iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** VIII **Month of publication:** Aug 2023

DOI: https://doi.org/10.22214/ijraset.2023.55519

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VIII Aug 2023- Available at www.ijraset.com

Protein Sequencing and Multi-Omics Data Analysis of Human Insulin

Uma Kumari[1], Aditi Bhatia[2]
[1]*Senior Bioinformatics Scientist,* [2]*Project Trainee, Bioinformatics Project and Research Institute, Noida-201301, India*

Abstract: ***Advancements in molecular biology have led to a deeper understanding of complex biological processes through the analysis of multi-omics data. In this study, we focused on the protein sequencing and multi-omics data analysis of human insulin, a pivotal hormone with critical roles in regulating glucose metabolism. In this study, we delved into the intricate realm of human insulin, a critical hormone regulating glucose metabolism, using advanced molecular biology techniques. Leveraging the Protein Data Bank (PDB) and the WorldWide Protein Data Bank (wwPDB), we explored 3D protein structures through sources like X-ray Crystallography and NMR spectrometry. RasMol software aided dynamic visualization of molecular structures, allowing manipulation and annotation. The Molecular Modelling Database (MMDB) facilitated structure identification and cross-species sequence alignment. The National Centre for Biotechnology Information (NCBI) provided genetic and biomedical insights, enhancing our understanding of insulin's molecular intricacies. Computational tools, including the FASTA format, accelerated sequence analysis, while BLAST unveiled evolutionary relationships and functional conservation. COBALT enabled incremental protein sequence alignment. Python-based PyMOL enriched analyses with specialized plugins, promoting efficient complex research.***

Keywords: Insulin; PyMol; Sequence alignment, Cobalt,structural analysis,Biological database

I. INTRODUCTION

Human insulin, a peptide hormone weighing approximately 5808 Da, is synthesized within the beta cells of the pancreas islets of Langerhans. It plays a vital role in governing glucose metabolism. Studies have indicated that a minor proportion of insulin might also be produced in specific neurons within the central nervous system. Following a meal, there is a surge in blood glucose levels, triggering the secretion of insulin in response. As insulin levels decrease, the liver releases glucose into the blood. Before the discovery of insulin, individuals with diabetes faced a lot of difficulties, as medical options were limited. Doctors had little means to effectively manage the condition, and the primary treatment involved placing patients on extremely restrictive diets with minimal carbohydrate intake. Although such measures could provide a few additional years, they were unable to ensure long-term survival for those with diabetes. In 1889, a pivotal discovery was made by Oskar Minkowski and Joseph von Mering, German researchers, who observed that removing the pancreas gland from dogs resulted in the animals having diabetes symptoms and eventually getting to the condition. This finding led to the hypothesis that the pancreas played a crucial role in producing "pancreatic substances" (later identified as insulin). In 1910, Sir Edward Albert Sharpey-Shafer put forth a compelling hypothesis that a single chemical was lacking in individuals with diabetes. He named this vital chemical "insulin," deriving the term from the Latin word "insula," which translates to "island." This reference was inspired by the pancreas' unique clusters of cells, known as islets of Langerhans, where insulin is produced and secreted. In the year 1921, the discovery of insulin was officially reported by Canadian scientists Frederick G. Banting and Charles H. Best. They identified and extracted insulin from the pancreas of a dog, marking a monumental achievement in diabetes research. Interestingly, during the same period, a Romanian physiologist named Nicolae C. Paulescu was also independently investigating a substance containing insulin, which he referred to as the "pancrein." Thus, these parallel efforts from different corners of the world contributed significantly to the understanding and potential treatment of diabetes through the use of insulin. Their remarkable progress continued as they managed to keep another dog, severely afflicted with diabetes, alive for another 70 days, and it was only when the insulin extract ran out that the dog finally died. Encouraged by this success, the researchers, in collaboration with their colleagues, J.B. Collip and John Macleod, embarked on an even more ambitious endeavor. They successfully developed a refined and pure form of insulin, this time sourced from the pancreas of cattle. In January 1922, a pivotal moment occurred when Leonard Thompson, a 14-year-old boy suffering from diabetes in a Toronto hospital, became the first person to receive an insulin injection. Within just 24 hours of observation, Leonard's dangerously high blood glucose levels dropped to nearly normal levels.

