IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

# A Review: Modeling Of Selected Protein Structure From NCBI Database Using BLAST & FASTA

**Ashutosh Saini and Swati Gupta**

Department of Life Sciences, Vivekananda Global University

**Abstract**

Insulin is a small globular protein hormone that consists of two polypeptide chains, A and B, that are connected by two disulfide bonds. The A chain is composed of 21 amino acids, while the B chain is composed of 30 amino acids. The two chains are held together by a third disulfide bond between the A7 cysteine residue and the B7 cysteine residue. The tertiary structure of insulin is stabilized by a network of hydrogen bonds, salt bridges, and hydrophobic interactions between the amino acid residues. The hormone also contains a hydrophobic core that is responsible for its stability and solubility. The insulin molecule adopts a compact conformation that allows it to fit into its receptor on the cell surface. Alterations in the structure of insulin, such as mutations or chemical modifications, can affect its biological activity and lead to insulin resistance or other metabolic disorders.

**Keywords:** Protein**,** Insulin, BLAST, FASTA.

## Introductions

Insulin is a critical protein hormone that regulates glucose metabolism in the body. It is primarily produced by beta cells in the pancreas and acts on target tissues, such as muscle, liver, and adipose tissue, to promote glucose uptake and utilization. In addition to its role in glucose metabolism, insulin also plays a vital role in lipid metabolism and protein synthesis, making it a crucial hormone for overall metabolic health. Recently, there have been significant research efforts to develop next-generation insulin analogs that have improved pharmacokinetics and pharmacodynamics to better serve the needs of patients with diabetes. These efforts have resulted in several promising candidates that offer improved efficacy, reduced hypoglycemia risk, and increased administration convenience.

Consists of two polypeptide chains, A and B, that are connected by two disulfide bonds. The A chain is composed of 21 amino acids, while the B chain is composed of 30 amino acids. The two chains are held together by a third disulfide bond between the A7 cysteine residue and the B7 cysteine residue.

Insulin is synthesized as a precursor molecule called preproinsulin, which is then processed into proinsulin by the removal of the signal peptide. Proinsulin is subsequently cleaved by a specific protease called prohormone convertase to yield the mature insulin molecule. In addition to insulin, the processing of proinsulin also generates a smaller peptide called C-peptide, which is secreted in equimolar amounts with insulin. The secretion of C-peptide provides a useful marker of endogenous insulin secretion in patients with diabetes. The biological function of C-peptide is not fully understood, but studies have suggested that it may have beneficial effects on blood vessel and kidney function.

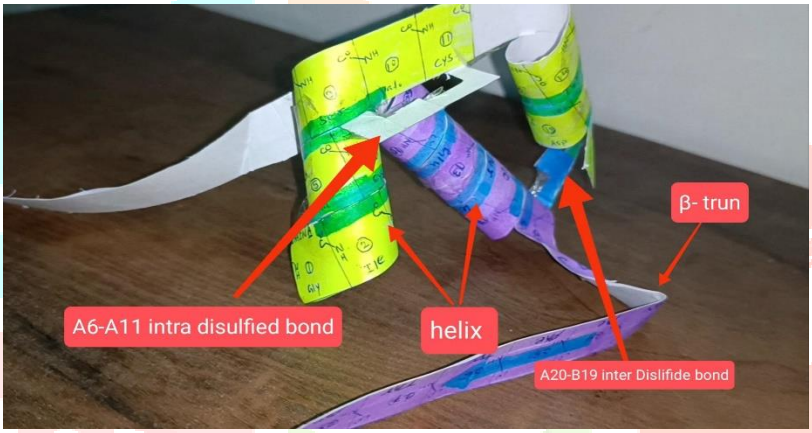



Fig. 1 - Insulin protein model.

