Equinox Issue

SCHOLARLY REVIEW ONLINE

Structural impotence of PDX-1 protein in the epigenetic control of insulin expression

By Jia Wang

Author Biography

Jia Wang is a senior student at Beijing City International School in China. She is passionate about genetics and protein structure. Throughout writing this paper, she has learned the importance of protein structure and their disordered regions in the cell signaling pathway controls. She plans to continue her career by pursuing molecular biology and bioinformatic majors to develop knowledge around epigenetics and the regulation of cell signaling in the future.

Abstract

This study investigates the role of pancreatic and duodenal homeobox 1 (PDX-1) as an important transcription factor in the epigenetic regulation of insulin expression. The research focused on PDX-1 as a disordered protein, and how the flexibility of its intrinsically disordered regions (IDRs) allows PDX-1 to interact with various transcription factors and epigenetic enzymes, facilitating histone modifications and insulin transcription. Particularly, the role of PDX-1 in processes like acetylation, methylation, and demethylation was discussed by detailing its interactions with key enzymes such as p300, Set7, and JMJD3. For instance, PDX-1's interaction with p300 leads to histone H4 hyperacetylation, enhancing insulin gene expression, while Set7-mediated methylation of H3K4 further promotes transcription. In addition, the demethylation of H3K27 by JMJD3 is also facilitated by the PDX-1 protein. Finally, structural predictions of PDX-1 using AlphaFold III and EMS Fold also highlight the importance of its IDRs in these interactions. The findings highlight the significance of PDX-1 in insulin regulation and offer insights into potential therapeutic targets for treating diabetes from an epigenetic approach.

Keywords: PDX-1, Diabetes, Epigenetic regulation, Histone modification, Insulin, Acetylation, Methylation, Demethylation, Intrinsically disordered proteins (IDPs), AlphaFold

Introduction

Diabetes is an endocrine system disease that is characterized by high levels of blood glucose resulting from the dysfunction of the pancreatic β cells (Molina et al., 2007). Within the cell insulin genes are rescripted and translated to create insulin protein, which is an important hormone maintaining glucose homeostasis, after translations, insulin is modified and released to the blood circulatory system. Insulin protein acts on the insulin receptor protein of the target cells, which allows glucose to be absorbed by the cell for cellular activities. However, for patients with diabetes, the pancreatic B cells fail to secrete functional insulin or insulin receptors, prohibiting normal glucose metabolism activities from happening (Molina et al., 2007).

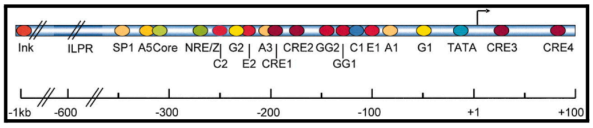
There are two common types of diabetes. Type I diabetes (T1D) primarily results from genetic causes, such as errors in gene sequences, including the PTP22, HLA complex, and INS gene (Seligman & Shearer, n.d.). Since gene sequence is translated to peptides, this creates insulin or insulin receptors with abnormal structures that cannot respond to blood glucose levels. Type II diabetes (T2D), on the other hand, is caused by the misregulation of the epigenetic system, which is about the control and regulation of gene expression rather than the gene itself. During transcription of the insulin gene, several proteins, known as the gene transcription factors, need to bind to specific regions on the DNA sequences, known as the promoter regions, to initiate the transcription of the insulin gene. These transcription factors are mostly glucose- and lipid-responsive and can be affected by factors like diet and lifestyle (Seligman & Shearer, n.d.)
Due to the increase in sugar and industrial fast-food products, increasing numbers of young people are suffering from insulin resistance, which highlights the importance of research into the T2D mechanism.
Recent research has suggested that the absence of transcription factors is highly associated with the increased risk of T2D (Elliott & Pavlínková, n.d.). Specifically, transcription factors such as PDX1 are also revealed to be involved in the regulation of transcription factors during the INS gene expression, as well as the chromatin modifications of methylation and acetylation of chromosomes, which also contribute to the development of T2D. However, the interaction between gene transcription factors is not well understood. Hence, the research will primarily focus on the epigenetic regulation of insulin expressions.

