Contents lists available at UGC-CARE

International Journal of Pharmaceutical Sciences and Drug Research

[ISSN: 0975-248X; CODEN (USA): IJPSPP]

Available online at www.ijpsdronline.com

Research Article

Deciphering FFAR4 Mediated Functional Pathways and Potential Drug Targets in Diabetes Mellitus: An Integrated Protein Interaction Network Analysis and Kinetic Simulation

Divya Jhinjharia, Samvedna Singh, Shakti Sahi*

School of Biotechnology, Gautam Buddha University, Greater Noida, Uttar Pradesh, India.

ARTICLE INFO

Article history:

Received: 22 March, 2024

Revised: 18 April, 2024

Accepted: 23 April, 2024

Published: 30 May, 2024

Keywords:

Free fatty acid receptor 4, Regulatory network, Drug target, Diabetes, Kinetic simulation.

DOI:

10.25004/IJPSDR.2024.160313

ABSTRACT

Type 2 diabetes is a metabolic disorder that affects people worldwide.The G-protein coupled receptor (GPCR) known as free fatty acid receptor 4 (FFAR4) has been shown to be a potential therapeutic target for type 2 diabetes mellitus (T2DM) and complications linked to obesity. The present study focuses on the pharmaceutical role of FFAR4 and its potential agonists by predicting anti-diabetic responses, including insulin secretion, glucose uptake and calcium ion concentration levels. We identified differentially expressed genes and elucidated their role extensively through analysis of pathways, molecular mechanisms and linked biological processes. In the present study, a systems biology approach was implemented to establish an interaction network between FFAR4 and its driver such as CASR and NR1H4, that highlighted their significance as potential prognostic and therapeutic targets. A mathematical model incorporating biological events mediated by these proteins is studied and simulated using kinetics law reactions. Furthermore, the kinetic simulations were conducted to assess the impact of drug molecules, namely comp35, comp50, compN1 and compN2, on FFAR4 function. The findings reveal FFAR4's potential as a therapeutic target for the treatment of type 2 diabetes mellitus.

Introduction

Type 2 diabetes mellitus (T2DM), a metabolic disease, is the leading cause of death and morbidity worldwide. It affects around 537 million people aged 20 to 79.[1] An in-depth understanding of the disease's underlying mechanism is crucial for developing better and highly effective therapy. Metformin, thiazolidinediones, sulfonylureas, glides, and DPP-4 inhibitors are used for therapy. Newer therapeutics, such as sodium-glucose cotransporter-2 (SGLT2) inhibitors, lower blood glucose levels by decreasing glucose reabsorption in the kidney, and glucagon-like peptide-1 receptor agonist (GLP-1) lower the rate of digestion by enhancing GLP-1's functional role, lowering blood sugar levels. Essential biochemical alterations in T2DM include decreased glucose tolerance, hyperglycemia, and glycosuria. Most of these drugs affect blood glucose control, avoiding hyperglycemia. However, these drugs show gastrointestinal side effects, increased cases of acute pancreatitis, cardiovascular disorders, and kidney failure. Therefore, it is imperative to explore newer targets for better therapeutic interventions.

GPR120 or free fatty acid receptor 4 (FFAR4), regulates whole-body metabolic homeostasis. The nutritional changes regulate energy balance and can lead to certain metabolic disorders, including diabetes, cardiovascular diseases and obesity. The absence or low levels of insulin results in an imbalance of glucose utilization by affecting energy production pathways. It leads to fat mobilization from adipose tissue and accumulating higher cholesterol levels, triglycerides (TG), free fatty acids, lipoproteins, and ketone bodies. Various studies have shown that FFAR4 is a potential drug target for controlling blood glucose

***Corresponding Author:** Dr. Shakti Sahi

Address: School of Biotechnology, Gautam Buddha University, Greater Noida, Uttar Pradesh, India.

Email ✉**:** shaktis@gbu.ac.in

Tel.: +91-120-2344275

Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

levels and enhancing tissue insulin sensitivity.[2] Higher expression of GPR120 in enteroendocrine cells elucidates its pharmacological activity. Activation of GPR120 enhances GLP1 secretion, which validates its role in insulin secretion in the pancreas. Identifying and classifying synthetic ligands associated with FFAR4 is integral to a better understanding of biological function. Some of the ligands reported for FFAR4, including NCG21, TUG-891, and GW9508, overlap with those for FFAR1 since they share structural similarities.[3] A selective agonist of FFAR4 has shown improvement in the tolerance of glucose levels, decreased hyperinsulinemia, and increased sensitivity to insulin in obese mice when fed on a high-fat diet,[4] clinical trials for the natural and synthetic ligands for FFAR4 are progressing for diabetes and obesity. Hence, a better understanding of pharmacokinetic and pharmacodynamic properties is vital for further development.

High-throughput methods facilitate understanding of the biological role of the FFAR4 receptor, discovering high-affinity synthetic ligands and studying their pharmacological aspects. Gene expression data based computational techniques can characterize genetic modifications at the genome level, allowing to study of differential gene expression and their potential physiological or pathological significance. This study uses different network analysis approaches to construct and analyze the biological network of FFAR4, highlighting its association with key drivers in the disease. These novel drivers can help in prognosis, detection as well as disease prevention. Various diseases, including cancers, focus on driver mechanisms such as epigenetic regulators, mutations, transcription factors and gene regulation for generating therapy and personalized treatment opportunities.[5,6] Our findings reveal the crucial links between disease progression and novel drivers. Furthermore, we have extensively studied the role of FFAR4 in diabetes and diabetes-linked disorders, which not only validates the potential of FFAR4 to be used as a therapeutic target for T2DM. It also provides insight into the disease's mechanisms and can enable the development of newer treatment opportunities for T2DM. Additionally, FFAR4 agonists can help in disease management; therefore, to establish this finding, we also performed kinetic simulations and monitored anti-diabetic responses of potential hit compounds.