## 3. Review of literature

### 3.1 Insulin and receptor interaction

Recent advances in structural biology have shed new light on the conformational changes that occur in insulin upon binding to its receptor. Studies have shown that insulin undergoes a conformational change from a closed, inactive state to an open, active state upon binding to its receptor. This conformational change is thought to be triggered by the binding of insulin to the alpha subunit of the receptor, which induces a conformational change in the beta subunit that results in the activation of downstream signaling pathways. The crystal structures of the insulin receptor in complex with insulin have provided insights into the molecular mechanisms of insulin signaling and have identified potential sites for therapeutic intervention in metabolic disorders such as type 2 diabetes.

### 3.2 T-R transition of insulin

The T-R transition is a significant conformational change that occurs in insulin upon binding to its receptor. In this transition, the insulin molecule undergoes a conformational change from a predominantly T-state (tensed or inactive state) to an R-state (relaxed or active state). The T-state of insulin is characterized by a compact conformation in which the two chains, A and B, are held together by three disulfide bonds and a network of hydrogen bonds, salt bridges, and hydrophobic interactions. In contrast, the R-state is characterized by a more extended conformation in which the A and B chains are separated by a distance of about 6 angstroms.

The T-R transition is triggered by the binding of insulin to its receptor on the cell surface. The binding of insulin to the alpha subunit of the receptor induces a conformational change in the beta subunit that results in the activation of downstream signaling pathways. This conformational change in the receptor is transmitted to the insulin molecule, leading to the disruption of the T-state conformation and the formation of the R-state conformation. The R-state conformation of insulin is thought to be essential for its biological activity, as it allows the hormone to bind to its receptor and activate downstream signaling pathways that mediate its effects.

### 3.3 Improved insulin stability through amino acid substitution

explores the impact of substituting specific amino acids in insulin on its stability (Gilliland, et al., 1992). The researchers discovered that certain substitutions, such as substituting asparagine with aspartic acid, enhanced the stability of the insulin molecule without reducing its biological activity. Insulin is a crucial hormone that regulates blood sugar levels in the body. However, insulin is highly unstable, and its degradation can lead to reduced effectiveness and potentially harmful side effects (Kumar & Kaur, 2019). The researchers in this study aimed to identify ways to enhance the stability of insulin without harming its function. By analyzing the insulin molecule, the researchers identified specific amino acids that could be involved in the molecule's stability. They then substituted these amino acids with other amino acids to determine whether they could enhance the stability of the insulin molecule. This finding has significant implications for the development of new insulin therapies for individuals with diabetes. By improving the stability of insulin, researchers may be able to develop more effective and longer-lasting treatments for diabetes. Additionally, this study provides critical insights into the structure and function of insulin, which may help researchers better understand how this vital hormone works in the body.

| Insulin type | Number of Amino acid | Amino acid sequence | Disulfide Bonds | Molecular Weight | Conserved sequences | Variable sequence |
|---|---|---|---|---|---|---|
| **A Chain** | 21 | GIVEQCCTSICSLYQLENYCN | A7-B7, A20-B19 | 5808 Da | GIVQCCSCSLYQLEYCN | E,T,I,N |
| **B Chain** | 30 | FVNQHLCGSHLVEALYLVCGERG FFYTPKT | A7-B7, A20-B19 | 5733 Da | FVQHLCGHLVEALYLVCGER GFFY | N, S, T, P, K, T |

### 3.4 $Zn^{+2}$ role in dimer and hexamer

The protein exists as a dimer, with two insulin molecules linked together by zinc ions at the dimerization interface. The role of zinc in insulin dimerization has been extensively studied, and it has been shown to play a critical role in stabilizing the dimeric form of insulin. Zinc ions coordinate with histidine residues in both A and B chains of the protein, forming a tetrahedral complex that links the two insulin molecules together. The formation of the zinc-histidine complex is crucial for the proper folding and stability of the insulin dimer, and any disruption in this coordination can lead to the formation of unstable insulin oligomers. The hexameric form of insulin is another important structure that is formed by the assembly of three insulin dimers. The hexameric form of insulin is stabilized by additional interactions between the A and B chains of each insulin dimer, as well as by the formation of an extensive network of hydrogen bonds and hydrophobic interactions between adjacent insulin molecules. The presence of zinc ions is also required for the proper assembly of the insulin hexamer, and disruption of zinc coordination can lead to the formation of unstable insulin oligomers. The importance of zinc ions in insulin dimerization and hexamerization is further highlighted by the fact that mutations in the zinc-binding sites of insulin can lead to the development of diabetes. In some cases, these mutations can destabilize the insulin dimer or hexamer, leading to the formation of amyloid fibrils that can clog the pancreas and impair insulin production. Therefore, understanding the role of zinc ions in insulin structure and function is crucial for developing new treatments for diabetes and other metabolic disorders.