This research will include a review of previous studies regarding the epigenetic control of T2D. Moreover, advanced protein interaction prediction models including the new Alpha fold III model, and EMS fold are used to present the interaction between some important transcription factors during the methylation and acetylation of the chromosome will provide a more visual understanding of the interaction of insulin genes, potentially providing valuable insights into the mechanisms underlying T2D and offering potential treatments that aim at improving insulin production.

Overview of Insulin Regulation Network

The expression of the insulin gene is regulated via the specific binding of transcription factors to the insulin promoter regions, which are located around nucleotides -340 to -90 (Wright et al., 1986). As shown in Figure 1, studies have revealed some sites such as A4, A3, A2, C1, E1, A1, and G1, to be important binding sites for regulatory elements due to their matching chemical and physical properties (Fig. 1). The binding activity of transcription factors is regulated mainly by glucose and fatty acid level in the bloodstream through different pathways.

Figure 1
Promoter regions of the homo insulin gene

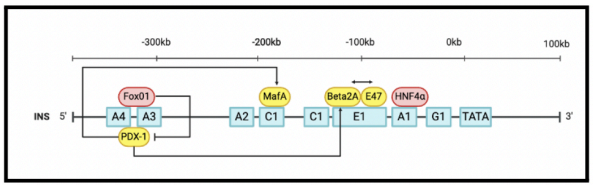


Reprinted from Wright et al. (1986). The figure displays the promoter region of INS. Color circles each represent possible binding locations of transcription factors on the INS gene.

Regulation of insulin gene through glucose metabolism involves transcription factors including pancreatic/duodenal homeobox-1 (PDX-1), a mammalian homolog of avian MafA/L-Maf (MafA), Beta2/Neuro D (B2), etc 90 (Wright et al., 1986). During glucose metabolism, as the blood sugar rises, the ratio between ATP and ADP increases, which triggers the influx of Ca+ which causes depolarization of B-cells. Therefore, allowing the transcription factors to enter the nucleus and bind to their specific insulin promoter region, such as the binding of PDX-1 to A3 site, MafA to A3, and NeuroD1/BETA2 to E1 site of the Insulin promoter region 90 (Wright et al., 1986). Hence, this is essential for enhancing the affinity of transcription factors to the DNA binding site.

On the other hand, the expression of insulin genes can also be affected by blood lipid levels, such as free fatty acids (FFAs) and triglycerides. The major transcription factors involved in the lipid-insulin signaling pathway include Hepatocyte Nuclear Factor 4 Alpha (HNF4α), MafA, and PDX-1, which are part of the glucose regulation of insulin expression and can also be affected by the lipid concentration in the bloodstream. This suggests that the two regulation networks work interchangeably to achieve the regulation of insulin secretions. Noticeably, as the lipid concentration in blood sugar rises, the Fox01 will bind to act as a competitive inhibitor that prevents the binding of PDX-1 to the A3 and A4 sites, causing a decrease in the binding activity of PDX-1.

Figure 2
Interaction of PDX-1 with other transcription factor and their position on the INS gene



The blue box indicates the region of the INS promoter region involved in binding activity with transcription factors. The red box indicates transcription factors involved in mainly lipid metabolism of INS expression, whereas the yellow box indicates transcription factors involved in glucose metabolism for regulating INS expression.

PDX-1

Pancreatic and duodenal homeobox 1 (PDX-1) is a transcription factor that is significantly involved in the regulation of insulin secretion. Both the A3 box and GG box of insulin promoter regions are conserved across vertebrate species (Hay & Docherty, 2006), suggesting the binding of PDX-1 to the A3 box in the expression of insulin to happen. PDX-1 is highly expressed by the pancreatic beta cells (Zhou et al., 2008). PDX-1 is also secreted by organs such as the stomach, which, considering that the regulation of insulin gene expression activity also influences both glucose and, therefore, also highlights the close relationship between T2D and diet, and PDX-1 as an essential regulator in both the lipid and glucose-dependent transcription of insulin.