Materials And Methods

Data Retrieval and Differential Gene Expression Analysis

In the present study, an analysis of differential gene expression was performed on GEO datasets from the GEO-NCBI database (https://www.ncbi.nlm.nih.gov/geo/) GSE118139, GSE166502, GSE156993, and GSE166467 with samples of human disease patients (Type 2 Diabetes) and control samples of healthy tissues. Normalization was performed to remove the batch effects of the datasets. This analysis to calculate differential gene calculations was done in control versus diseased groups using the R package, Limma (https://www.bioconductor.org/packages/release/bioc/html/limma.html).[7] Benjamini & Hochberg's false discovery and Limma precision weights were performed to calculate DEGs with significant fold change. The threshold to consider DEGs was logarithmic fold change >1.0 and *p-value* <0.05.

Gene Ontology and Pathway Studies

Enrichment studies were carried out to understand the biological characteristics of gene sets and genes for high throughput transcriptome and genome data. DAVID (https://david.ncifcrf.gov/), a Database for Annotation, Visualization, and Integrated Discovery, was used to understand the biological significance of genes. Biological processes, cellular components, and molecular functions of genes were studied. The Kyoto Encyclopedia of Genes and Genomes (KEGG) [8] database was used to study enriched pathways systematically.

Protein-protein interaction

Protein-protein interaction analysis was performed. The search tool for the retrieval of interacting genes (STRING, version 11.5; string-db.org/), currently covering 24'584'628 proteins from 5,090 organisms,[9] was used to construct the PPI network. STRING associations with a minimum confidence score of 0.4 were used to construct relationships among proteins. Multiple proteins of significant differential expression from the aforementioned analysis were used to build a network of human PPI. Genomic context predictions, automated text mining, (conserved) co-expression, high-throughput lab experiments and literature were considered fundamental principles for determining the interaction between proteins. Interacting proteins and hub protein networks were analyzed based on the following criteria. Topology and characterization of proteins: The protein-protein interaction network is characterized based on key topological parameters. These parameters are crucial in determining the connectivity between the nodes to understand their biological significance of interaction. Network Analyzer in Cytoscape was used to construct a topological property-based interaction network. The node degree of a protein in the network was calculated based on the number of interactions it has with other nodes or proteins. Probabilistic degree distribution was calculated throughout the whole network.

Identification of sub-network and hub proteins

The topologically distributed PPI network was further studied to identify crucial interactions, sub-networks, and hub proteins contributing to the network. An approach to determine sub-networks or highly contributing clusters

from the complete network of 5332 proteins is performed using molecular complex detection (MCODE).[10] Cytoscape (https://cytoscape.org/), an open-source application for integrating data and visualization of intricate networks and integrating data into many other attributes, employed MCODE-based clustering. The sub-networks or clusters in the complex networking shared by proteins and their neighbors were identified in a three-step process: (a) weighting- a score is given to more interconnected nodes; (b) molecular complex prediction- starting with the highest weighted node, also known as a seed, iteratively move out, adding more nodes that have a specific value above a threshold and (c) post-processing- applies filters, thus improving cluster quality.

In addition to this, an approach to identify bottleneck nodes, CytoHubba was used. These networks were analyzed on edge percolated component, maximum neighborhood component, bottleneck, maximal clique centrality and stress-based on shortest paths, which gave us highly contributing proteins (hub proteins). Furthermore, enrichment studies of selected MCODE and CytoHubba clusters were studied with the attributes of biological processes, molecular function, and cellular compartments.

FFAR network construction and kinetic simulations

We constructed another protein interaction network keeping FFAR4 as a seed to understand better the essential drivers' association and the disease progression's biological mechanism. We performed topological analysis to identify and classify related and novel hubs, elucidating the essential link between FFAR4 and critical drivers of T2DM. Enrichment studies identified the interacting drivers and pathways associated with strengthening our understanding of hubs and their underlying mechanisms. Based on these studies, we constructed a biological pathway of the FFAR4 signaling mechanism in the diabetic pathway using the topological PPI as a frame, revealing the crucial linkage in disease progression and the importance of FFAR4 as a drug target in diabetes. For kinetic simulations, firstly, with the help of a process diagram editor called Cell Designer,[11] we drew the biochemical network comprising the signaling pathways associated with FFAR4 and its drivers. We divided the components of this network into different compartments like the nucleus, cytoplasm, macrophages, and cells. Secondly, to monitor the effect of kinetic reactions concerning simulation time, we used COPASI, a software application, to study biochemical network dynamics.[12]

Results

Identification of DEGs in Diabetes Mellitus

Differential gene expression analysis of GSE118139,[13] GSE166502, GSE166467,[14] and GSE156993[15] was done using R. Preprocessing and normalization of data files allowed us to remove batch effects and redundancy across the data. Normalized data were then processed for FDR calculation using Limma package in R. Top tables for each dataset were analyzed and significant genes with less than a 0.05 *p-value* and differential expression in terms of fold change>1.0 for upregulation and <-1.0 for downregulation were selected for further analysis.

Screening of genes and their corresponding proteins was performed. In all, 7041 DEGs were selected for further analysis. About 3880 genes were upregulated, while 3161 were downregulated. The volcano plots for each dataset