"Human insulin" refers to a synthetic form of insulin that is artificially produced in laboratories to closely mimic the insulin naturally found in humans. This revolutionary development received pharmaceutical approval in 1982, marking a significant milestone in diabetes treatment. Before the advent of human insulin, individuals relied on animal insulin, typically sourced from purified porcine (pork) insulin, for diabetes management. The introduction of human insulin brought about a major advancement in diabetes care, offering a more precise and effective alternative for patients. Human insulin is synthesized in laboratories by cultivating insulin proteins within Escherichia coli (E. coli) bacteria**[1,2]**.

Human insulin is of two forms, a short-acting (regular) form, and an intermediate-acting (NPH) form. NPH is Neutral Protamine Hagedorn insulin, also called isophane insulin. Regular human insulin exhibits certain undesirable characteristics, including a delayed onset of action and variable peak and duration of action when administered. Consequently, an increasing number of medical practitioners are opting for alternative insulin options, and the prescription of Regular insulin is gradually declining. The delayed onset of action necessitates injecting insulin and waiting before meals. Moreover, the unpredictable duration of action increases the risk of hypoglycemia long after the meal has concluded. NPH (Neutral Protamine Hagedorn) is an intermediate-acting form of human insulin utilized to manage blood sugar levels between meals and fulfill overnight insulin needs. Some examples of human insulin are- regular (short-acting): Humulin S, Act rapid, Insuman Rapid, NPH (intermediate-acting): Humulin I, Insuman basal, Insulatard, Premixed human insulins: Humulin M2, M3, and M5, Insuman Comb 15, 25 and 50. Premixed insulins are formulations that combine regular and NPH insulin. These premixed insulins are offered in various mixing ratios to cater to different patient needs. For instance, Humulin M3 contains a blend of 30% short-acting insulin and 70% intermediate-acting insulin, while Humulin M5 consists of an equal combination of 50% short-acting and 50% intermediate-acting insulin. Short-acting (regular) insulin works approximately 30 minutes after injection and starts affecting between 2 to 3 hours post-injection. The duration of its activity spans up to 10 hours. On the other hand, intermediate-acting (NPH) insulin takes around 2 to 4 hours to initiate its action, peaking between 4 and 10 hours after injection. Its duration of activity extends up to 18 hours. There may be some side effects of human insulin such as hypo unawareness, tiredness, and weight increase. These side effects may not be found while taking animal insulin**[3, 4]**.

Type 1 diabetes is an autoimmune disease that destroys the beta-cells in the pancreas that produces insulin. So this could prevent the body from producing enough insulin to regulate blood glucose levels. It is also called juvenile diabetes or insulin-dependent diabetes. The symptoms expressed in the individuals might be above-average thirst, tiredness during the day, or unexplained weight loss. Type 2 diabetes mellitus is a metabolic disorder that may eventually result in hyperglycemia (high blood glucose level). It is mostly caused due to the body being unable to metabolite glucose. The high level of blood glucose can damage the organs of the body. Its symptoms are increasing hunger, frequent urination, etc**[5, 6]**.

II. MATERIALS AND METHODS

PDB is Protein Data Bank. The Protein Data Bank has functioned solely and served data on the 3-D protein structures, nucleic acids and intricate molecular assemblies. The WorldWide Protein Data Bank (wwPDB) organization oversees the management of the PDB archive and makes sure that it is easily available globally. It comprises data collected from various sources, including X-ray Crystallography and nuclear magnetic resonance (NMR) spectrometry. RasMol is a software tool designed in such a way that can view PDB files and visualization of graphics. It can be used to rotate the molecules, change mold type, color, etc[7,8]. It has two windows:

1) Displays the image of the molecule
2) Command line window (commands can be given)

```
PDB: 1B9E_D
>pdb|1B9E|D Chain D, PROTEIN (INSULIN)
FVNQHLCGEHLVEALYLVCGERGFFYTPKT
```

The Molecular Modelling Database (MMDB) is a database that identifies 3D structures. The analysis of 3D protein structures frequently offers insights into the sequence seen in protein. Additionally, aligning 3D structures results into multiple sequences. MMDB is updated once a week**[9, 10]**. NCBI is the National Centre for Biotechnology Information, known as NCBI. It gives access to all the users to access genetic and biomedical information. The research conducted by the institution focuses on biomedical issues at the molecular level with computational methods.