PDX-1 regulatory pathway

Research conducted by Tuttle et al. (2020) suggests that the expression of PDX-1 is heavily dependent on glucose metabolism. In response to increasing blood sugar levels, glucose metabolism processes produce signaling molecules that activate kinases such as Akt and MAPK, which phosphorylate PDX-1. This phosphorylation allows the DNA-binding terminal of PDX-1 to interact with the nuclear import receptor, importin β, on the nuclear membrane, allowing the transcription factor to enter the nucleus through the pore complex and perform DNA-binding activities (Hartl et al., 1994). Apart from the glucose regulation pathway, PDX-1 regulation can also be stimulated by growth factors such as IGF-1 and mTOR, which modulate the phosphorylation status of PDX-1 similarly to Akt and MAPK kinases (Smith et al., 2019).

As evaluated in the overview of the insulin regulation pathway, PDX-1 is involved in lipid and glucose regulation of insulin secretion. PDX-1 binds directly to Beta2 (NeuronD1), which stabilizes the formation of the transcription complex. PDX-1 also binds to an adjacent promoter region with E47. This interaction enhances the affinity of both transcription factors to their DNA-binding regions. In high-level expression of insulin, PDX-1 and MafA bind cooperatively to the A3/A4 region of the insulin promoter (Tuttle et al., 2020).

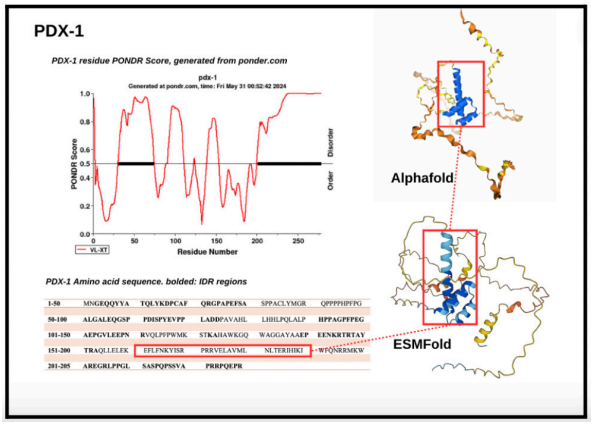
Structure of PDX-1

PDX-1 is structurally unique compared to other transcription factors, as it is considered a disordered protein, with the majority of its structure lacking a specific conformation, known as an intrinsically disordered region (IDR). The Prediction of Nature Disorder Region (PONDR) score indicates the level of disordered portions of the protein that lack specific conformations (Fig. 2, Fig. 3).

Out of the 283 amino acids, the IDR region accounts for almost 50% of the protein structure (Uversky, 2019).

The 283 amino acids in human PDX-1 are organized into three structurally distinct domains: the DNA binding domain (73-133), the N-terminal, and the C-terminal. The DNA-binding domain, which consists of only 60 residues, has specific sequences that enable it to bind to the DNA promoter region, with the core DNA motif located at the center of the protein complex, highlighted in red in Figure 3 below. The remaining amino acids, which form the N and C terminals, are crucial for interactions with other coactivators and transcription factors, occurring as long helical strands of polypeptides surrounding the DNA-binding domain (Babu, 2016; Ward & Jones, 2010).

Figure 3
PDX-1 structural prediction generated by Alphafold III and ESMFold Model.



The graph to the left shows the PDX-1 residue PONDR score that indicates the level of the order of PDX-1 protein. The table below indicates, with the bolded region indicating IDR regions. The two protein structures to the right are generated from AlphaFold III and ESMFold. The helical structures that are similar in two structures are boxed.