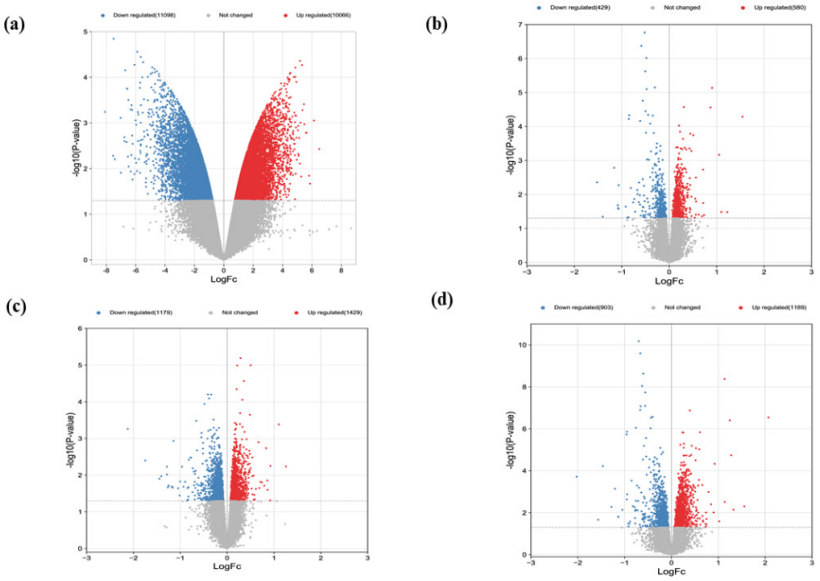

Fig. 1: Volcano plots depicting the upregulated (red) and downregulated (blue) genes in four datasets, (a) GSE118139, (b) GSE166502, (c) GSE156993, (d) GSE166467, grey color depicts the genes which fall below the threshold value, i.e., P value (less than 0.05) and logFC (-1 < 0 < +1)

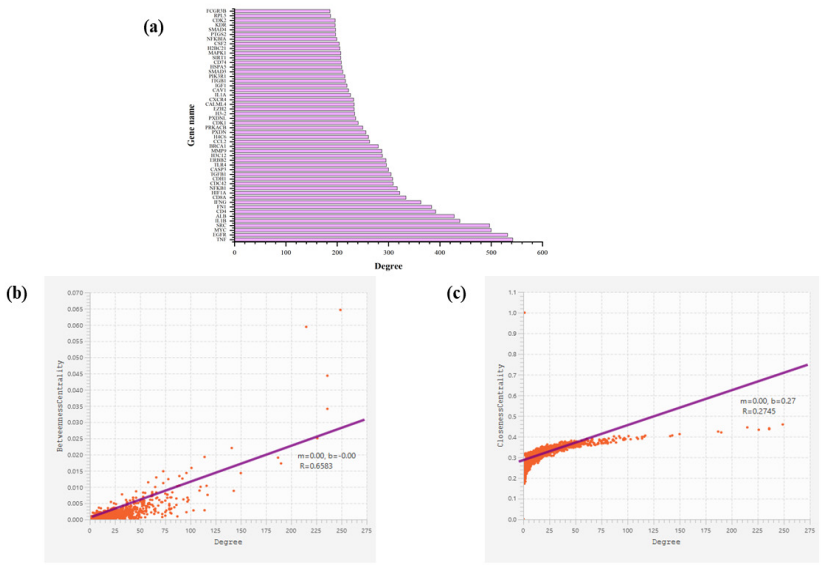

Fig. 2: (a) Top 50 nodes of the network ranked based on the number of node degrees. TNF, IL1B, EGFR and FN1 depicted in bar graphs are the highest degree nodes showing their significance in the protein-protein interaction network and (b) the behavior of betweenness centrality and (c) closeness centrality as a function of degree for diabetic pathway

with significant (P value) versus magnitude of change (fold change) were plotted. The number of upregulated genes (red) is observed on the right. In contrast, the downregulated ones (blue) can be visualized on the left, and the most statistically significant genes are placed toward the top of the plot based on their fold change value (Fig. 1).

Enrichment and Ontology of DEGs

After the compilation of the genes that were differentially expressed genes, we evaluated genes according to biological processes, cellular components, molecular functions, and biological pathways. The main molecular functions associated with the DEGs were phospholipid binding, G-protein coupled receptor activity, transmembrane transporter activity, antigen binding, signaling receptor binding and olfactory receptor binding. The GO terms based on biological processes ranked by fold enrichment were regulation of signal receptor activity, positive regulation of inflammatory responses (IL1B, FFAR3, FFAR4, NR1H4, IL1A, NLRP3), positive regulation to kinase activity, homeostasis process (FFAR3, FFAR4), positive regulation of protein metabolic processes (PCSK1, 1L1R2, CASP3) and calcium-mediated signaling (TNF, EDN1, CD4CCR7). Also, the pathways associated were the TNF-signaling pathway, pathways in cancer, insulin secretion (INSR, MAPK1, TNF, SOCS1, MAFA, IRS2), P13-Akt signaling pathway, Type 2 diabetes mellitus (MAFA, TNF, SOCS3, INSR, KCNJ11), MAPK signaling pathway (VEGFC, TNF, NLK, PGF, TGFB2), gastric cancer (WNT2, WNT4, TGFB1, CDH1), calcium signaling pathway (PLCB3, VEGFC, HGF, EGFR), cAMP signaling (ADCY5, MAPK1, GCG, PLN), MAPK signaling, pancreatic secretion, glucagon signaling and pancreatic cancer (SMAD3, SMAD4, EGFR, TGFB1, VEGFA).

Protein-protein Interaction Network Analysis

Examining networks is helpful since it allows one to investigate the connections among various proteins. To find critical protein interactions and biological modules involved in type 2 diabetes. A network of protein interactions between proteins was created for differentially expressed genes was constructed. A protein-protein interaction (PPI) of 5332 proteins corresponding to DEGs was built using the STRING plugin in Cytoscape. The interaction network with a combined score >0.4 was set as the cut-off criterion. Owing to the vast number of proteins in the PPI network, several cluster forms were identified in the parent network. There are less than six steps between two nodes, also known as the shortest path. We also observed “scale-free networks” in which most nodes are connected to a low number of neighbors and a small number of hubs that provide high connectivity to the network. The network was analyzed based on its topological properties.