### 3.5 Next generation insulin analog

To maintaining the structural integrity of insulin, it is essential to consider the clinical goals when developing next-generation insulin analogs. Insulin therapy is a cornerstone of diabetes management, but it can be challenging for patients to achieve optimal glycemic control due to the risk of hypoglycemia and the need for frequent injections. Therefore, the clinical goals of next-generation insulin analogs are to improve glycemic control, reduce the risk of hypoglycemia, and enhance patient convenience. Several approaches have been taken to achieve these goals, including the development of ultra-rapid-acting insulin analogs that can mimic the kinetics of endogenous insulin and the introduction of new delivery systems, such as oral insulin and inhalable insulin. Additionally, next-generation insulin analogs that target specific aspects of insulin signaling, such as the insulin receptor or downstream signaling pathways, offer new avenues for improving insulin therapy (Jarosinski *et al.*, 2022). One approach to developing next-generation insulin analogs is the use of ultra-rapid-acting insulin analogs. These analogs have a faster onset of action and a shorter duration of action than currently available rapid-acting insulin analogs, better mimicking the kinetics of endogenous insulin. Another approach to improving insulin therapy is the development of new delivery systems, such as oral insulin and inhalable insulin.

## 4. Material and method

For the protein structure analysis of insulin following software and website used

**RCSB PDB**

### 4.1 NCBI

All type of protein's details i.e., amino acid sequence, protein data bank (pdb) was collected from NCBI website for research work.

### 4.2 BLAST

An alignment between a nucleotide or protein sequence, known as a "query," and nucleotide or protein sequences within a database, known as "subject" sequences, is created using a computer suite known as BLAST, which stands for Basic Local Alignment Search Tool.

### 4.3 FASTA

Filtering, transforming, annotating and analysing biological sequence data were made simple and powerful with the using of FAST (FAST Analysis of Sequences Toolbox), an open-source command-line toolkit.

### 4.4 Rasmol

A molecular graphics tool called RasMol is designed to visualise proteins, nucleic acids, and tiny molecules. The programme aims to demonstrate, instruct, and produce photos of publication quality.

## METHODOLGY

**Following steps used for our research work**

1 First, we opened internet browser and search RCSB PDB, and open that official protein data bank website.

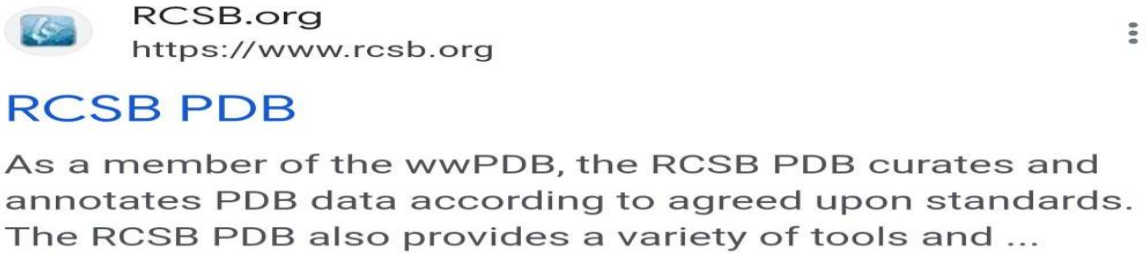
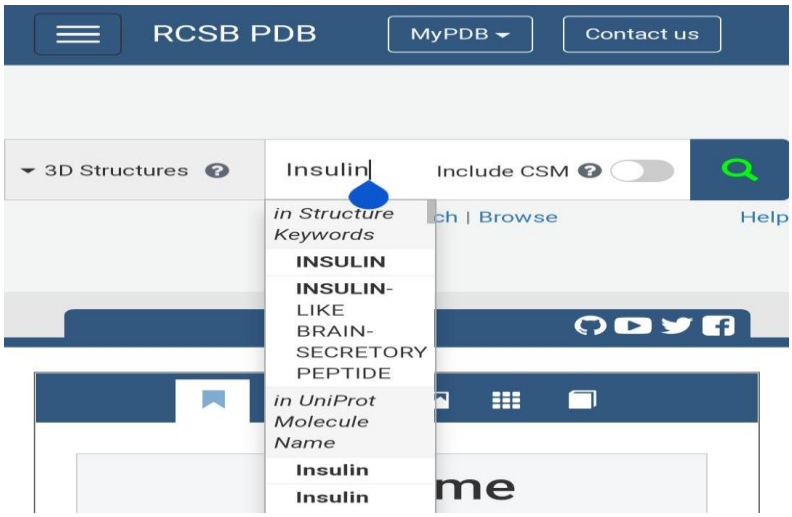



