Designing Inhibitors against *HOX* domain mutations of PDX-1 and studying its association in Diabetes

Allam Appa Rao[1], Anupam Bhattacharya[2], Amit Kumar[3]*, Amita Kashyap[4], GR Sridhar[5]

[1]Jawahar Lal Nehru Technological University
Kakinada, AP, India

[2]BioAxis DNA Research Centre Private Limited,
Lucknow, UP, India

[3,4] BioAxis DNA Research Centre Private Limited,
Hyderabad, AP, India

[5]EDC, Vishakapatnam

*Corresponding Author; Telefax: +9140-24034503

Abstract

Type 1 diabetes mellitus was formally known as IDDM, type I, or juvenile onset diabetes. Type 1 DM can occur at any age. In this study,we analyzed the involvement of HOX domain of PDX-1 protein.The homeodomain transcription factor, pancreas duodenum homeobox (PDX)-1, encoded by PDX-1 gene, which is a transcriptional activator of several genes, including insulin, somatostatin, glucokinase, islet amyloid polypeptide, and glucose transporter type 2 and essential for pancreas development, insulin production, and glucose homeostasis.[1,13]. HOX domain has a length of 63aa and control developmental patterns and cell differentiation in vertebrates by acting positive or negative regulators[4,9,16]. Different approached had been applied to identify the mutational hot spot region of HOX domain and calculate mutational frequency of the amino acids which resides in the hotspot region. Binding site of the domain had been identified and found that THR208, GLN246 ,VAL247, ASN253 involved in interaction with ligand. Potential Inhibitors had been screened on the basis of various criteria and bioactivity score had been calculated. Energy optimization was done by applying AMBER force field and steepest descent method. Docking was performed by CCDC GOLD, Molegro, HEX, and Argus lab to find the best potent inhibitor and increase the accuracy of the docking process. Sitagliptin showed satisfactory result on both docking and bioactivity analysis. It showed a GOLD fitness score of 49.8386 and had a moldock score of -122.919 with a ligand efficiency -4.33692. Compound had a bioactivity score of 0.56 for protease inhibitor. Sitagliptin showed good binding affinity to the target, which helps to work the pancreas in proper way and to secret insulin.

Keywords: *Diabetes ,IDDM, HOX domain regulation, PDX-1, Homeodomain function, docking, Disorder region*

. 1. INTRODUCTION

Pancreas duodenum homeobox-1 (PDX-1) (also known as insulin promoter factor-1, islet/duodenum homeobox-1, somatostatin transactivating factor-1, insulin upstream factor-1 and glucose-sensitive factor) is a transcription factor encoded by a Hox-like homeodomain gene. In humans and other animal species, the embryonic development of the pancreas requires PDX-1, as demonstrated by the identification of an individual with pancreatic agenesis resulting from a mutation that impaired the transcription of a functionally active PDX-1 protein. In adult subjects, PDX-1 is essential for normal pancreatic islet function as suggested by its regulatory action on the expression of a number of pancreatic genes, including insulin, somatostatin, islet amyloid polypeptide.[24] We focus on HOX domain of PDX-1 protein because it is evolutionarily conserved which containing transcription factors that activate and repress gene expression in a precise temporally and spatially regulated manner during development and differentiation. Pancreatic-duodenal homeobox 1 (PDX-1) is a Hox-type protein that is a critical requirement for normal pancreas development and for proper differentiation of the endocrine pancreas. HOX factors are evolutionarily conserved homeodomain-containing transcription factors that are critical for the determination of the body plan during early development and for the regulation of key differentiation-specific genes that determine cellular specificity. Proper spatial and temporal expression of HOX

factors is required to orchestrate anterior-posterior segmentation, limb development, and organogenesis. HOX proteins regulate gene transcription by a well-coordinated and tightly regulated balance of gene activation and repression[18,20].

The HOX domain is 63 amino acids in length, the position of HOX domain in pancreas duodenum homeobox (PDX)-1 started from 146 and ends with 209 residue of amino acid.