The topological analysis implemented included: 1) Centrality analysis which estimates how vital a node or edge is for connectivity, and the nodes with high clustering values are considered more biologically significant. The centrality measured the degree of nodes and global centrality of the complete network. The network was analyzed considering both closeness centrality and betweenness centrality. Based on these parameters, tumor necrosis factor (TNF) (542), epidermal growth factor (EGFR) (532), MYC proto-oncogeneMYC (500), SRC proto-oncogeneSRC (497), interleukin 1 beta IL1B (439), and fibronectin-1 (FN1)(384) were observed with high degree and centrality (Fig.2a). 2) The measure known as closeness centrality, or CC, quantifies the potential flow at

Table 1: Protein-protein interaction clusters of diabetic-associated proteins along with their predicted cluster score

Clusters	*Score*	*Proteins*
1	26.838	*ADIPOQ, LEP, IGF2, SMAD3, SMAD4, EDN1, VEGFA, ILIA, FN1*
2	22.13	*IFNG, IL1B, KCNA5, TNF, SOCS1, C3, BRAF, CASP3, EGFR, NLRP3, MYC, C3*
3	13.768	*ADCY1, ADCY2, ADCY5, GCG, PLCB3, PRKCG, SHC1, APC, BCAR1, ALB, SRC, MMP9*
4	9.496	*IRS1, PFKM, PSMD9, RBP4, GRIK3, GNAS, GNAO1*
5	6.145	*RGS5, ABCC8, KCNN4, TRPM4, PTPRN, CPE, G6PC2, GNG2, BRCA1*
6	5.204	*SOCS4, PIK3R1, MAPK10, SNAP25, SYT7, GCNT1, MAPK12, CAMK2*
7	5.156	*INSR, GYS1, CD36, NCF2, PIK3CG, RASA1, GBE1, ARF1, SHC3, GRM4, TMPRSS6, GRK1, KCNJ12*
8	4.6	*PFKFB1, ARV1, UGT8, B4GALT6, GBA, GALC, PLD6*
9	3.935	*PLN, ATP1A1, PDE38, FXYD1, KCNJ3, KCNJ5*

which information might move from one node to another. It measures the farness, i.e., how far/close the nodes are from each other. The closeness centrality value for the aforementioned higher-degree nodes lies between 0.48 to 0.43. ARMH4, CD320, PRRT4, and IGFL3 are among the other higher CC values of 1, demonstrating their significance with the rest of the network (Fig.2b).

Based on the betweenness centrality (BC) communication flow, nodes with high betweenness centrality are interesting. As they are on the communication channel and have the ability to regulate information flow, nodes with high betweenness centrality are interesting. The BC value represents important nodes for signaling pathways and can form a target for drug discovery. Some important nodes with high BC values are ARMH4, EGFR, MYC, SRC, TNF, and ALB (Fig. 2c).

Sub-networks and hub proteins in T2DM

A cluster is a collection of nodes more linked to one another than the rest of the network. The complete PPI network of 5332 proteins was analyzed for clustering, yielding 66 clusters with high-weighted nodes (Table 1). These 66 clusters are deduced based on weighting, molecular complex prediction, and post-processing. All 66 clusters were studied individually and examined based on their topological parameters and biological significance. Enrichment analysis of these sub-networks revealed that the proteins (nodes) were present in various biological pathways such as the P53 pathway, pancreatic cancer, breast cancer, diabetic nephropathy, TGFß signaling pathway, AGE-RAGE signaling pathway in diabetes IL-17 signaling pathway, and TNF-alpha signaling *via* NF-kB. Based on these findings, we shortlisted nine clusters with enrichment *p-value* >0.05 that had proteins enriched in diabetes and diabetes-related disorders and had significantly high node degrees and cluster-based confidence scores (Table 2). Proteins such as SMAD3 and SMAD4 in cluster 1 are critical drivers in diseases such as pancreatic cancer. It has been reported that at the time of diagnosis, up to 80% of patients with pancreatic cancer have either newly diagnosed T2DM or impaired glucose tolerance.[16] Cluster 2 proteins such as IL-1β perform a variety of other roles in controlling insulin production and promoting β cell death, which can ultimately result in type 2 diabetes.[17] NLRP3 and TNF are also reportedly involved in diabetes-related inflammatory response. Complement component C3 is also discussed as the biomarker of diabetic neuropathy, and diabetes and coronary calcium scores have also been linked to C3.[18] Proteins in these clusters such as TNF, IL1B, EGFR, SRC, FN1 and INSR were also observed in 50 high-degree nodes selected. Therefore, we selected these proteins for further analysis.

Another approach to infer this network and identify critical proteins was done by implementing five different algorithms, including maximum neighborhood component (MNC), maximal clique centrality (MCC), edge percolated

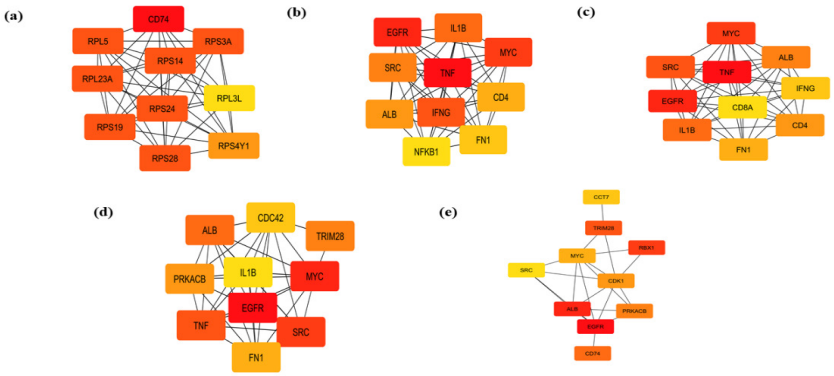

Fig. 3: Identification of top ten bottleneck proteins ranked by five algorithms, namely, (a) MCC, (b) EPC, (c) MNC, (d) stress, and (e) bottleneck where edge thickness corresponds to the degree of interaction between the differentially expressed proteins. Proteins like TNF, IL1B, FN1 and SRC were strongly interconnected, indicating their significance in the design mechanism

Table 2: Identified 15 hub proteins based on centrality, clustering analysis, and selection criteria parameters (node degree, expression) and their enrichment studies indicating various biological processes, molecular functions, and pathways involved.