**Fig. 3.** Protein data bank

**2** After open website, search insulin on the top.




**3** where we were download FASTA sequence, from right corner of the desktop.

We were also download PDB format of that file that help in visualize protein 3D structure in different angels and parameter in rasmol application. We can also extract aur FASTA sequence from uniport official website, in sequence and isoform section after search insulin:

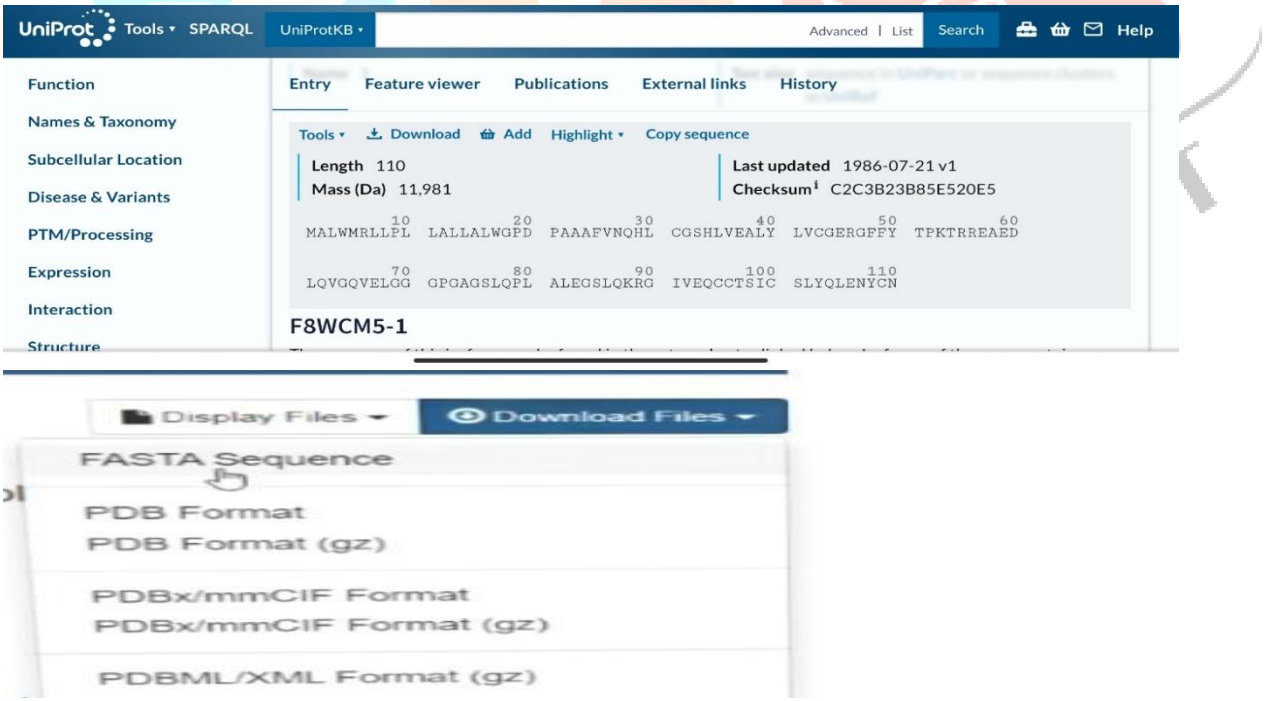
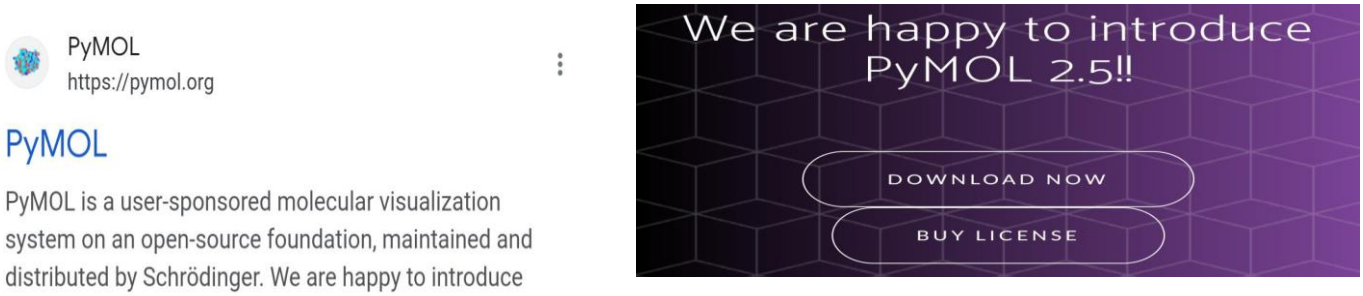