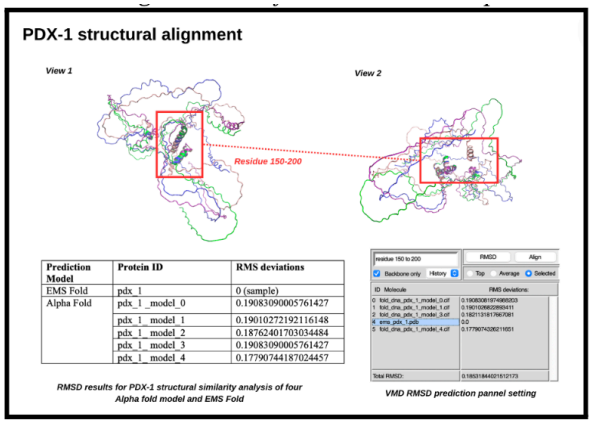
Unlike the DNA-binding region, which possesses a specific structure, both the N and C terminals are IDR regions that lack specific conformation. In human PDX-1, the IDR accounts for almost 50% of the protein structure (Fig. 3). This property offers flexibility to these domains, allowing them to interact with a diverse range of biological molecules (Uversky, 2019; Ward & Jones, 2010). Notably, the C-terminal is primarily involved in proteasomal degradation, which is stimulated under certain conditions to prevent disease-associated mutations by cleaving the C-terminal. The N-terminal, which is also disordered, has been suggested to perform essential functions in preproinsulin gene activation (Babu, 2016). Nevertheless, both of these domains are involved in the epigenetic modification of the insulin gene, including histone modification and transcription factor regulation (Babu, 2016).

Evaluating the accuracy of prediction

To evaluate the accuracy of the prediction, the protein models were loaded into VMD. Then, the RMS deviations (RMSD), which quantify the structural similarity of the prediction results, were calculated using RMSD analysis. According to the literature, an RMSD value ≤ 2.0 Å is generally considered a good result, suggesting that the protein structures are very similar to each other (Schröder et al., 2018).

Figure 4

The RMSD alignment result for PDX-1 structural predicted by Alpha Fold and EMS Fold



The protein structure was algin from residue 150-200. View 1 and View 2 display the protein chain prediction results in different colors. The red block highlights the regions that align perfectly (residue 150-200). The graph below shows the numerical RMS deviation from the prediction.

In Graph 5, the models from AlphaFold were compared with EMSFold, with the PDX-1 prediction from EMSFold being the sample. Of the alignment results, sequence 150-200, which is the DNA-binding domain, is the most conserved, as shown in the figure; the three alpha helices overlap perfectly. However, the polypeptide chains around the three alpha helices are poorly conserved, which indicates, to some extent, the flexibility of disordered regions. It should be noted, however, that poor alignment may also result from inaccurate predictions.

Moreover, referring to the RMSD values, the prediction result indicates that, compared to the EMSFold prediction, model 4 deviates the least, with an RMSD of around 0.178. In contrast, model 1 deviates the most, with an RMSD of around 0.191. This suggests that the most confident prediction result from AlphaFold III differs the most compared to EMSFold. Therefore, the results suggest that the central helical structures are mostly accurately predicted by the model, while the prediction

results for the disordered regions, due to their flexibility and lack of structural conservation, are less accurate (Schröder et al., 2018).

PDX-1 epigenetic modification

Due to its unique disordered structure, PDX-1 can interact simultaneously with DNA promoter sequences and other transcription factors, facilitating complex epigenetic modifications. These modifications include chromatin remodeling processes such as methylation, demethylation, and acetylation, which regulate gene expression. For instance, DNA methylation of the PDX-1 gene has been shown to reduce its expression in pancreatic islet cells, impacting insulin production. Histone modifications, like acetylation and deacetylation, also play a critical role in regulating chromatin structure and, consequently, the transcriptional activity of the insulin gene. When DNA is tightly wrapped around histones, transcription is inhibited, while loosening of the chromatin through acetylation allows transcription to proceed (Liu et al., 2021; Covington et al., 2009)