Gene name	*Gene description*	*Node degree*	*Expression value (Logfc)*	*Pathway and processes*
EGFR	Epidermal Growth Factor Receptor	468	-3.213	Regulation of monooxygenase activity, glial cell proliferation, positive regulation of BIK/NF-kappa B signalling and positive regulation of protein localization to the cell periphery
IL1B	Interleukin 1 beta	321	-5.431	Negative regulation of protein secretion, interleukin-10 signaling, neuroinflammatory response, regulation of glucose transmembrane transport
TNF	tumor necrosis factor	439	-3.008	Type 2 diabetes, neuroinflammatory response, regulation of glucose import and regulation of glucose transmembrane transport,
IGF1	insulin-like growth factor 1	193	4.180	Neuroinflammatory response, regulation of glucose transmembrane transport, regulation of glucose import
NR1H4	nuclear receptor subfamily 1 group H member 4	88	-3.269	Type 2 diabetes mellitus, bile secretion, inflammation, cellular glucose homeostasis, positive regulation of insulin signaling pathway
KCNJ11	potassium voltage-gated channel subfamily J member 11	68	3.836	Type 2 diabetes mellitus, positive regulation of protein localisation to the cell periphery, negative regulation of protein secretion
INSR	insulin receptor	81	-1.526	Type 2 diabetes mellitus, regulation of glucose transmembrane transport, regulation of glucose import
ABCC8	ATP binding cassette subfamily C member 8	72	4.266	Type 2 diabetes mellitus, glial cell proliferation
CCL2	C-C motif chemokine ligand 2	191	-2.221	Interleukin-10 signaling, lipopolysaccharide-mediated signaling pathway, regulation of synaptic transmission, glutamatergic and positive regulation of phagocytosis,
MMP9	Matrix metallopeptidase 9	218	1.661	Positive regulation of vascular-associated smooth muscle cell proliferationIL-17 signaling pathway, regulation of neuroinflammatory response
PCSK1	proprotein convertase subtilisin/Kexin type 1	56	3.236	Response to glucose, positive regulation of protein secretion, response to fatty acids, response to calcium ions, obesity
ADCY5	Adenylate cyclase 5	77	1.338	Pancreatic secretion cAMP signaling pathway, phospholipase D signaling pathway, insulin secretion
SRC	SRC Proto-oncogene, non-receptor tyrosine kinase	497	-1.74	ERK Signaling Pathway, G-Protein-Coupled Receptors Signaling to MAPK/Erk Pathway, Jak/Stat Signaling: IL-6 Receptor Family Pathway
FN1	Fibronectin 1	384	-2.35	Focal adhesion, CCL18 signaling pathway, Epithelial to mesenchymal transition in colorectal cancer, Focal adhesion: PI3K-Akt-mTOR-signaling pathway
BRCA1	Breast and ovarian cancer susceptibility protein 1	280	1.18	P53 pathway, WNT pathway, Breast cancer pathway,

component (EPC), and bottleneck and stress based on shortest paths (Fig. 3). In topology analysis, proteins such as EGFR, IL1B, MYC, SRC, and TNF were the highest-degree proteins. They were also present in the top sub-networks based on the MCODE above analysis. These proteins were also seen in the top ten bottleneck proteins. Fibronectin 1 (FN1), one of the top ten proteins in this analysis, was also one of the highest degree nodes per the centrality analysis parameter of topology analysis. FN1 is an immune-related biomarker of diabetic neuropathy[18] and is often upregulated in the case of T2DM. After the network analysis, which included topological analysis, MCODE sub-networks, and bottleneck proteins, we selected 15 proteins that were present in all three analyses as hub proteins (Table 2). This validates that these proteins play a significant role in the pathways of diabetes.

FFAR interacting hubs and their role in the T2DM pathway

Based on network analysis, TNF, IL1B, SRC, MYC, EGFR, and FN1 were identified driver nodes/proteins. Gene expression markers like FN1 and C3 are often related to a signaling pathway and are regulated by a receptor protein. To understand the signaling mechanism of diabetes, we studied pathways with these high-degree nodes. As per the enrichment and network analyses, common nodes of these proteins, including FFAR4 were identified. The driver

Table 3: List of identified 23 FFAR4-interacting proteins involved in diabetic and other signaling pathways

Gene Symbol	*p-value*	*logFC*	*Regulation*
GPR142	0.022429	4.00172	Upregulated
CASR	0.048276	2.08709	Upregulated
FFAR4	0.018967	2.52194	Upregulated
FFAR3	0.027759	2.91793	Upregulated
TNF	0.045199	-3.00836	Downregulated
IL1B	0.000302	-5.431862	Downregulated
NR1H4	0.008618	-3.26955	Downregulated
PCSK1	0.003408	3.2366	Upregulated
TMSRSS6	0.0068	2.96765	Upregulated
MC3R	0.014839	2.718761415	Upregulated
ADIPOQ	0.034181	-1.806016555	Downregulated
P2RX2	0.035094652	1.015308842	Upregulated
NLRP3	0.032024105	-1.579400022	Downregulated
GCG	0.003195746	1.995321275	Upregulated
CCK	0.003823549	2.548849707	Upregulated
TAS2R1	0.048122311	1.614065789	Upregulated
SLC6A14	0.037189944	-1.913543613	Downregulated
ADRB1	0.021825591	1.123891992	Upregulated
GLP1R	0.017455646	1.829134171	Upregulated
TLR4	0.001640258	2.591891037	Upregulated
GNAQ	0.024097992	-1.311354527	Downregulated
LEP	0.009915991	3.90395877	Upregulated
GNAS	0.018258255	1.80558955	Upregulated

genes served as the missing links interacting with FFAR4. FFAR signaling pathway in T2DM showed the involvement of these nodes. We created another PPI network to increase our understanding of the critical proteins involved in this signaling pathway (Fig. 4a). This network was created with the parent nodes connecting to FFAR4 in the complete network of 5332 proteins and the high-degree hubs we identified from the network analysis. Using FFAR4 as the seed, the other nodes linking to it were identified (Table 3). 23 significant nodes, including NR1H4, KCNJ11, CASR, INSR, ABCC8, CCL2, MMP9, ADCY5, and PCSK1 were identified. The darker lines in the network depicted high confidence of interaction between FFAR4 and FFAR3, TNF, and IL1B (Fig. 4b). These interacting proteins share common pathways, like inflammation, homeostasis, and calcium signaling, with FFAR4. They are also reported to be co-expressed; for example, FFAR4-dependent release of GLP-1 leads to improved glycemic control from the entero-endocrine cells. The enrichment studies identify four main functional groups to be involved, namely, type II diabetes mellitus (ABCC8, INSR, KCNJ11, NR1H4, TNF, FN1), negative regulation of protein secretion (ABCC8,