Fig. 4 uniport

**4 PyMOL**

Firstly, we were download and installed PyMOL desktop version 2.5 (current version 2023), from




official website pymol.org.

**Fig. 5 PyMOL**

## Application of software used for research work

For analysis of the protein that we saw previously in protein data bank, in 3D using PyMOL, first we have to copy that protein unique code from protein data bank (PDB), before open PyMOL. This process shown in fig.).

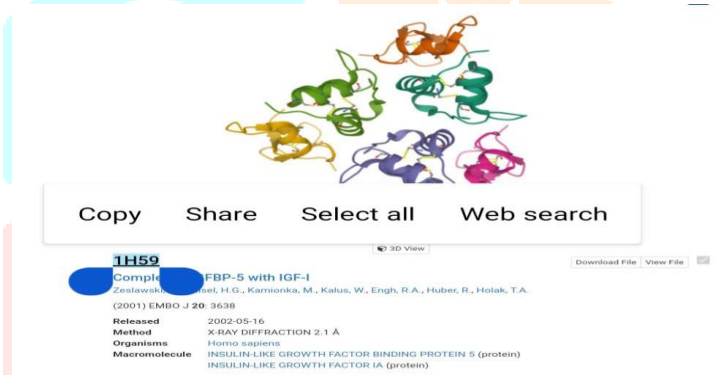



**Fig. 6 PDB**

Source- Zeslawski, W., Beisel, H.G., Kamionka, M., Kalus, W., Engh, R.A., Huber, R., Holak, T.A.(2001) EMBO J **20**: 3638)

So, in this picture we can see the code: - 1H59.

This code use in PyMOL software for import 3D insulin protein. Code import with fetch 1H59.

Above FASTA sequence use in protein BLAST, for comparison with another species insulin protein Amino acid sequence.