Human mutation in the A chain of the PDX-1 protein are associated with the type 1diabetes mellitus, emphasizing the functional importance of this region, the available data support that the domain contain disorder region which are responsible for improper development of pancreases or responsible for β-cells not to secrete insulin, if insulin is secreted it don't function properly and this lead to the disease in the early age. In our work we predict the tertiary structure of the protein (pancreas duodenum homeobox) and found the fold recognition that is functional region of the protein, next we found the disorder region in the domain which is responsible for disease, by tools (RONN, Disopred, and DISPROT, DisEMBL, GLOBPLOT 2).and search for the binding site by (Q site finder, pocket finder, castp) to bind the drug compound at particular amino acid by docking software's like (Argus lab, Hex, Molegro virtual docker, and GOLD).

In this report we identify a single amino acid which lies in the (pancreas duodenum homeobox (PDX)-1) in HOX domain range which have mutation, by this we can say that this amino acid is responsible for improper development of pancreas and decrease the PDX-1 which is necessary for β-cell maturation which produce insulin.

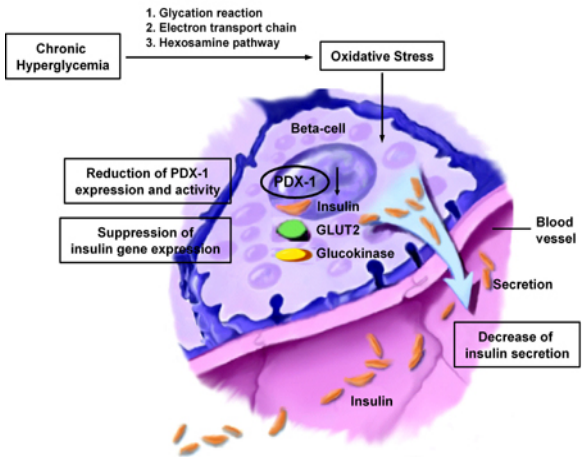

Fig1. Figure showing the position and regulation of PDX-1 (Ref:Kaneto H et al. Involvement of oxidative stress and the JNK pathway in glucose toxicity. Rev Diabet Stud. 2004; 1(4): 165-74.)

2.MATERIALS AND METHODS

2.1 Disorder prediction

Disorder region in the protein sequence were predicted by the tools of bioinformatics like (RONN, Disopred, DISPROT, DisEMBL, and GLOBPLOT 2). Which show the disorder or mutational region in the protein sequences.

2.2 Binding site and Active site identification

The binding site or active site identification is the method of finding the site where the ligand binds, for binding site prediction we use three tools (Q site finder, pocket finder, and Castp) by which we select the best binding site for the drug.

2.3Energy Minimization

Energy minimization was carried out of both receptor and ligand molecule. Energy minimization was done by using AMBER force field and applying steepest descent method to get the optimal energy of the molecules. The energy minimization of ligand was done by using Tripos force field.

2.4 Docking

Docking is a method which predicts the preferred orientation of one molecule to a record when bound to each other to form stable complex knowledge of the preferred orientations in turn may be used to predict the binding strength of association or binding affinity between two molecules. Docking is frequently used to predict the binding orientations of small molecules drug candidates to protein targets in order to in turn predict the affinity and activity of the small molecule. The receiving molecule that primarily binds to a small molecule or another protein or a nucleic acid called receptor. A molecule that forms the complementary partner in the docking process called ligand, we used several software for docking i.e. Arguslab, molegro, Hex, gold.

3. RESULTS AND DISCUSSION

3.1 Sequence retrieval

The protein sequence was retrieve from the GeneCard by selecting the sequences from whole involved sequences in the diabetes mellitus type 1 disease these sequences are screened and we fund the sequence which is playing the major role in disease i.e. pancreas duodenum homeobox (PDX)-1.

3.2 Domain Analysis

Domain identification was done by using bioinformatics database (SMART) and found HOX domain which is 63 amino acid residues long, the domain within the protein sequence starts at position 146 and ends at position 208.

3.3 Disorder Prediction

Disorder prediction in the protein was done by the tools (RONN, Disopred, DISPROT, DisEMBL, and GLOBPLOT 2). The results show the major disorder region in the protein sequence, the disorder region which lies in the domain range were selected for docking the candidate drug.