In bioinformatics, the FASTA format serves as a text-based method for representing nucleotide or amino acid sequences. It employs single letter codes to denote nucleotide or amino acids. It was first described by Lipman and Pearson in 1985. It is a DNA and protein sequence alignment software and is a fast homology search tool**[11, 12]**. It is quite similar to BLAST but in this tool it speeds up the sequence. Hence, called FASTA (FAST- fast; A- alignment). BLAST is Basic Local Alignment Search Tool. A sequence search program that is accessible through a web interface or can function as an independent tool. Multiple types of BLAST are available to facilitate to comparisons between all possible combinations of nucleotide or protein databases. BLAST offers statistical data to aid in interpreting the biological significance of the alignment**[13, 14, 15]**. COBALT is a multiple sequence alignment tool. It identifies set of pairs that are extracted from the database. It performs incremental multiple alignment sequences of protein**[16]**. PyMOL is a python based software and has the advantage of embedding many useful python based plugins to enhance its functionalities for research work like drug designing and further calculations**[17, 18]**. Molecular visualization is a crucial aspect of structural biology, enabling researchers to comprehend the complex arrangement of atoms and molecules that underlie biological processes. RasMol, first released in 1992 by Roger Sayle, emerged as one of the pioneering tools in this arena. It empowered scientists to visualize and manipulate molecular structures obtained from various experimental techniques like X-ray crystallography, NMR spectroscopy, and cryo-electron microscopy **[19]**.

III. RESULT AND DISCUSSION

A. Structural Analysis

Human insulin is an important hormone and it helps in regulation of glucose metabolism. It comprises of two peptide chains, A and B. The aim of this analysis was to gain insights into its bond (disulfide) interactions and the arrangement of the chains. As already known, there are two windows in RasMolsoftware, one is for displaying the structure and the other for giving commands.

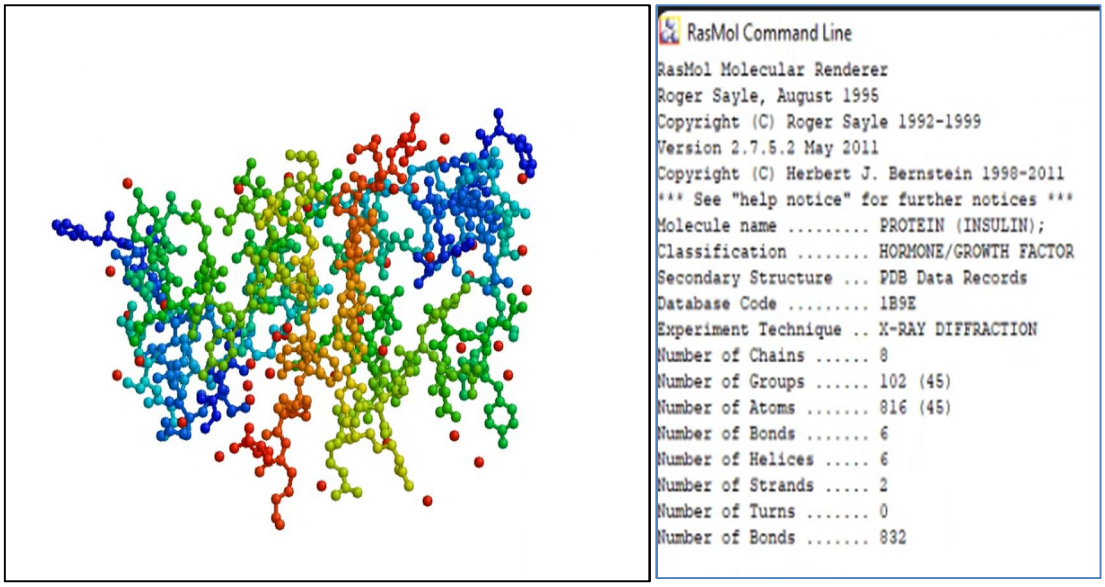


Figure 1: Structure analysis in ball and stick model (defined residues)

Table1: Result obtained from RasMol

Atom	Number of atoms observed	Color shown by atoms
Carbon	518	Gray
Sulfur	12	Yellow
Nitrogen	130	Blue
Chlorine	No atoms present	Green
Oxygen	201	Red

By giving commands in the command line window, user can explore how many atoms are comprised in the structure. The above data (output) has been found.