**4.7: Building paper model of 3D insulin**

## 5. Result and Discussion

### 5.1 BLAST

We applied the FASTA sequence of human insulin from uniport INS_HUMAN (P01308)

**5.2 FASTA** sequence run in BLAST (blastp)

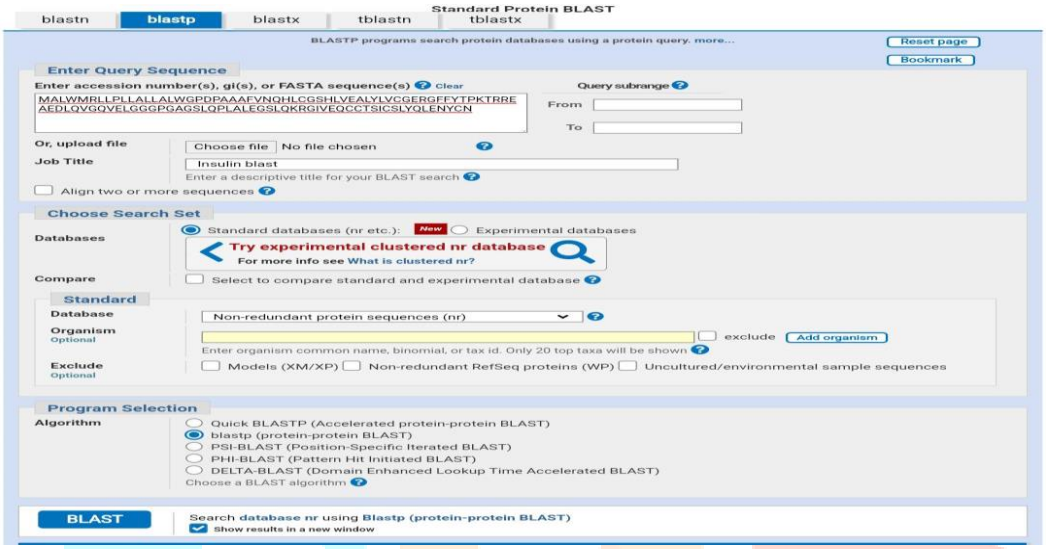



**Fig. 7 BLAST**

Insulin protein sequence have highly conserve sequence as we can see in blastp result, even in lower identical species with human.

**In A chain** of insulin have mostly conserved sequence except A4(E), A8(T), A10(I), A18(N).

A4(E) is substitute mostly with Asp(D) in Mastomys coucha, **Arvicanthis niloticus**, Mus musculus. A8(T) is substitute with Ala(A), A10(I) is substituted with Val(V) and A18(N) is substitute with His(H) in **Prionailurus bengalensis.**

In insulin construct A1(G) is substitute with Cys(C).

**Fig. 8. BLASTp of H_INS in which red circle was shown A-chain amino acids of insulin.**

**In B chain** of insulin have mostly conserve sequence except B3(N), B9(S), B27(T), B28(P), B29(K), B30(T).

B3(N) is substitute with K(Lys) in Arvicanthis niloticus, **Mastomys coucha.** B9(S) is substitute with P(pro) in


Rattus losea and Rattus rattus.

B27(T) and B28(P) are substitute with A(Ala) Amino acid in Lipotes.

B29(K) is substitute with M(met) and K(Lys) respectively in Arvicanthis niloticus and Mastomys coucha.

B30(T) substitute with S(Ser) in Arvicanthis and Mastomys coucha. B30(T) also substitute with A(Ala) in Hyaena hyaena and Prionailurus bengalensis