The Role of PDX-1 in Acetylation of Histone Protein

Acetylation is the process of attaching an acetyl group to the histone protein, which loosens the DNA wrapping around the histone protein and allows gene expression. Acetyltransferase (HAT) enzymes such as p300 and CBP are crucial for the signal communication of key transcription factors. Studies have revealed that PDX-1 plays a crucial role in assembling HATs in insulin secretion. The interaction between PDX-1 and p300 leads to the hyperacetylation of histone H4, a hallmark of gene transcription enhancers across cells. This interaction occurs through the physical binding of p300 to the IDR region of PDX-1, specifically at the TAZ1 and TAZ2 domains, forming a PDX-1/p300 complex. The hydrophobic characteristics of the p300 residue allow it to bind to PDX-1 despite poor sequence conservation, stabilizing the IDR region of PDX-1 and transitioning it to a more ordered state (Liu et al., 2021)

Although specific insights into how PDX-1 supports histone H4 acetylation are not well understood, research suggests that knocking down the PDX-1 gene in insulinoma cells results in a decrease in p300 promoter activity, supporting the essential role of PDX-1 in hyperacetylation of histone H4 (Covington et al., 2009). This complex also enhances the binding affinity of PDX-1 to the A3 site, conserved among all mammalian insulin gene promoters, thereby increasing insulin gene expression.

The Role of PDX-1 in Methylation of Histone Protein

Methylation is a type of chromatin-level epigenetic modification where a methyl group is added to histone proteins, altering the DNA structure and gene activity. For example, methylation at H3K4 loosens the DNA wrapping around histones and is considered a marker of active gene expression, while methylation at H3K9 and H3K27 tightens DNA structure, repressing gene expression. Research suggests that PDX-1 is involved in the former regulation. In insulin regulation, PDX-1 recruits PTM enzymes, such as methyltransferase Set7, which adds a methyl group to H3 lysine 4 (H3K4me1), promoting insulin gene transcription (Liu et al., 2021)

Moreover, Set7 interacts with the IDR terminal of PDX-1, playing a crucial role in regulating chromatin structure and gene expression at insulin-related promoters, particularly by adding methyl groups to lysine residues Lys123 and Lys131, which are substrates for Set7 (Covington et al., 2009).

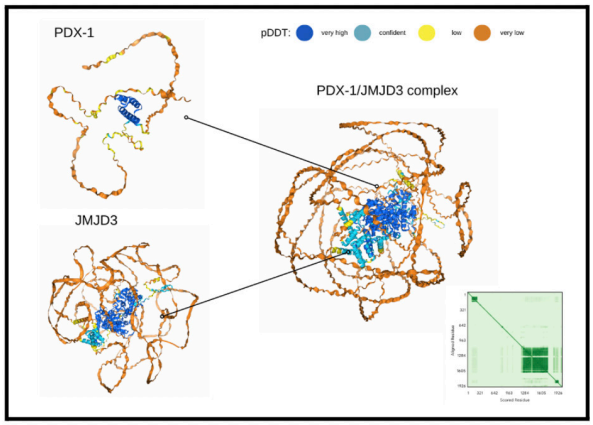
The Role of PDX-1 in Demethylation of Histone Protein

In contrast to methylation, demethylation involves removing methyl groups from histone proteins, often leading to increased gene expression. Under high glucose conditions, PDX-1 interacts with JMJD3, a demethylase enzyme that removes repressive methyl groups from H3K27 by phosphorylating FOXO1, enhancing the transcription of the insulin gene and increasing insulin synthesis (Liu et al., 2021).

Stimulating the Interaction of PDX-1 with Various Transcription Factors using Alpha Fold III To visualize the interaction between PDX-1 and various gene transcription factors, the protein interaction production model Alpha fold III is applied to estimate the protein-protein interaction. Alpha III is an advanced protein structure prediction model that generates the three-dimensional structure of the protein from their amino acid sequences, providing a visualized image of the protein, therefore, facilitating the study of the protein structure and function. Considering that the research primarily focuses on diabetes for the human species, thus only the protein structure for the Homo species is selected for testing, Therefore, the amino acid sequence of protein PDX-1, p300, JMJD3, and se7 is downloaded from the Unipro and ran using the Alpha fold server, yielding result as shown below.