3.4 RONN

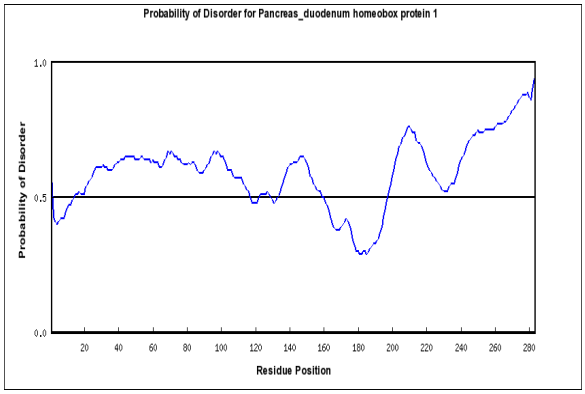

Fig 2. The graph show the higher peak of disorder region in the protein sequence at 207 amino acid residue which lies between the domain ranges.

3.5 DisEMBL

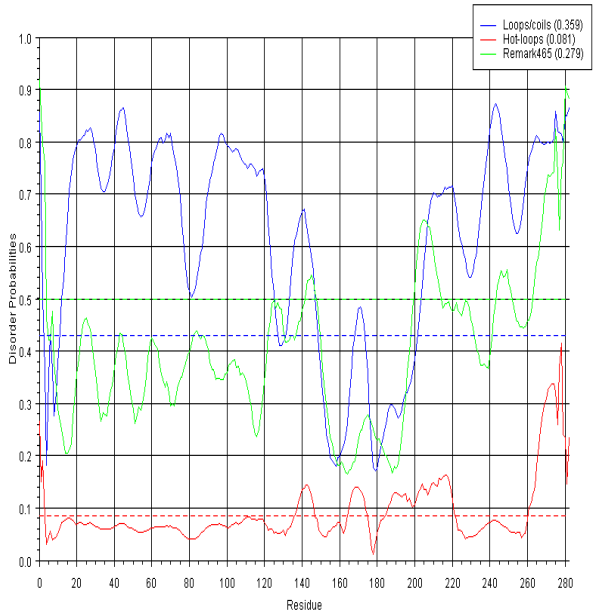

Fig.3. In above graph we see the higher peak which comes in domain range this shows that the region has higher disorder in the protein sequence.

3.6 Disopred

This is the tool which is used for finding the disordered region in the protein sequence.

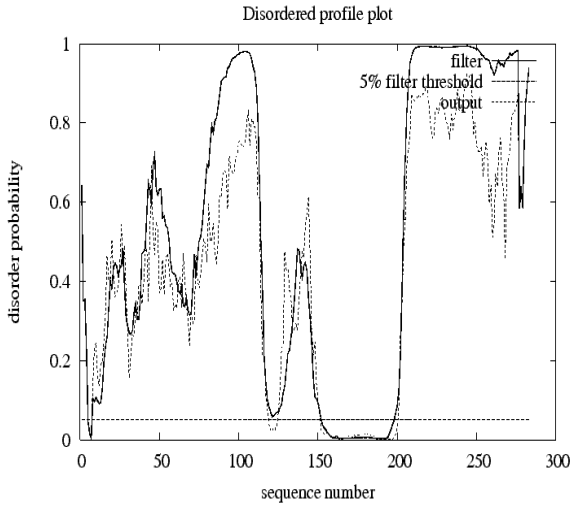

Fig.4. The graph shows the higher peak of disorder region in the protein sequence.

3.7 Binding site and Active site identification

The binding site and active site identification is the method of finding the site where the ligand binds. Using bioinformatics tools and approach we found the site which is good for ligand to bind and it comes in the domain region which concludes that the mutation in that site is responsible for type 1 diabetes mellitus, tools we used for this aaproach (Q site finder, pocket finder). All three tools identified the same binding site ARG (arginine) at position 207 in the protein sequence. This amino acid residue is targeted with the drug compound to cure mutation.

3.8 Energy minimization of target

Energy minimization was carried out by using sybyl software. AMBER force filed was applied and energy minimization was done by steepest descent method.