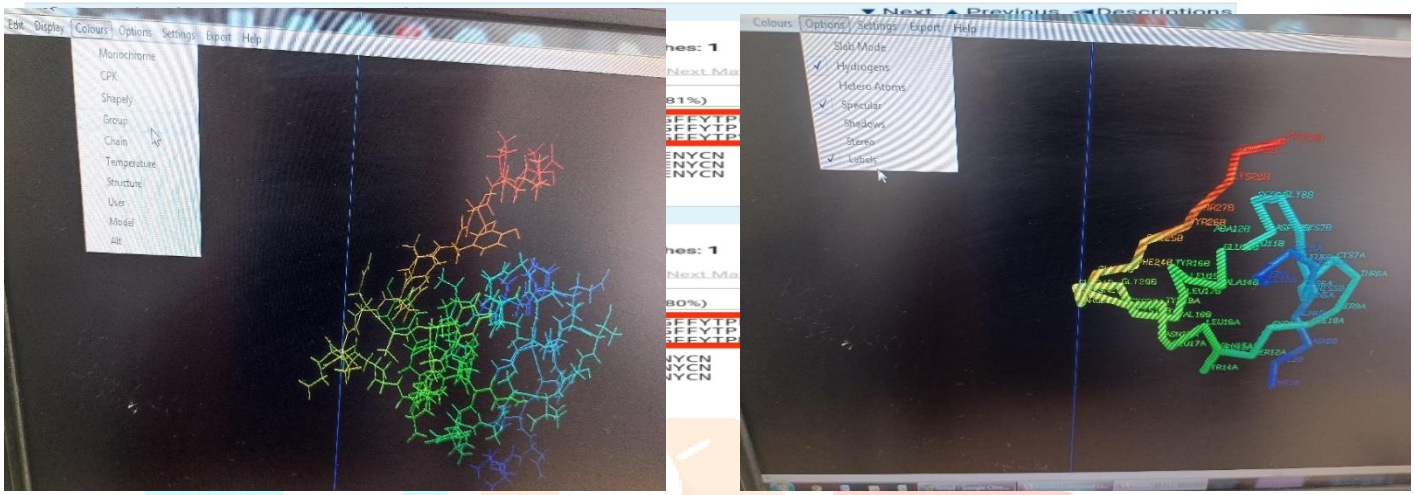

**Fig.9 BLASTp of H_INS in which red circle shown B chain amino acid.**

### 5.3 RasMol

From protein data bank (PDB) we download insulin protein pdb format file for 3D structure analysis in rasmol, we can import that particular pdb format file in rasmol software, and can analysis the structure in 3D, in different angels and we were also tag to the protein with amino acid in rasmol. Structure can be visualized in different form like stick, wire, backbone, boll and ribbon form as aur desire, and can understand the structure variability.

**Fig.10 RasMol 3D structure analysis of insulin with different forms.**

Chain A have 2 helix, and chain B have only one helix and connect by two inter Dislifide bond A20-B19 and



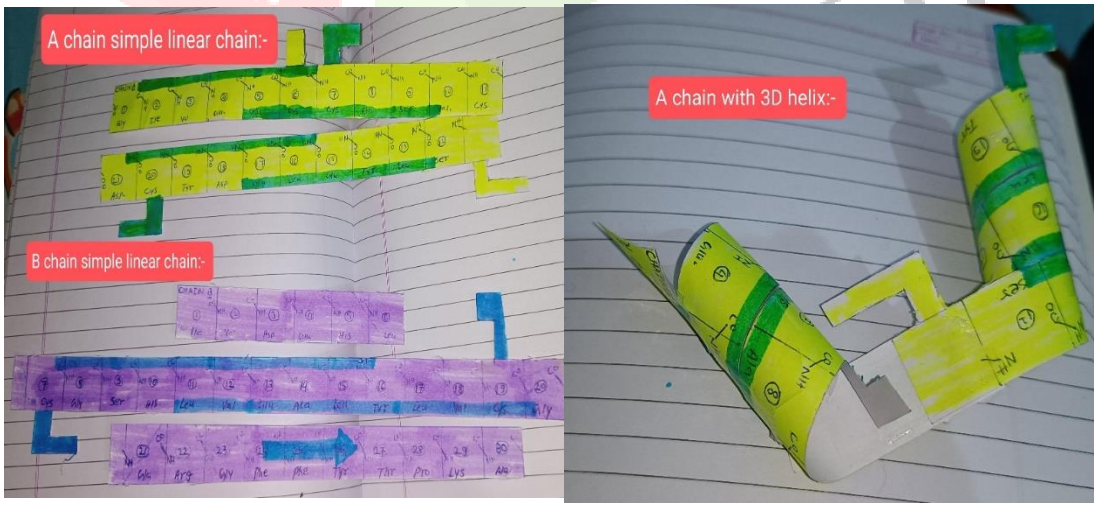
A7-B7.