For structure analysis, the 3D structure can be displayed according to the user and can similarly change the color.

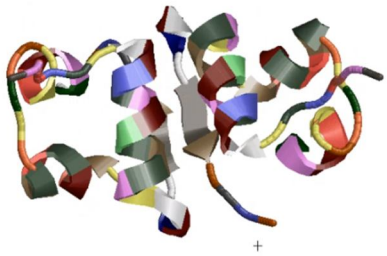


Figure 2: RasMol result observed in shapely form (protein structure)

Red color represents glutamic amino acid, Light blue color represents adenine amino acid, Dark blue color represents arginine amino acid, Green color represents alanine amino acid, Dark gray represents proline acid, Yellow represents cysteine acid , and Olive green represents leucine acid.

Figure 3: PyMol result observed in cartoon and electrostatic form (protein structure)

The PyMOL visualization in various forms enhanced the understanding of its conformation, interactions, etc. This serves as a foundation for further investigation into molecular behavior.

B. Sequence Analysis

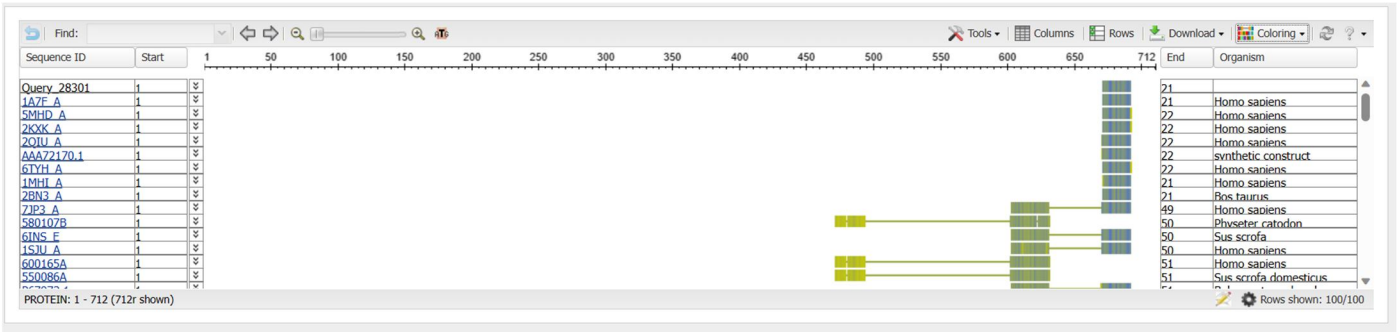


Figure 4: BLOSSUM 62 RESULT CHAIN A

BLOSSUM displays the degree of match of residue related to each alignment positions. Here, blue represents better match and green represents worst match. BLOSSSUM 62 is a default scoring matrix. In this, amino acid sequences exhibited a minimum of 62% identity when tow proteins were aligned.

