res.* (2014) 42. Database issue: p. D336–46.
- [26] M.A. Marti-Renom, et al., Comparative protein structure modeling of genes and genomes, *Annu. Rev. Biophys. Biomol. Struct.* 29 (2000) 291–325.
- [27] A. Sali, T.L. Blundell, Comparative protein modelling by satisfaction of spatial restraints, *J. Mol. Biol.* 234 (3) (1993) 779–815.
- [28] A. Fiser, R.K. Do, A. Sali, Modeling of loops in protein structures, *Protein Sci.* 9 (9) (2000) 1753–1773. <http://spdbv.vital-it.ch/>.
- [29] <http://molprobity.biochem.duke.edu/>.
- [30] V.B. Chen, et al., MolProbity: all-atom structure validation for macromolecular crystallography, *Acta Crystallogr.* 66 (Pt 1) (2010) 12–21.
- [31] I.W. Davis, et al., MolProbity: all-atom contacts and structure validation for proteins and nucleic acids, *Nucleic Acids Res.* (2007) 35. Web Server issue: p. W375–83.
- [32] D. Sehnal, et al., Mol\* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures, *Nucleic Acids Res.* 49 (W1) (2021) W431–W437.

- [39] N.L. Sim, et al., SIFT web server: predicting effects of amino acid substitutions on proteins, *Nucleic Acids Res.* (2012) 40. Web Server issue): p. W452-7.
- [40] P. Jones, et al., InterProScan 5: genome-scale protein function classification, *Bioinformatics* 30 (9) (2014) 1236–1240.
- [41] M. Blum, et al., The InterPro protein families and domains database: 20 years on, *Nucleic Acids Res.* 49 (D1) (2021) D344–D354.
- [42] M. Hirose, et al., MASCO: multiple alignment system for protein sequences based on three-way dynamic programming, *Comput. Appl. Biosci.* 9 (2) (1993) 161–167.
- [43] <http://emboss.bioinformatics.nl/>.
- [44] P. Rice, I. Longden, A. Bleasby, EMBOSS: the European molecular biology open software suite, *Trends Genet.* 16 (6) (2000) 276–277.
- [45] S. El-Gebali, et al., The Pfam protein families database in 2019, *Nucleic Acids Res.* 47 (D1) (2019) D427–D432.
- [46] <http://scop.mrc-lmb.cam.ac.uk/scop/>.
- [47] J.G. Glanville, et al., Berkeley Phylogenomics Group web servers: resources for structural phylogenomic analysis, *Nucleic Acids Res.* (2007) 35. Web Server issue): p. W27-32.
- [48] N. Krishnamurthy, et al., PhyloFacts: an online structural phylogenomic encyclopedia for protein functional and structural classification, *Genome Biol.* 7 (9) (2006) R83.
- [49] N. Krishnamurthy, D. Brown, K. Sjolander, FlowerPower: clustering proteins into domain architecture classes for phylogenomic inference of protein function, *BMC Evol. Biol.* (7 Suppl 1) (2007) S12.
- [50] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Res.* 32 (5) (2004) 1792–1797.
- [51] R.C. Edgar, MUSCLE: a multiple sequence alignment method with reduced time and space complexity, *BMC Bioinf.* 5 (2004) 113.
- [52] R. Hagopian, et al., SATCHMO-JS: a webserver for simultaneous protein multiple sequence alignment and phylogenetic tree construction, *Nucleic Acids Res.* (2010) 38. Web Server issue): p. W29-34.
- [53] D.P. Brown, N. Krishnamurthy, K. Sjolander, Automated protein subfamily identification and classification, *PLoS Comput. Biol.* 3 (8) (2007) e160.
- [54] <http://www.ncbi.nlm.nih.gov/COG/>.
- [55] M.Y. Galperin, et al., COG database update: focus on microbial diversity, model organisms, and widespread pathogens, *Nucleic Acids Res.* 49 (D1) (2021) D274–D281.
- [56] R.L. Tatusov, et al., The COG database: a tool for genome-scale analysis of protein functions and evolution, *Nucleic Acids Res.* 28 (1) (2000) 33–36.
- [57] L.J. McGuffin, et al., The Genomic Threading Database: a comprehensive resource for structural annotations of the genomes from key organisms, *Nucleic Acids Res.* (2004) 32. Database issue): p. D196-9.
- [58] <http://www.psорт.org/>.
- [59] J.L. Gardy, et al., PSORT-B: improving protein subcellular localization prediction for Gram-negative bacteria, *Nucleic Acids Res.* 31 (13) (2003) 3613–3617.
- [60] P. Horton, et al., WoLF PSORT: protein localization predictor, *Nucleic Acids Res.* (2007) 35. Web Server issue): p. W585-7.
- [61] <http://gnm.csb.pitt.edu/oGNM.php>.
- [62] L.W. Yang, et al., oGNM: online computation of structural dynamics using the Gaussian Network Model, *Nucleic Acids Res.* (2006) 34. Web Server issue): p. W24-31.
- [63] D. Kozakov, et al., The ClusPro web server for protein-protein docking, *Nat. Protoc.* 12 (2) (2017) 255–278.
- [64] <http://www.genome.ad.jp/kegg>.
- [65] M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes, *Nucleic Acids Res.* 28 (1) (2000) 27–30.
- [66] M. Kanehisa, The KEGG database, *Novartis Found. Symp.* 247 (2002) 91–101, discussion 101-3, 119-28, 244-52.
- [67] M. Kanehisa, Y. Sato, KEGG Mapper for inferring cellular functions from protein sequences, *Protein Sci.* 29 (1) (2020) 28–35.
- [68] M. Kanehisa, et al., KEGG as a reference resource for gene and protein annotation, *Nucleic Acids Res.* (D1) (2016) 44, p. D457-62.
- [69] M. Tanabe, M. Kanehisa, Using the KEGG database resource, *Curr. Protoc. Bioinf.* (2012) 1.12.1–1.12.43, <https://doi.org/10.1002/0471250953.bi0112s38>. Supplement 38, (Chapter 1): p. Unit1 12, <https://pubmed.ncbi.nlm.nih.gov/22700311/>.
- [70] M. Kanehisa, et al., KEGG for integration and interpretation of large-scale molecular data sets, *Nucleic Acids Res.* (2012) 40. Database issue): p. D109-14.
- [71] D. Otasek, et al., Cytoscape Automation: empowering workflow-based network analysis, *Genome Biol.* 20 (1) (2019) 185.
- [72] <http://www.cytoscape.org>.
- [73] F. Supek, N. Skunca, Visualizing GO annotations, *Methods Mol. Biol.* 1446 (2017) 207–220.