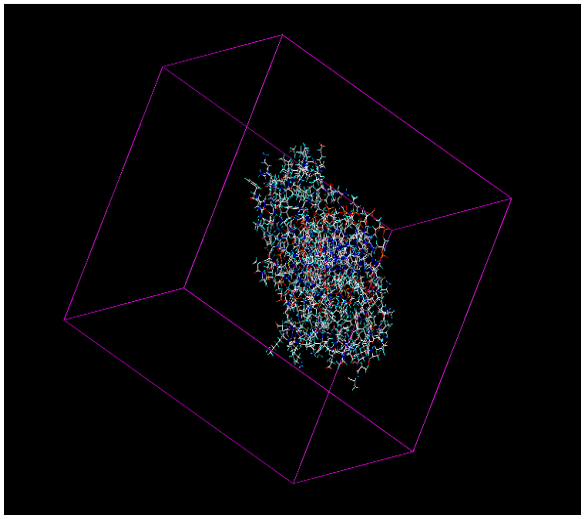

Fig 5. Recptor molecule inside a periodic bundary condition.

Initial energy of the target:

Table 1: Energy Breakdown of the Intial Energy calculation of Receptor Molecule.

Bond Stretching Energy	199.703
Angle Bending Energy	340.456
Torsional Energy	942.007
ImproperTorsional Energy	0.938
1-4 van der Waals Energy	1063.472
van der Waals Energy	27.774
1-4 Electrostatic Energy	261.677
Electrostatic Energy	-4494.384
Total Energy	-1658.357 kcals/mol

Optimized Energy of the Target:

Table 2: Energy breakdown of the optimized energy of receptor

Bond Stretching Energy	128.966
Angle Bending Energy	348.500
Torsional Energy	928.180
Improper Torsional Energy	8.524
1-4 van der Waals Energy	808.579
van der Waals Energy	-1661.864
1-4 Electrostatic Energy	236.970
Electrostatic Energy	-6343.784
Total Energy	-5545.928 kcals/mol

Initial Energy of Ligand:

Force field applied: Tripos

Table 3: Energy Breakdown of the Initial Energy of the Ligand Molecule.

Bond Stretching Energy	0.022
Angle Bending Energy	0.000
Torsional Energy	0.000
Out of Plane Bending Energy	0.000
1-4 van der Waals Energy	0.000
van der Waals Energy	348916573.709
1-4 Electrostatic Energy	0.000
Electrostatic Energy	552.743
Total Energy	348917126.474 kcals/mol

Optimized Energy of the Ligand:

Table 4: showing the optimized energy of the ligand.

Bond Stretching Energy	0.000
Angle Bending Energy	0.000
Torsional Energy	0.000
Out of Plane Bending Energy	0.000
1-4 van der Waals Energy	0.000
van der Waals Energy	-3.823
1-4 Electrostatic Energy	0.000
Electrostatic Energy	0.000
Total Energy	-3.823 kcals/mol

3.9 Docking Result

Docking is a method which predicts the preferred orientation of one molecule to a second when bound to each other to form a stable complex, Applying different approaches and software of bioinformatics like (Argus lab, HEX, Molegro Virtual docker, GOLD) we use to dock the inhibitor compounds to the protein binding site where the compound binds properly and show the result.

Argus lab result:

Table 5: Showing the different poses of ligand with docking energy in Argus lab.

Poses	Energy (kcal/mole)
Pose 1	-7.66
Pose 2	-7.05
Pose 3	-6.09
Pose 4	-5.56

Hex Result:

Ligand	ETOTAL	ESHAPE	EFORCE
Sitagliptin	-263.12	-263.12	0.00

Table 6: The above table showing the docking result performed by Hex software. The inhibitor having the docking energy of -263.12 kcal/mole.