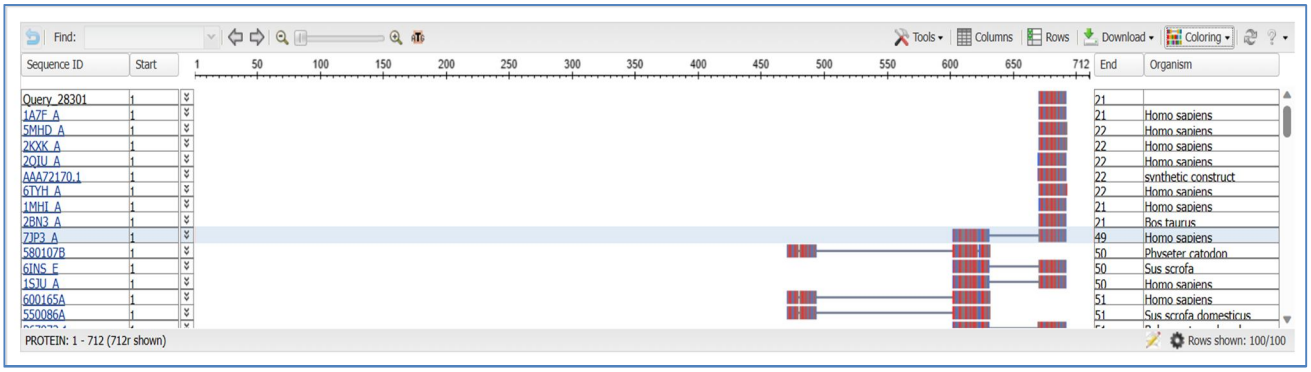


Figure 5: Cobalt result (hydropathy scale) – hydrophobic amino acidsshown by red color while hydrophilic amino acids are shown by blue color

The Cobalt Hydropathy Scale is a method used to assess the hydrophobicity or hydrophilicity of amino acid residues in a protein sequence. It assigns a numerical value to each amino acid based on its relative hydrophobic or hydrophilic nature.

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Chain B, PROTEIN (INSULIN) [Homo sapiens]	Homo sapiens	102	102	100%	1e-25	100.00%	30	1B9E_B
Chain B, R6 INSULIN HEXAMER [Homo sapiens]	Homo sapiens	97.8	97.8	100%	1e-23	96.67%	30	1AI0_B
insulin B chain [synthetic construct]	synthetic construct	97.8	97.8	100%	1e-23	96.67%	31	AAA72171.1
Modified A22Gly-B31Arg Human Insulin [Homo sapiens]	Homo sapiens	97.8	97.8	100%	1e-23	96.67%	31	2LGB_B
Structure of Glargine insulin in 20% acetic acid-d4 (pH 1.9) [synthetic construct]	synthetic construct	97.8	97.8	100%	1e-23	96.67%	32	6K59_B
Human Insulin Mutant A22Gly-B31Lys-B32Arg [Homo sapiens]	Homo sapiens	97.8	97.8	100%	1e-23	96.67%	32	2KXK_B
THREE-DIMENSIONAL SOLUTION STRUCTURE OF AN INSULIN DIMER. A STUDY OF THE B9(ASP) MU...	Homo sapiens	96.5	96.5	96%	3e-23	96.55%	30	1MHI_B
insulin [Homo sapiens]	Homo sapiens	97.8	97.8	100%	1e-22	96.67%	51	600165A

Figure 6: Protein description of human insulin (BLAST)

It is a statistical theory to produce a bit score and E value for each alignment score. The maximum score is the highest alignment score calculated from the sum of the rewards for matched nucleotides and penalties for mismatches and gaps while the total score means that the sum of alignment scores of all segments from the same subject sequence. Query cover [age] is the percentage of the query length that is included in the aligned segments. The lower the E value or closer to zero – the more significant match.

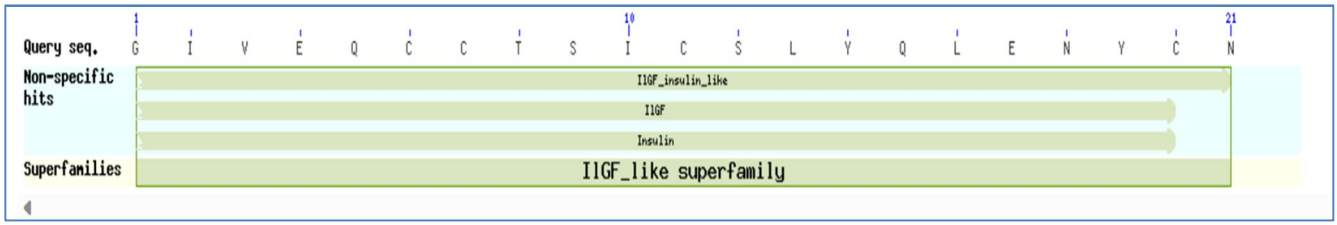


Figure 7: Conserved domain result

Conserve domain of chain A protein (insulin).