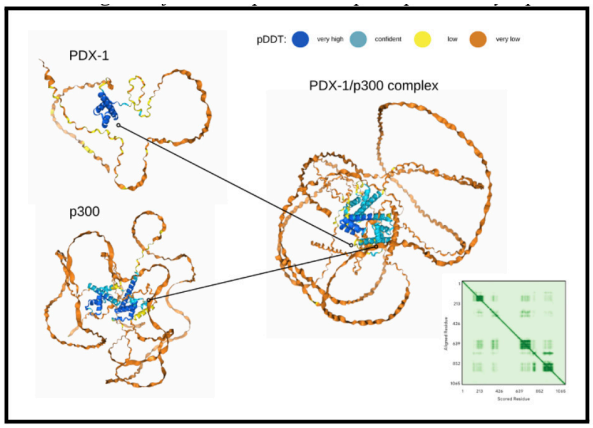
Figure 5

Structure alignment for PDX-1 protein with JMJD3, predicted by Alpha Fold III.



The color scale bar highlights the level of prediction accuracy.
Figure 6

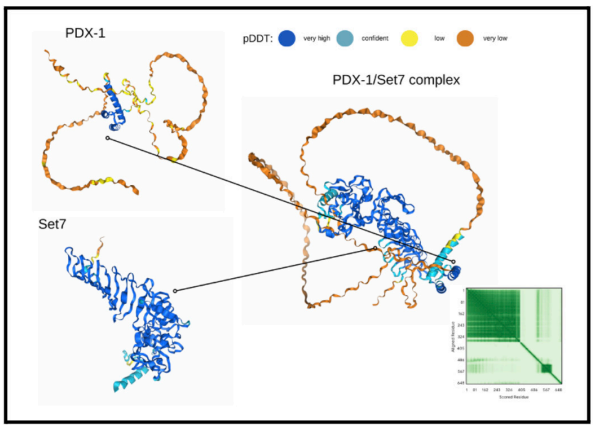
Structure alignment for PDX-1 protein with p300, predicted by Alpha Fold III



The color scale bar highlights the level of prediction accuracy.

Figure 7

Structure alignment for PDX-1 protein with Set7, predicted by Alpha Fold III



The color scale bar highlights the level of prediction accuracy.

Per-atom confidence estimates (pLDDT) indicate the accuracy of the predicted structure, with higher pLDDT values corresponding to more reliable predictions. Each sequence and protein residue is highlighted with a specific color corresponding to its pLDDT value. According to Jumper et al. (2021), a pTM value greater than 0.5 suggests that the prediction model likely resembles the true structure of the protein.

Among the three prediction results, the PDX-1/Set7 complex has the highest-scored residues, potentially due to the ordered structure of Set7. The low prediction accuracy in certain regions can be attributed to the intrinsically disordered regions (IDRs) of the complex. As shown in the figures, the regions located at the center of the complex highlight the interaction between the alpha helix structure of PDX-1 and its co-enzymes. These regions are mostly ordered, as indicated by the blue color coding, specifically in the residue numbers 156-206, which represent the ordered N-terminus. This ordered structure may be necessary to stabilize the complex and reduce degradation, facilitating gene regulation.

The IDR regions of the three-protein complex are located on the exterior, enclosing the ordered region at the center. The exposure of the IDR regions to the environment suggests their significant role in transcriptional regulation, as these regions are the first to interact with chromosomes and other transcription factors. Unlike the ordered regions, IDR regions are loosely packed and occupy most of the complex's space, providing flexibility for different interactions, such as facilitating the binding of BetaA2 to the INS promoter for PDX-1 and depositing methylation marks for the p300/Set7 complex. However, low pLDDT values for the exterior layers of the complex indicate that the prediction result may be less accurate in these IDR regions. Thus, further studies are necessary to confirm the arrangement of polypeptide bonds at the exterior of the protein.