### 5.4 Build paper model of insulin 3D

By some instructions (pdb101) we make a 3D protein structure model with paper, that also labled with amino acid, and show helix, and intra and inter Dislifide bond.

We were made a model contains A chain and B chain of insulin. following picture shown the

**Fig. 11 A and B chain of insulin protein at paper model.**



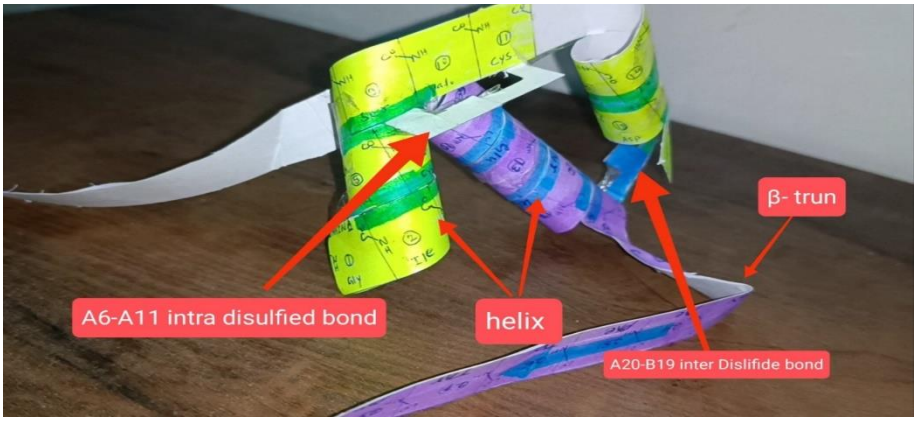



**Fig :12 Paper model of insulin protein 3D structure.**

Source:- https://pdb101.rcsb.org/learn/paper-models/insulin-activity-page

## Conclusion

Insulin is a protein hormone that plays a critical role in regulating blood sugar levels. The 3D arrangement of insulin consists of a pair of polypeptide chains, namely an A chain and a B chain, which are joined by disulfide bonds. The A chain consists of 21 amino acid residues, while the B chain contains 30 amino acid residues. The two chains are held together by two disulfide bonds, one between cysteine residues at positions A7 and B7, and another between cysteine residues at positions A20 and B19.The overall structure of insulin is a compact globular protein, with the A chain forming an alpha-helical structure and the B chain forming a beta-sheet structure. The two chains are arranged in a characteristic cross-linked structure, known as the insulin fold. The 3D structure of insulin is critical for its biological activity and has been extensively studied to understand its function and to design drugs for the treatment of diabetes.

## Reference

Gilliland, L. K., Norris, S. R., Marbet, A. J., & Shoemaker, K. R. (1992). Improved insulin stability through amino acid substitution. Protein Engineering, Design and Selection, 5(6), 629-635. https://doi.org/10.1093/protein/5.6.629

Kumar, R., & Kaur, N. (2019). Insulin: molecular structure and function. In Molecular and Cellular Endocrinology (pp. 1-12). Springer. https://doi.org/10.1007/978-981-13-6737-2_1

Dodson, G., & Steiner, D. (1998). The role of assembly in insulin's biosynthesis. Current Opinion in Structural Biology, 8(2), 189-194.

Weiss, M.A. (2009). Insulin receptor binding to insulin: an insider's view. Biochemistry, 48(51), 9969-9976.

Whittingham, J.L., & Dodson, E.J. (1998). The structure of insulin: a small protein with a big punch. Current Opinion in Structural Biology, 8(2), 189-194.

Gloyn, A.L. (2003). The search for maturity-onset diabetes of the young genes. Diabetes Care, 26(2), 337-338.

Jarosinski, M. A., Chen, Y. S., Varas, N., Dhayalan, B., Chatterjee, D., & Weiss, M. A. (2022). New Horizons: Next-Generation Insulin Analogues: Structural Principles and Clinical Goals. The Journal of clinical endocrinology and metabolism, 107(4), 909–928. https://doi.org/10.1210/clinem/dgab849.