Human insulin is a pivotal hormone produced by the pancreas that plays a crucial role in regulation of glucose metabolism in the human body. It helps in maintaining stable blood sugar levels, which is essential for overall health and well-being. Insulin is an integral part of the endocrine system and is involved in various physiological processes, like growth and development.Human insulin is synthesized within the beta cells of the pancreas, specifically in clusters called the islets of Langerhans.These cells are responsible for monitoring blood glucose levels. Firstly, structure analysis in RasMol was done. It involves using the software to explore the 3D arrangement of molecular structure. To perform computational modeling, the target information, like the PDB ID (1B9E) was obtained from National Center for Biotechnology Information (NCBI) database. All the data all over the world is submitted in NCBI, EBI and DDBJ. The 3D structure of the human insulin (target) was obtained from the PDB (Protein Data Bank).

It began by loading a molecular structure in RasMol, its common file format is PDB (Protein Data Bank) which can be easily viewed in RasMol. Once the structure is loaded, RasMol displays the molecule as a graphical representation in the viewing window. Different visualization styles can be used to display different aspects of the structure like, wireframe, ball and stick, space filling, etc. RasMol also allows to color atoms and residues. Also, desired atoms and residues can be selected to highlight them. It's quite a powerful tool to gain insights about the structure of biomolecules. PDB plays a very important role in research work as it provides standardized database for users. The data in the PDB is expressed in the PDB format that includes atomic coordinates, atom types, bond lengths, and other essential information required to describe the molecular structure. Each entry in the PDB is related to a unique molecule. Insulin also plays a very important role in lipid (fat) metabolism. It helps in the storage of fatty acid in adipose tissue. It also facilitates the uptake of potassium ions into cells. The MMDB (Molecular Modeling Database) is used to store and provide access to experimentally determined macromolecular structures, including proteins, nucleic acids, and their complexes. Now, the retrieved FASTA sequences of the protein from MMDB were put to BLAST to analyze similar sequences. The analysis came out to be 96% identical between human insulin structure [homo sapiens] and other sequences. By using COBALT, it provided a graph that shows the matches and mismatches in alignment. It showed the matches in blue across the sequences and gaps in yellow color. The molecular structure file was loaded In PyMOL (by drag and drop). It also offers various visual representations like ribbon, stick and ball, wireframe, etc. The molecule can be colored in a desired way. So the structure is observed deeply by using PyMOL software.

IV. CONCLUSION

In conclusion, this project explored the potential of human insulin through the utilization of powerful software and tools. This research project exemplified the synergy between computational tools and experimental data. By using RasMol and PyMOL, we explored the 3D structure of human insulin. The utilization of BLAST and MMDB explored a wide perspective, enabling us to compare the human insulin sequence with other sequences in databases, highlighting its significance. This research project contributes to a deep and better understanding of human insulin by using software tools.