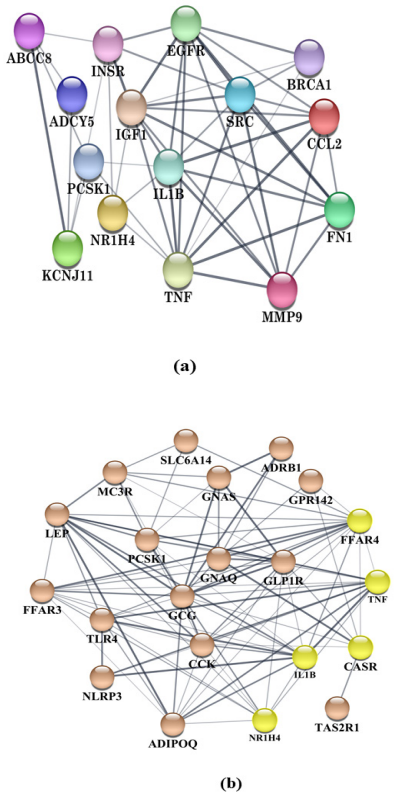

Fig. 4: (a) Network of 15 hub proteins selected from network analysis and (b)23 FFAR4 interacting nodes, including proteins TNF, IL1B, and NR1H4 from the hub protein network

IL1B, KCNJ11), regulation of glucose transmembrane transport (ADCY5, IGF1, IL1B, INSR, MMP9, TNF, FN1) and regulation of neuroinflammatory response (ABCC8, CCL2, IGF1, IL1B, INSR, MMP9, TNF). The biological processes, molecular function, and pathways corresponding to the hub proteins and FFAR4 interacting proteins are shown in Fig. 5. The final kappa score group for this PPI is 97, which defines term-term interactions shown as edges on the network and associates terms and pathways into functional groups based on shared proteins.

Emerging drivers (NR1H4 and CASR) interacting with FFAR4 - giving an insight in the FFAR4 mediated type 2 diabetic pathway

Proteins such as NR1H4 and CASR show novel interactions and linkages in the FFAR-mediated diabetic signaling pathway. This interaction has never been reported before in direct relation with FFAR4. This association can potentially have a significant impact on FFAR4-mediated diabetes.

CASR, while signaling through the $G\alpha_{i/0}$ subunit, inhibits the cAMP formation. cAMP is essential for calcium signaling through protein kinase A (PKA). It inhibits insulin secretion and limits the mitogenic action in the nucleus (Fig. 6). $G_{\beta\gamma}$ subunit works actively by activating

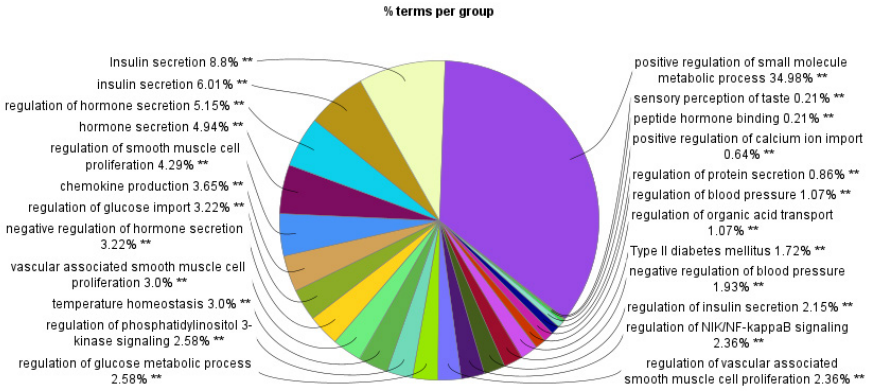

Fig. 5: KEGG pathways, biological processes and molecular functions associated with the hub proteins and FFAR4 interacting proteins wherein diabetes-associated processes can be observed as per cent terms per group in the pie-chart representation

Ras and initiating other signaling cascades like MAPK and ERK1/2. Other subunits of CASR, like $G\alpha_{q/11}$ and $G\alpha_{12/13,}$ employ conventional signaling pathways. FFAR4 signals through $G\alpha_{q/11}$ subunit by phosphorylation event at PLC dissociating into PIP2, IP3, and DAG, further activating protein kinase C (PKC), which increases the calcium concentration and in turn triggers pathways like JNK, IKK, MAPK. As the disease progresses, the body's ability to produce insulin becomes less responsive, and the outcome of insulin resistance also may lead to inflammation. β-arrestin bound to the FFAR4 is responsible for exerting inflammatory effects. TNF, IL1B, and NLRP3 are proteins involved in the inflammatory process. These proteins boost the production of several proteins that inhibit the insulin signaling pathway. NR1H4 encodes the FXR bile acid receptor, which is involved in both bile acid production and inflammation. FXR in white adipocytes suppresses the inflammatory cytokines that cause insulin resistance.