Conclusion

In summary, Type 2 Diabetes is an endocrine system disease that is characterized by the error in control and regulation of gene expression of insulin expression. According to reviews, regulation of insulin expressions is achieved mainly through the interaction between gene transcription factors. Review previous studies regarding the epigenetic control of T2D by PDX-1, the unique structure of PDX-1, that allows it to regulate the secretion of insulin through methylation and acetylation of chromosomes. Alpha fold prediction model for PDX-1/p300, PDX-1/Set7, and PDX-1/JDJ3 indicates that IDR regions located at the external of the complex, enclosing the ordered regions are the center. In addition, the IDR regions are lossy packed, which provides flexibility for the complex to adjust to the environment, potentially indicating the role of the IDR region in adjusting the conformation based on varied conditions. However, the prediction result using the Alpha fold model is only a simulation, thus the arrangement of the polypeptide bonds at the exterior of the protein may require further studies to confirm.

References

Babu, M. M. (2016). The molecular basis for cellular function of intrinsically disordered protein regions. Nature Reviews Molecular Cell Biology, 17(12), 746-754. https://doi.org/10.1038/nrm.2016.116

Covington, H. E., Maze, I., LaPlant, Q. C., Vialou, V. F., Ohnishi, Y. N., Berton, O., ... & Nestler, E. J. (2009). Antidepressant actions of histone deacetylase inhibitors. Journal of Neuroscience, 29(37), 11451- 11460. https://doi.org/10.1523/JNEUROSCI.1758-09.2009

Hartl, F. U., Lecker, S., Schiebel, E., Hendrick, J. P., & Wickner, W. (1994). The binding cascade of SecB to SecA to SecY/E mediates preprotein targeting to the E. coli plasma membrane. Cell, 79(8), 1129-1140. https://doi.org/10.1016/0092-8674(94)90012-4

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., ... & Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. Journal of Molecular Biology, 433(20), 167- 183. https://doi.org/10.1016/j.jmb.2021.06.002

Liu, Y., Li, S., Wang, X., Zhang, Y., & Chen, G. (2021). Epigenetic regulation of PDX-1 in Type 2 Diabetes Mellitus. Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy, 14, 165-175. https://pubmed.ncbi.nlm.nih.gov/33098197/

Schröder, M., López-García, J., & Sebastián, E. (2018). Applications of RMSD in structure prediction and protein docking. Molecules, 23(5), 1038. https://doi.org/10.3390/molecules23051038

Sharma, A., Kumar, R., & Ghosh, S. (2020). PDX-1 is essential for insulin gene transcription and is activated by glucose-stimulated calcium influx. Journal of Biological Chemistry, 295(50), 17029-17039. https://doi.org/10.1074/jbc.RA120.013423
Smith, A. G., & Saunders, R. A. (2019). Regulation of PDX-1 by growth factors: Mechanisms and implications. Journal of Biological Chemistry, 294(29), 11293-11305. https://doi.org/10.1074/jbc.RA119.008259

Tuttle, R. L., Briaud, I., & Shalev, A. (2020). Regulation of PDX-1 by glucose metabolism: The role of Akt and MAPK signaling pathways. Journal of Biological Chemistry, 295(24), 8109-8119. https://doi.org/10.1074/jbc.RA120.013423

Uversky, V. N. (2019). Intrinsically disordered proteins in cell communication and signaling. Cell Communication and Signaling, 17(1), 31. https://doi.org/10.1186/s12964-019-0354-1

Ward, J. J., & Jones, D. T. (2010). Intrinsically disordered proteins and regions predicted by PONDR-FIT. Nucleic Acids Research, 38(9), 1619-1622. https://doi.org/10.1093/nar/gkq479