GOLD result:

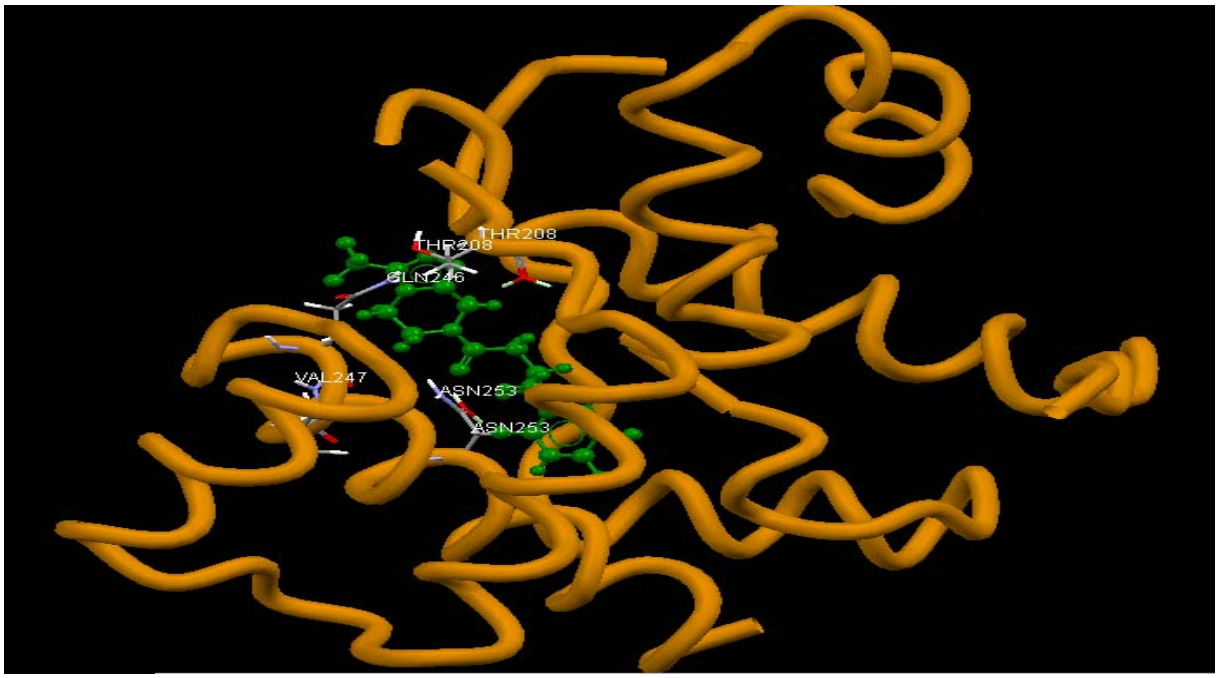

Fig. 6. Above figure showing the hydrogen bond interaction and short contacts between ligand and the target, involving the interaction between the amino acids i.e. (THR208, GLN246 ,VAL247, ASN253).

Gold fitness score:
The ligand and target showing the GOLD fitness score of 49.8386 which show best fitting to the target region.

Molegro Pose Results

Table 7: The following table shows the MolDock score, Rerank score, ligand Effiency (LE1) and Hbond interaction between the ligand and target

.

s.no	MOL DOC SCORE(GRID)	MOL DOC SCORE	RERANK	TORSION	Heavy Atoms	LEI	HBond
1	-123.612	-122.919	-61.9012	5	28	-4.33692	-1.99759
2	-118.504	-111.965	-14.6408	5	28	-4.17575	-0.701358
3	-117.883	-113.137	-79.1273	5	28	-4.18887	-0.826193
4	-116.154	-109.609	-74.0989	5	28	-4.16604	-5.13766
5	-113.403	-105.752	-71.7545	5	28	-3.95598	-3.68488

Ligand Information:

Table 8: Different properties of ligand. CID: compound Id, MW: Molecular weight(Dalton), HD:H-bond donor, HA: H-bond acceptor, RN: ring number, TPSA: Topological polar surface area, RotB: number of rotatable bonds

CID	MW	HD	HA	HATOM	RN	Logp	TPSA	volume	RotB
11476287	443.77	2	8	29	3	1.5	77.05	311.65	5

Bioactivity of ligand:

Table 9: Table showing the bioactivity of the ligand.

The bioactivity result showed that the ligand is a protease inhibitor and pdx-1 protein belongs to protease family. So it might be a potent compound to block the action of Hox domain of PDX-1.

GPCR Ligand	0.25
Ion Channel modulator	-0.27
Kinase inhibitor	0.01
Nuclear Receptor ligand	-0.60
Protease Inhibitor	0.56
Enzyme Inhibitor	-0.06

4. CONCLUSION

In the study of homeodomain transcription factor, pancreas duodenum homeobox (PDX)-1 we have studied that it is essential for pancreas development, insulin production, and glucose homeostasis, the Hox domain found in the protein sequence is a critical requirement for normal pancreas development and for proper differentiation of the endocrine pancreas and β-cell maturation, regulation and maturation of pancreas are likely to be important for the insulin production and glucose homeostasis in human blood, in our study we see that the HOX domain which lie in

PDX-1 protein is responsible for proteins control, developmental patterns and cell differentiation in vertebrates which help in pancreas development.

Here we identified the mutational region in the protein sequence which is responsible for improper maturation of pancreas. Further we find the binding region, the region to which our drug bind and found the single amino acid arginine (ARG) which is at 207 position in protein sequence.

while docking we target the domain mutational region in which the amino acid ARG is present to that we bind the drug and observed that the drug **Sitagliptin** has good hydrophobic effect based on LogP value which bind with target sequence showing the less energy, this show that the drug is best to correct the disorder in the protein sequence, the docking result were identified by different docking software like GOLD, Argus lab, molegro, Hex. Finally we conclude that the drug is showing the higher binding affinity between the ligand and protein so we can say that this drug is best to correct the mutation or to increase the insulin production in the pancreas for type 1 diabetes mellitus patients.

Acknowledgment

This project was solely funded by Department of Science and Technology, Govt of India under IRHPA Scheme.

References

[1] Bonner-Weir S, Baxter LA, Schuppin GT & Smith FE 1993 A second pathway for regeneration of adult exocrine and endocrine pancreas. A possible recapitulation of embryonic development. *Diabetes* **42** 1715–1720.

[2] Characterization of Pancreatic Transcription Factor Pdx-1 Binding Sites Using Promoter Microarray and Serial Analysis of Chromatin Occupancy.David M Keller ,Athanasios Arsenlis ,Jacques Drouin , Christopher V.E. Wright , Haiyan Wang,Claes B. Wollheim ,Peter White, Klaus H. Kaestner and Richard H. Goodman.

[3] Dobson, M. (1776). "Nature of the urine in diabetes". *Medical Observations and Inquiries* **5**: 298–310.

[4] Dor Y, Brown J, Martinez OI & Melton DA 2004 Adult pancreatic beta-cells are formed by self-duplication rather than stem-cell differentiation. *Nature* **429** 41–46.

[5] Glick E, Leshkowitz D & Walker MD. Transcription factor BETA2 acts cooperatively with E2A and PDX1 to activate the insulin gene promoter. Journal of Biological Chemistry 2000 2752199–2204.

[6] GlobPlot: exploring protein sequences for globularity and disorder. Run Linding. Robert B. Russell. Victor Neduva and Toby J.Gibson Received February 14, 2003; Revised and Accepted March 20, 2003

[7] Harris MI, Flegal KM, Cowie CC, *et al.* (April 1998). "Prevalence of diabetes, impaired fasting glucose, and impaired glucose tolerance in U.S. adults. The Third National Health and Nutrition Examination Survey, 1988–1994". *Diabetes Care* **21** (4): 518–24

[8] Hideaki Kaneto, Yoshihisa Nakatani, Dan Kawamori,Takeshi Miyatsuka, and Taka-aki Matsuoka. Involvement of Oxidative Stress and the JNK Pathway in Glucose Toxicity Rev Diabet Stud, 2004, 1(4):165-174.

[9]Iimura T, Pourquie O. Hox genes in time and space during vertebrate body formation. Dev Growth Differ 2007;49(4):265-75

[10] Jennifer M. Oliver-Krasinski[1,2], Margaret T. Kasner[2], Juxiang Yang[1,2], Michael F. Crutchlow[1,2], Anil K. Rustgi[2,3], Klaus H. Kaestner[1,3] and Doris A. Stoffers[1,2.] The diabetes gene *Pdx1* regulates the transcriptional network of pancreatic endocrine progenitor cells in mice.

[11] Jonathan J. Ward, Liam J. McGuffin, Kevin Bryson, Bernard F. Buxton and David T. Jones. The DISOPRED server for the prediction of protein disorder, Received on *February 19, 2004; revised and accepted on March 20, 2004 Advance Access publication March 25, 2004.*

[12] Leonard J, Peers B, Johnson T, Ferreri K, Lee S, Montminy MR 1993 Characterization of somatostatin transactivating factor-1, a novel homeobox factor that stimulates somatostatin expression in pancreatic islet cells. Mol Endocrinol 7:1275–1283

[13] Marie-Jose´ e Boucher, Me´ lanie Simoneau, and Helena Edlund. The Homeodomain-Interacting Protein Kinase 2 Regulates Insulin Promoter Factor-1/Pancreatic Duodenal Homeobox-1Transcriptional Activity

[14] Nathan DM, Cleary PA, Backlund JY, *et al.* (December 2005). "Intensive diabetes treatment and cardiovascular disease in patients with type 1 diabetes". *The New England Journal of Medicine* **353** (25): 2643–53.

[15] Piper DE, Batchelor AH, Chang CP, Cleary ML, Wolberger C. Structure of a HoxB1-Pbx1 heterodimer bound to DNA: role of the hexapeptide and a fourth homeodomain helix in complex formation. Cell 1999; 96:587-97. [PMID 10052460].

[16] Qian J, Kay EN, Towle HC & Olson LK. Upstream stimulatory factor regulates Pdx-1 gene expression in differentiated pancreatic b-cells Biochemical Journal 1999 341 315–322.

[17] Qiu Y, Guo M, Huang S & Stein R 2002 Insulin gene transcription is mediated by interactions between the p300 coactivator and PDX-1, BETA2, and E47. *Molecular and Cellular Biology* **22** 412–420.

[18] Ryan EA, Lakey JR, Paty BW, Imes S, Korbutt GS, Kneteman NM, Bigam D, Rajotte RV & Shapiro AM 2002 Successful islet transplantation: continued insulin reserve provides long-term glycemic control. *Diabetes* **51** 2148–2157.

[19] Scott MP. A rational nomenclature for vertebrate homeobox (HOX) genes. Nucleic Acids Res 1993;21(8):1687-8.

[20] Sophie Gray1, Hardev S Pandha1, Agnieszka Michael1, Gary Middleton2, Richard Morgan *HOX* Genes in Pancreatic

Development and Cancer11Postgraduate Medical School, Faculty of Health and Medical Sciences, University of Surrey;2Royal Surrey County Hospital NHS Trust. Guildford, United Kingdom.

[21] Watada H, Kajimoto Y, Kaneto H, *et al.* (1997). "Involvement of the homeodomain-containing transcription factor PDX-1 in islet amyloid polypeptide gene transcription.". *Biochem. Biophys. Res. Commun.* **229** (3): 746–51.

[22] Ying Liu, Raymond J. MacDonald and Galvin H. Swift. DNA Binding and Transcriptional Activation by a PDX1·PBX1b·MEIS2b Trimer and Cooperation with a Pancreas-specific Basic Helix-Loop-Helix Complex.

[23] Zheng Rong Yang1, Rebecca Thomson2, Philip McNeil3 and Robert M. Esnouf2. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Received on July 9, 2004; revised on June 7, 2005; accepted on June 8, 2005 Advance Access publication June 9, 2005.*

[24] sHongxiang Hui1 and Riccardo Perfetti1, Pancreas duodenum homeobox-1 regulates pancreas development during embryogenesis and islet cell function in adulthood. Division of Diabetes, Endocrinology and Metabolism, Cedars-Sinai Medical Center and 2University of California, Los Angeles, California, USA (Correspondence should be addressed to R Perfetti, Division of Endocrinology and Metabolism, Cedars-Sinai Medical Center, 8723 Alden Drive, SSB #290, Los Angeles, CA 90048, USA; Email: perfettir@cshs.org)