REFERENCES

[1] Wilcox G. (2005). Insulin and insulin resistance. The Clinical biochemist. Reviews, 26(2), 19–39.
[2] Rahman, M. S., Hossain, K. S., Das, S., Kundu, S., Adegoke, E. O., Rahman, M. A., Hannan, M. A., Uddin, M. J., & Pang, M. G. (2021). Role of Insulin in Health and Disease: An Update. International journal of molecular sciences, 22(12), 6403. https://doi.org/10.3390/ijms22126403
[3] Irl B Hirsch and others, The Evolution of Insulin and How it Informs Therapy and Treatment Choices, *Endocrine Reviews*, Volume 41, Issue 5, October 2020, Pages 733–755, https://doi.org/10.1210/endrev/bnaa015
[4] Gao, P., Hu, Y., Wang, J., Ni, Y., Zhu, Z., Wang, H., Yang, J., Huang, L., & Fang, L. (2020). Underlying Mechanism of Insulin Resistance: A Bioinformatics Analysis Based on Validated Related-Genes from Public Disease Databases. Medical science monitor : international medical journal of experimental and clinical research, 26, e924334. https://doi.org/10.12659/MSM.924334
[5] Rao, A. A., Tayaru, N. M., Thota, H., Changalasetty, S. B., Thota, L. S., &Gedela, S. (2008). Bioinformatic analysis of functional proteins involved in obesity associated with diabetes. International journal of biomedical science : IJBS, 4(1), 70–73.
[6] Rentería, M. E., Gandhi, N. S., Vinuesa, P., Helmerhorst, E., &Mancera, R. L. (2008). A comparative structural bioinformatics analysis of the insulin receptor family ectodomain based on phylogenetic information. PloS one, 3(11), e3667.
[7] Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). The Protein Data Bank. Nucleic acids research, 28(1), 235–242. https://doi.org/10.1093/nar/28.1.235
[8] Stephen K Burley and others, RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy, *Nucleic Acids Research*, Volume 47, Issue D1, 08 January 2019, Pages D464–D474, https://doi.org/10.1093/nar/gky1004
[9] Thomas Madej and others, MMDB: 3D structures and macromolecular interactions, *Nucleic Acids Research*, Volume 40, Issue D1, 1 January 2012, Pages D461–D464, https://doi.org/10.1093/nar/gkr1162
[10] Madej, T., Lanczycki, C. J., Zhang, D., Thiessen, P. A., Geer, R. C., Marchler-Bauer, A., & Bryant, S. H. (2014). MMDB and VAST+: tracking structural similarities between macromolecular complexes. Nucleic acids research, 42(Database issue), D297–D303. https://doi.org/10.1093/nar/gkt1208
[11] Sayers, E. W., Agarwala, R., Bolton, E. E., Brister, J. R., Canese, K., Clark, K., Connor, R., Fiorini, N., Funk, K., Hefferon, T., Holmes, J. B., Kim, S., Kimchi, A., Kitts, P. A., Lathrop, S., Lu, Z., Madden, T. L., Marchler-Bauer, A., Phan, L., Schneider, V. A., … Ostell, J. (2019). Database resources of the National Center for Biotechnology Information. Nucleic acids research, 47(D1), D23–D28. https://doi.org/10.1093/nar/gky1069
[12] NCBI Resource Coordinators (2013). Database resources of the National Center for Biotechnology Information. Nucleic acids research, 41(Database issue), D8–D20. https://doi.org/10.1093/nar/gks1189
[13] McGinnis, S., & Madden, T. L. (2004). BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic acids research, 32(Web Server issue), W20–W25. https://doi.org/10.1093/nar/gkh435
[14] Boratyn, G. M., Camacho, C., Cooper, P. S., Coulouris, G., Fong, A., Ma, N., Madden, T. L., Matten, W. T., McGinnis, S. D., Merezhuk, Y., Raytselis, Y., Sayers, E. W., Tao, T., Ye, J., &Zaretskaya, I. (2013). BLAST: a more efficient report with usability improvements. Nucleic acids research, 41(Web Server issue), W29–W33. https://doi.org/10.1093/nar/gkt282