Kinetic Simulation of the biochemical pathway

Diabetes is an intricate disease. Multiple processes are involved in the occurrence of this disease, leading to the involvement of numerous diabetic pathways such as the AMP-activated protein kinase (AMPK) pathway, the insulin signaling pathway, renin angiotensin aldosterone system (RAS) pathway, peroxisome proliferators activated receptors (PPAR) regulation and TGFß signaling pathway. To decipher the signaling mechanism involved in the cellular events associated with FFAR4in type 2 diabetes, we constructed a biochemical pathway (Fig. 7). The ultimate goal of constructing this pathway was to understand the effect of insulin secretion, glucose uptake, calcium ion concentration and other cellular processes on the various species involved in it. The biological information of interacting species (gene, protein, and other molecules) was collected through the aforementioned PPI network analysis, KEGG regulatory pathways of diabetes and diabetes-linked disorders, as well as through the literature.[19,20]

A mathematical computational framework of the signaling pathway of FFAR4 was constructed and visualized in COPASI. The pathway consists of 55 metabolites and 60 reactions. The reactions occur in the pancreatic beta cell, nucleus, and cytoplasm. To study the kinetic behavior of the components of this pathway, we divided it into individual sub-pathways, the FFAR4 signaling pathway and the CASR and NR1H4 associated pathway. The concentrations were assigned to each component of the biochemical pathway in nanomolar. The initial concentrations for species were set by referring to the databases and literature. The kinetic laws implemented were mass action kinetics (drug-receptor interactions) and the Michaelis. Menten equation (kinases) that could explain the changes in intracellular metabolite concentrations within a pancreatic β-cell.

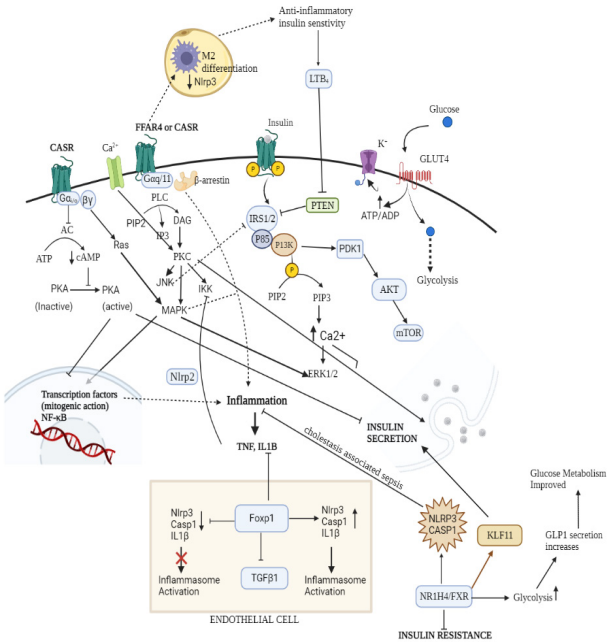

Fig. 6: Pathway depicting different signaling mechanisms carried out by identified drivers, CASR, FFAR4, and NR1H4 in type 2 diabetes

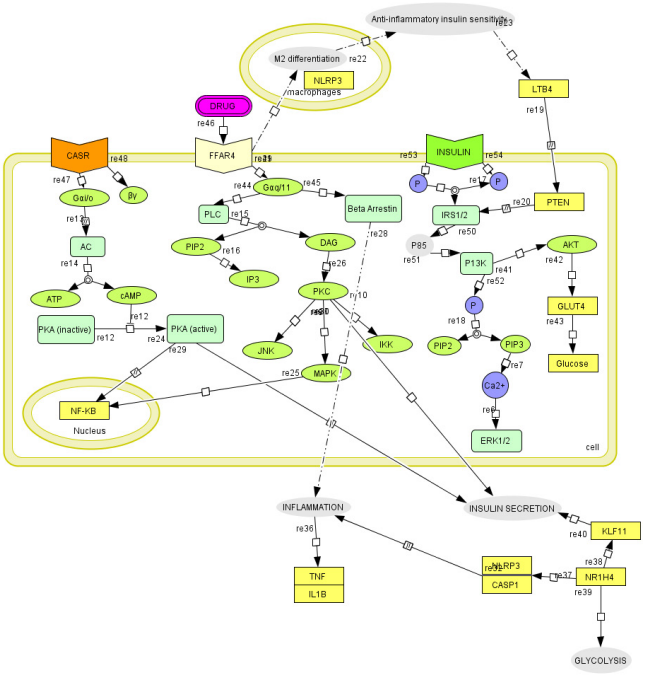

Fig. 7: Biochemical pathway including the components (gene, proteins, species, drug, ion, and other molecules) in different compartments with the assigned kinetic laws for studying the effect of insulin secretion and glucose uptake

Kinetic simulations are plotted as time *vs* concentration and the effect of glucose, insulin, calcium concentration, and cAMP levels were monitored.

- *FFAR4drug-induced effects:*

In our earlier study, we identified hit compounds through robust high-throughput virtual screening and molecular dynamics studies that may be potential agonists for FRAR4. Through high-throughput virtual screening and machine learning techniques, we screened 4 lakh compounds and shortlisted 52 potential hits.[21] The study focused on the structure modeling and identification of hit compounds, including the ones that are used in the present study (comp50, comp35, compN1 and compN2). These selected hit compounds were shortlisted based on non-bonded interactions, binding affinity and pharmacokinetic profiling, including Lipinski rule evaluation, protein plasma barrier, blood-brain barrier, PAINS and hepatotoxicity parameters. Additionally, findings were validated using molecular dynamic simulations and analysis such as free binding energy calculation, principal component analysis and cross-correlation matrices.

In the present study, the kinetic simulation of four selected compounds are used to examine the effect of these compounds on FFAR4. The compounds selected are comp50, comp35, compN1 and compN2, as given in Table 4. The kinetic simulations of these test compounds: comp50, comp35, compN1 and compN2 were compared with the known agonist TUG-891 (44-75nM) for FFAR4 at different concentrations and time intervals. The initial concentrations of comp50, comp35, compN1 and compN2 were 970, 150, 73 and 9.67 nM, respectively. We analyzed the improved effect of insulin secretion, glucose uptake, and calcium concentrations at time intervals 25, 50, 100, and 120 ns by the presence of drug molecules (Table 5).

We observed the signaling in multiple ways to understand the effects and outcomes on the pathway, including (1) without the drug (diseased condition), (2) in the presence of a known agonist, TUG-891, (Diseased state with the reference drug) and (3) Disease conditions with the

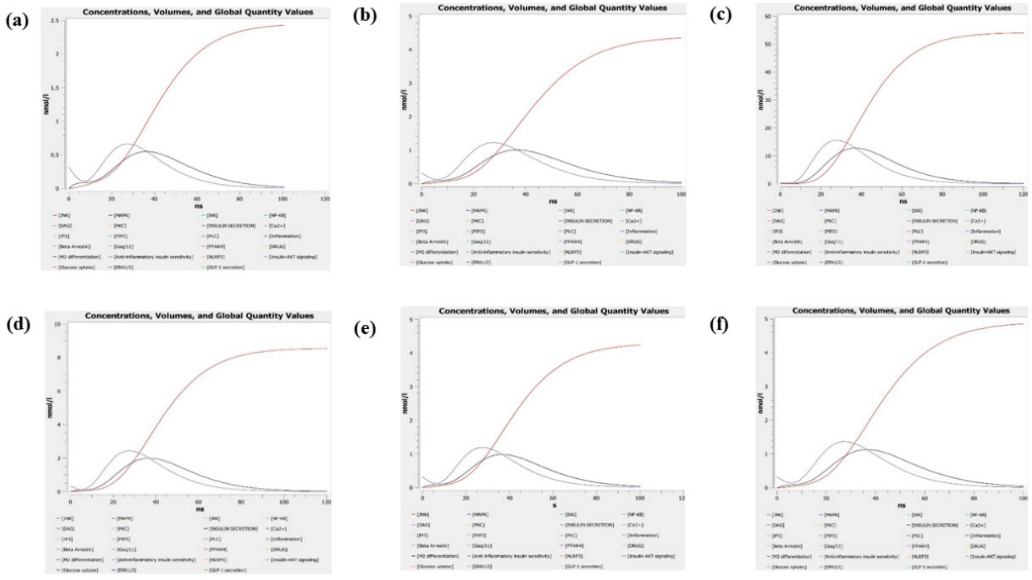

Fig. 8: Kinetic simulations of FFAR4-induced signaling pathway to study the effect of insulin secretion (black), glucose uptake (red), and calcium concentration (violet), (a) in the absence of drug, i.e., diseased condition, and in the presence of (b) TUG891, (c) comp50, (d) comp35, (e) compN1 and (f) compN2

Table 4: Selected four compounds for kinetic simulations with IUPAC names, molecular formulas and molecular weight for studying their effect on FFAR4

Molecule	*Molecular formula*	*Molecular weight (g/mol)*	*IUPAC Name*
Comp50	$C_{20}H_{21}NO_2$	322.4	*N-(3-benzamidophenyl)thiophene-2-carboxamide*
Comp35	$C_{22}H_{20}N_2O_2$	344.4	*3-methyl-N-[3-[(3-methyl benzoyl)amino]phenyl]benzamide*
CompN1	$C_{22}H_{20}N_2O_3$	360.4	*N-[3-[(2-methoxybenzoyl)amino]phenyl]-2-methylbenzamide*
CompN2	$C_{22}H_{20}N_2O_4$	376.4	*2-methoxy-N-[3-[(2-methoxybenzoyl)amino]phenyl]benzamide*

Table 5: Effect of the drug concentrations (in nM) w.r.t time intervals (in ns) on insulin secretion, glucose uptake, and calcium concentration in the biochemical pathway

Drug molecules	*Concentration (nM)*	*Time interval (ns)*	*Effect in the presence of drug*		
			Insulin secretion	*Glucose uptake*	*Calcium conc.*
TUG891	75	25	0.72	0.61	1.20
		50	0.7	3.03	0.47
		100	0.025	4.35	0.025
		120	stable	stable	stable
Comp50	970	25	8.8	6.26	15.18
		50	9.8	35.52	6.5
		100	0.6	53.43	0.3
		120	stable	stable	stable
Comp35	150	25	1.44	1.164	2.39
		50	1.54	5.64	0.99
		100	0.05	8.47	6.025
		120	stable	stable	stable
CompN1	73	25	0.95	0.78	1.603
		50	1.05	3.86	0.67
		100	0.04	5.77	0.015
		120	stable	stable	stable
CompN2	9.67	25	0.81	0.69	1.35
		50	0.86	3.27	0.55
		100	0.03	4.8	0.03
		120	stable	stable	stable

selected potential hits (comp50, comp35, compN1 and compN2). The results are plotted in concentration *vs*. time interval graphs (Fig. 8). The three effects that are monitored in these plots are glucose uptake (red), insulin secretion (black) and calcium concentration (violet). In diseased conditions, the glucose uptake can be observed going increasingly high even after simulating it to 100 ns. Meanwhile, in the presence of TUG-891, the glucose is controlled and reaches a plateau phase after 100 ns at a concentration of 4.35 nM, showing a stable system (Fig. 8a and b). Similarly, the other four novel hits exhibited similar responses when simulated with kinetic reactions (Fig. 8c-f). Furthermore, comp50, comp35, and compN1 attained stability at a higher concentration than TUG-891 and compN2. In contrast, compN2 showed an effect equivalent to TUG-891 on glucose uptake. We observed that insulin and calcium concentrations were consumed adequately after completing the 100ns period in all the simulation graphs. Based on this analysis, we conclude that compN2 has the potential to show therapeutic effects in type 2 diabetes therapy.

Kinetic simulations of driver proteins CASR and NR1H4

Based on the literature, we set the initial concentrations of CASR (0.4 nM) and NR1H4 (0.38 nM) and a hypothetical concentration for the drug to act at 3.2 nM to induce biological effects in the signaling pathway. The CASR signals through three subunits: Gαq/11, Gs, and Gαi/o. The Gαi/o is responsible for AC's activation and inhibition, affecting cAMP production. Inhibiting AC results in reduced cAMP production; hence, not activating PKA, and insulin is not secreted enough (Fig. 9a). However, if we activate cAMP production by activating AC, it results in the activation of PKA. Hence, insulin secretion also increases.