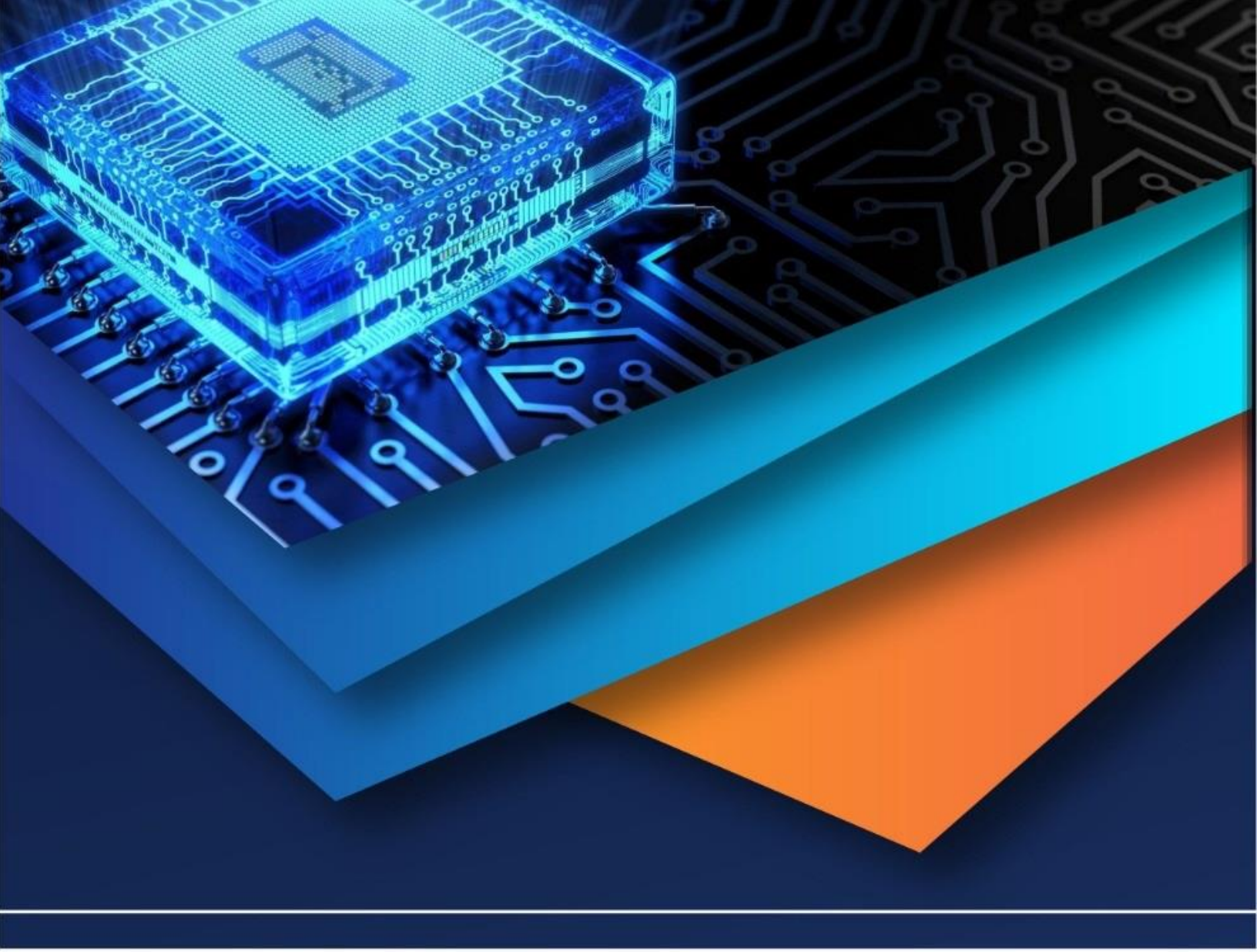
[15] Mark Johnson and others, NCBI BLAST: a better web interface, *Nucleic Acids Research*, Volume 36, Issue suppl_2, 1 July 2008, Pages W5–W9, https://doi.org/10.1093/nar/gkn201

[16] Jason S. Papadopoulos , RichaAgarwala, COBALT: constraint-based alignment tool for multiple protein sequences, *Bioinformatics*, Volume 23, Issue 9, May 2007, Pages 1073–1079, https://doi.org/10.1093/bioinformatics/btm076

[17] Seeliger, D., & de Groot, B. L. (2010).Ligand docking and binding site analysis with PyMOL and Autodock/Vina. Journal of computer-aided molecular design, 24(5), 417–422.https://doi.org/10.1007/s10822-010-9352-6

[18] Luciano Porto Kagami, Gustavo Machado das Neves, Luís Fernando SaraivaMacedoTimmers, Rafael Andrade Caceres, Vera Lucia Eifler-Lima,Geo-Measures: A PyMOL plugin for protein structure ensembles analysis,Computational Biology and Chemistry,Volume 87,2020,107322, ISSN 1476-9271, https://doi.org/10.1016/j.compbiolchem.2020.107322.

[19] Pikora, M., &Gieldon, A. (2015). RASMOL AB - new functionalities in the program for structure analysis. ActabiochimicaPolonica, 62(3), 629–631.https://doi.org/10.18388/abp.2015_972



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY