Sodium-glucose Cotransporter-2 Inhibitors as Modulator of Dipeptidyl Peptidase-4 in Diabetes

Misbahuddin M Rafeeq[1,*], Shahnaz Haque[1], Ziaullah M Sain[1], Ahmad Alzamami[2], Norah A Alturki[3], Mutaib M Mashraqi[4], Youssef Saeed Alghamdi[5], Rashed Ahmed Alniwaider[6], Ayyub Patel[7]

[1]Department of Pharmacology, Faculty of Medicine, Rabigh, King Abdulaziz University, Jeddah, SAUDI ARABIA.
[2]College of Applied Medical Science, Clinical Laboratory Science Department, Shaqra University, SAUDI ARABIA.
[3]College of Applied Medical Science Clinical Laboratory Science Department King Saud University, Riyadh, SAUDI ARABIA.
[4]Department of Clinical Laboratory Sciences, College of Applied Medical Sciences, Najran University, Najran, SAUDI ARABIA.
[5]Department of Biology, Turabah University College, Taif University, Taif, SAUDI ARABIA.
[6]Department of Pathology and Laboratory Medicine, Ministry of National Guard Hospital and Health Affairs (MNGHA), SAUDI ARABIA.
[7]Department of Clinical Biochemistry, College of Medicine, King Khalid University, Abha, SAUDI ARABIA.

ABSTRACT

Background: Genetic disorders such as diabetes have severe implications on human health. Mutation or aberrant activity of different proteins are associated with diabetes. The hyperactivation of the peptidase function of dipeptidyl peptidase-4 (DPP4) strongly correlates with the elevated level of blood glucose in diabetic patients. **Aim:** Preventing the activity of DPP4 by small molecule modulators is an excellent approach that proposes to curb the aggressiveness of diabetes. Blocking the DPP4 function quantitatively raises glucagon-like peptide 1 (GLP-1) in the blood that finally lowers the level of glucose in circulating fluids. **Materials and Methods:** In this study, we have conducted an elaborate investigations of the sequence-based structural properties of DPP4 protein by using various computational methods in order to find protein's antigenic and drug-binding regions. **Results:** Using the dataset of sodium-glucose transport protein 2 (SGLT2) inhibitors, we have identified a set of molecules that are predicted to bind DPP4. We have characterized the dipeptide ubenimex as the most potential modulator of DPP4. **Conclusion:** Based on the findings of current study, we concluded that our study has decoded the inhibitory module of DPP4 by the approach of structure-guided drug identification.

Key words: SGLT2, DPP4, Docking, Diabetes, Ubenimex.

INTRODUCTION

Diabetes is a major type of metabolic disorder that results in the accumulation of sugar, mostly glucose in the blood, leading to harmful effects on different organs of the human body. It is one of the significant causes of fatality in various age groups across the world. According to the World Health Organization (WHO) reports, approximately 1.6 million people succumbed to diabetes in 2019, and a population of more than 422 million people is symptomatically affected by this disease. In different reports, it is also referred that diabetes financially overwhelms the medical and medicinal expenses to a whopping amount of USD 727 billion (Muc, Saracen, and Grabska-Liberek, 2018; Saifulsyahira, Salmiah, Juni, and Sciences, 2018). Thus, diabetes remains a major concern in human health, research, and the pharmaceutical industry. The common forms of diabetes in humans are mainly categorized into two different groups based on the insulin-responsive characteristics of the disease. Type-I diabetes, also known as insulin-dependent diabetes, ensues due to perturbations of insulin production and secretion from the beta cells of pancreatic islets. Type-II diabetes, though independent of insulin production, is correlated to the nonresponsive behavior of target cells to insulin stimulus. While genetic factors underscore

both types of diabetes, the manifestation of genetic alterations is exploited at different cellular events. For example, the translational machineries and exocytosis-related proteins of cells are less-to-nonfunctional in type-I diabetes, resulting in malfunctioning of the endocrine system of the pancreas. On the other hand, the effector proteins and signal transducers of insulin, such as insulin receptors, glucose transporters, endocytosis-related proteins, and protein kinase,s are nonresponsive to insulin due to mutations and fluctuations in the cellular conditions. In either of the types, the overall effect is hyperglycemia. The molecular mechanisms underlying the induction and progression of diabetes in human is diverse and interconnected. While some environmental effects, such as food, lifestyle habits, and obesity, are correlated to diabetes, it has a strong genetic predisposition. Mutation and expressional abnormalities of certain genes in the sugar catabolism, transport, and reception processes in target cells are involved in the development of diabetes. For example, low expression or functional inactivation of glucose transporters GLUT1, GLUT4, causes increased blood glucose accumulation, which predispose to prediabetic phenotype or type 2 diabetes. Alteration of intracellular metabolic flux also leads to diabetes. Gluconeogenesis, aberrant cellular secretion process and inactive insulin receptors are also reported to perturb the physiological homeostasis of glucose metabolism.

Dipeptidyl peptidase-4 (DPP4) or CD26 is a cell surface-residing serine protease (exopeptidase) that cleaves dipeptides. DPP4 has the specificity for the dipeptides, which contain an alanine or proline in the second amino acid position. DPP4 functions in the intracellular signaling pathways that finally regulate metabolism and cell survival. The involvement of DPP4 in these processes is attributed to its proteolytic activity against growth factors, hormones, and trophic factors. While DPP4 activity is coherently linked to certain types of cancers like leukemia, recent studies have also pointed out the importance of DPP4 in regulating diabetes. The upregulation of DPP4 expression or activity leads to the enhancement of diabetic potential in the patients. On the other hand, reduced activity of DPP4 causes abnormalities of the cell surface receptors that include but are not limited to sugar transporters. A correlation of DPP4 activity is attributed to CD5 function (Connell *et al.*, 2012; Osman, 2018; Rodrigo, Lauret-Braña, and Pérez-Martnez; Sena, Bento, Pereira, Marques, and Seiça, 2013). Therefore, inhibition of DPP4 activity

accounts for a promising approach to reducing the glucose accumulation effects in diabetes.

DPP4 inhibitors, a typical class of molecules that lower the blood glucose level, is used as a conventional medication against diabetes. Food and Drug Administration (FDA) have approved several DPP4 and SGLT2 inhibitors for glycemic controls in type-2 diabetes (T2DM). However, some of these drugs have side effects, including genital infections (Cuypers, Mathieu, and Benhalima, 2013; Fadini, Bonora, and Avogaro, 2017; Fala and benefits, 2015). Thus, new formulation on drug repurposing and combinatorial therapy has become more popular to treat the conventional forms of diabetes. Both SGLT2 and DPP4 inhibitors can be utilized to treat T2DM patients who are unresponsive to insulin therapy (Gooßen, Gräber, and metabolism, 2012; Savarese *et al.*, 2016; Singh, Singh, and metabolism, 2016). While different DPP4 inhibitors like vildagliptin and sitagliptin are currently being used to treat T2DM, an investigation on more effecient DPP4 inhibitor is warranted. Therefore, in the present study, we aim to identify modulators of DPP4 from the SGLT2 class of inhibitors. We have used an array of computational approaches to annotate ubenimex as a strong candidate for DPP4 inhibition.

METHODOLOGY

Phylogenetic Analysis

Homologous sequences of DPP4 were extracted using PHMMER[1] search considering Homo sapiens DPP4 protein (Uniprot ID: P27487) as a query. The non-redundant sequences were then subjected to structure-guided multiple sequence alignment using PROMALS3D.[2] The aligned sequences were then utilized for generating a neighbor-joining phylogenetic tree using Mega.[3] For phylogeny test, the bootstrapping method was utilized (1000 repeats), and the final cladogram in circular view was plotted using Tree View version 1.4.2.

Sequence feature Calculations

DPP4 sequences from 5 species, *Homo sapiens* (P27487), *Mus musculus* (P28843), *Rattus norvegicus* (P14740), *Gallus gallus* (A0A1D5PJA5), and *Danio rerio* (A0A2R8QSL0) were used to predict secondary structures (using PSIPRED version 4.0,[4,5] domain organization (InterProScan server,[6] transmembrane helix (TMHMM server[7] and antigenic region (EMBOSS server.[8] The predictions obtained from different servers were plotted as two-dimensional representations using IBS illustrator.[9]

Structural feature calculations

The monomeric structure of DPP4 protein from the known crystal structure of the protein in complex with heterocyclic ligand N7F (PDB ID: 4A5S[10] was extracted and further used to analyze the structural features. The physicochemical properties such as hydrophobicity and electrostatic potential were calculated utilizing the Pymol plugin (APBS module.[11] The sequence properties such as domain organization and antigenic regions were also plotted onto the structure to understand the proteins's probable docking site.

Protein preparation for molecular docking

The monomeric structure from the known crystal structure of DPP4 in complex with heterocyclic ligand N7F (PDB ID: 4A5S) was extracted after removing bound ligand (N7F) and additional cofactors (NAG, MAM, and SO_4). The monomer so obtained was subjected to energy minimization using YASARA online server. The final structure was then utilized for high throughput molecular docking.[12]

High throughput screening of SGLT2 inhibitors against DPP4

The basic backbone of the SGLT2 inhibitor known as Phlorizin was used as a query to identify analogous ligands from the Zinc database. The ligands obtained (~450) were categorized as endogenous, metabolic compounds, and natural products, according to the Zinc database. A high throughput screening was performed for all the ligands using the blind docking method on the MTiAutoDock server.[13] The docked poses of the ligand were analyzed using Pymol.

RESULTS AND DISCUSSION

Evolutionary conservation of Human DPP4

Evolutionary analysis of any protein from a species provides insight into its convergence or divergence with respect to proteins of other species. A phylogenetic tree derived from multiple sequence alignments from different species provides insight into the sequence conservation across the species, and thus the protein can be explored as a probable drug target. The homologous sequences of DPP4 from various species were extracted using Hiden Markov Model-based search against the reference proteome database (PHMMER). The non-redundant sequences so obtained were then subjected to structure-guided multiple sequence alignments using PROMALS3D. It is also pertinent to understand the structural conservation of the target protein and identify the ligand binding sites. The structure-guided alignment
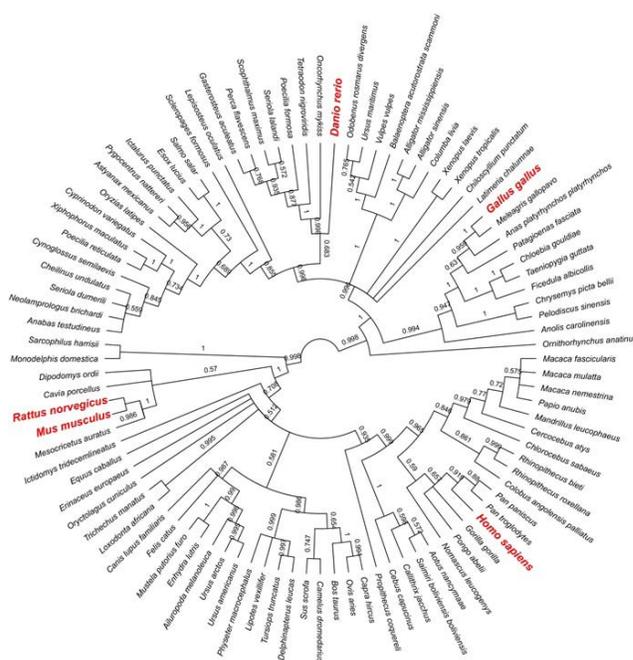

Figure 1: Phylogenetic analysis of orthologous sequences of DPP4.

Structure-guided alignments of orthologous DPP4 sequences were subjected to phylogenetic analysis to generate a neighbor-joining tree. In this unrooted tree, the branch label depicts the fraction of bootstrap values obtained after the phylogeny test. The closely related species, i.e. *Homo sapiens, Mus musculus, Rattus norvegicus, Gallus gallus,* and *Danio rerio,* are colored in red to depict their relative positioning in this well-conserved cladogram.

of approximately 100 homologous sequences with more than 25% sequence identity was then subjected to phylogenetic analysis. The neighbor-joining tree generated after utilizing the bootstrap method as a test of phylogeny depicts the variation in insertion/deletion across the species (Figure 1). However, DPP4 of five model organisms, i.e., *Homo sapiens, Mus musculus, Rattus norvegicus, Gallus gallus,* and *Danio rerio,* were placed equidistant in the cladogram (marked in red), thus indicated variation as well as similarity with each other. Overall, the cladogram depicts a well-conserved protein in terms of structural homogeneity.

The structure-guided multiple sequence alignment (Figure S1) of the model organisms considered from this phylogenetic analysis (i.e., *H. sapiens*, *M. musculus*, *R. norvegicus*, *G. gallus,* and *D. rerio*) depicted conservation of the protein with only a few insertion/deletions.

Sequence features of DPP4 from different species

It is essential to understand protein's secondary structure, domain organization, and antigenic regions to identify drug target sites in the protein. To determine that the variability observed through the phylogenetic analysis translates into structural composition/domain organi-
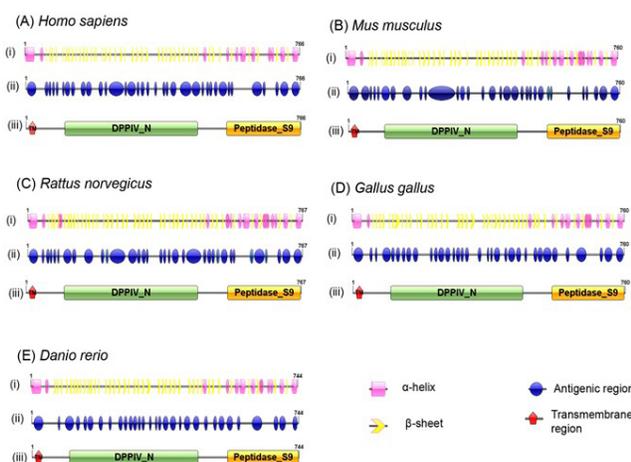
Figure 2: Sequence features of DPP4 from different species.
Prediction of sequence features from 5 different species (*H. sapiens*, *M. musculus*, *R. norvegicus*, *G. gallus,* and *D. rerio*). Prediction of (i) secondary structure using PSIPRED, (ii) antigenic regions using EMBOSS program, (iii) transmembrane helix, and domain using TMHMM and InterProScan, respectively were plotted as 2D representations using IBS illustrator.

zation, the sequence features of DPP4 were calculated for 5 model species i.e., *H. sapiens*, *M. musculus*, *R. norvegicus*, *G. gallus* and *D. rerio,* as depicted in Figure 2.

The secondary structure composition of the DPP4 predicted using PSIPRED depicts a β-propeller region followed by an α-helix region in all orthologous sequences. The domains predicted using InterProScan for the five species sequences consisted of two PFAM domains, i.e., DPPIV_N and Peptidase_S9. The transmembrane helix prediction performed using the TMHMM server[7] predicted transmembrane helices in the N-terminal region (5-29) of DPP4. The secondary structure, domain composition, and transmembrane helical regions were observed to be conserved in all these species. However, the antigenic regions that were predicted had variations in the number and length of regions.

Determination of the antigenic regions in a protein helps to identify those regions that are mainly on the surface-exposed. Analysis of the data of experimentally determined antigenic regions on the protein revealed that the antigenic patch consists of surface-exposed hydrophobic residues such as Cys, Leu, and Val. The minimum length of the antigenic region was set as six amino acids, which led to the prediction of 35 antigenic regions in *H. sapiens*, 36 in *M. musculus*, 39 in *R. norvegicus*, 35 in *G. gallus,* and 37 in *D. rerio.* The longest, as well as highest scoring antigenic region, consists of 36 amino acids; 5-40 in *H. sapiens*'s DPP4, as depicted in Figure 2. This stretch of DPP4 was predicted to be the highest-scoring antigenic region in all the model species. Since this region also includes amino acids from 5-29, which

constitutes the transmembrane region of DPP4, therefore, this high scoring antigenic region (5-40) was not ideal for performing docking analysis against the ligand library. Therefore, additional analysis at the structural level was carried out to identify the probable docking site.

Physiochemical properties of Human DPP4

The physicochemical properties were mapped onto the surface of Human DPP4 by using the monomeric structure of Human DPP4, which was extracted from the known crystal structure utilizing PDB ID 4A5S. The protein in this crystal model was present as a dimeric protein bound to ligand (N7F) as well as additional cofactors (NAG, MAM, and SO$_4$). The monomeric structure was then energy minimized and then analyzed for physicochemical properties such as hydrophobicity and electrostatic potentials. The sequence features obtained were also mapped onto the energy minimized structure of DPP4 to predict the probable regions of ligand binding for high throughput screening through molecular docking.

Figure 3(A) depicts the energy minimized structure of Human DPP4 (40-766) colored according to the secondary structure constitution (α-helices: pink and β-sheets: yellow). They corroborate well with the secondary structure predicted using sequence information. The domains predicted by InterProScan[6] were plotted onto the structure (DPPIV_N: green and Peptidase_S9: orange), as shown in Figure 3(B).

Physicochemical properties such as hydrophobicity provide the basis of interaction between the protein and the ligand. A more hydrophobic site is preferred to be the docking site for the ligand over a hydrophilic site. Figure 3(C) depicts the hydrophobicity mapped for each residue on human DPP4 to identify hydrophobic sites (marked red on the protein structure). The electrostatic potential/charge on the surface of the protein also plays a crucial role in determining both the composition as well as the orientation of the ligand in molecular docking studies. Hence, the electrostatic potential of the residues through their three-dimensional arrangement in the crystal structure was calculated and plotted utilizing the APBS plugin in pymol (Figure 3D). As depicted in Figure 3(D), the distribution of charge is more electronegative (red) on the surface of the protein than electropositive (blue).

The antigenic regions predicted using the EMBOSS plugin were also mapped onto the structure of Human DPP4 to identify the stretch of amino acids that are present together in the three-dimensional space and are also surface-exposed for ligand binding. As depicted in
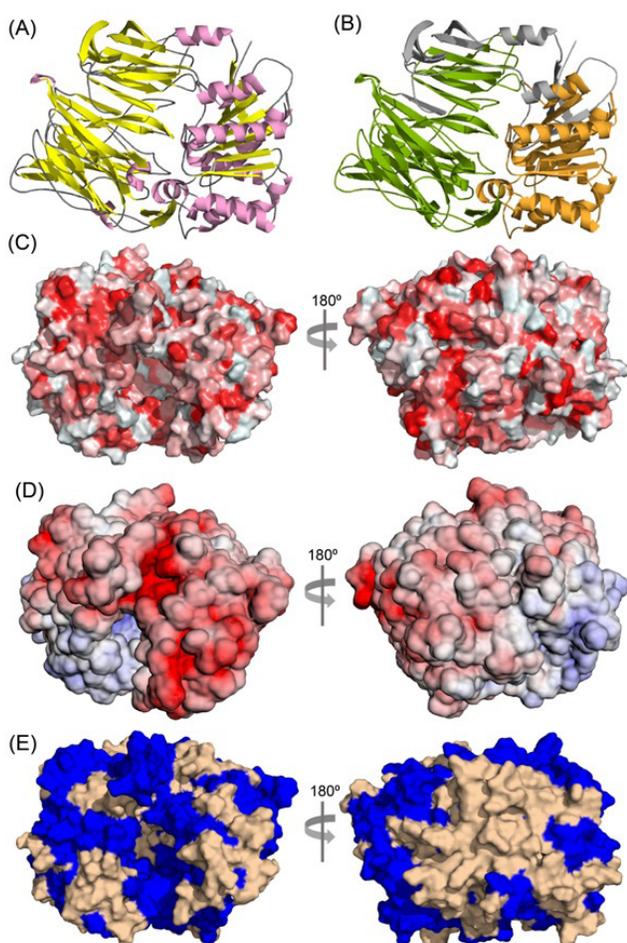
Figure 3: Structural features of Human DPP4.
The monomeric structure of the DPP4 was extracted from PDB (ID 4A5S), and various physicochemical properties were calculated. (A) Energy minimized monomer structure of DPP4 with α-helices (pink) and β-sheets (yellow). (B) Predicted domain organization for DPP4; DPPIV_N (green) and Peptidase_S9 (orange). (C) Hydrophobicity predicted for each residue shown in 180° rotated surface view (red: high hydrophobicity, white: low hydrophobicity) (D) Electrostatic potential predicted for each residue shown in 180° rotated surface view (red: negatively charged residues, blue: positively charged residues) (E) Antigenic regions predicted on DPP4 (blue) shown in 180° rotated surface view.

Figure 3(E), 12 of the 35 predicted antigenic regions are present on the protein surface, while 16 of the remaining stretches are less than 10% surface-exposed or are forming cavity in the protein. The seven predicted antigenic regions were found to be completely buried in the core of the protein. The antigenic regions which were fully or partially surface-exposed were more likely to form the binding site for the ligand (marked in blue in Figure 3E).

High throughput screening of SGLT2 inhibitor analogous ligands against DPP4

The Zinc database was used to search for analogous of SGLT2 inhibitors to conduct the high throughput ligand-screening against DPP4. The base structure

Phlorizin (Zinc ID: ZINC000044021884) was selected, which is experimentally-determined inhibitor against SGLT2, to search for analogous compounds.

Phlorizin is a natural product that was isolated from the bark of apple trees.[14] It is known to reduce the plasma glucose levels by blocking glucose absorption from renal and intestinal cells by inhibiting SGLT1 and SGLT2.[15] Through *in-vitro* studies, it has been established that Phlorizin exhibits Ki values of 140 and 11nM against SGLT1 and SGLT2, respectively.[16,17]

Nearly 500 analogous compounds to Phlorizin were extracted from the Zinc database, which were categorized as either endogenous, metabolites, natural products, or bioactive compounds. These ligands were used to screen against the energy minimized monomeric structure of DPP4 in a blind docking setting using MTiOpen-AutoDock server. Although, predicted antigenic site as well as the physicochemical properties, suggested certain probable regions of the interaction of the ligand with DPP4, a blind method was selected to filter out the best possible conformation. A blind docking method provided, 10 different binding conformations for each of the selected ligands. Hence, Phlorizin was used to dock onto the DPP4 structure to define the search space for the subsequent high throughput screening.

Based on the lowest energy conformation obtained using Phlorizin, the docking site was compared with the predicted-antigenic regions for DPP4. The final search space defined by 4 antigenic regions (360-365, 404-420, 431-462 and 464-486), were selected (Figure 4), which were also found to be similar in other species used in this study (Figure S1).

The molecular structure of human DPP4 was obtained from ligand-bound N7F (Figure S2A). It was also found that the ligand-binding region of DPP4 with heterocyclic ligand N7F is different from the predicted binding site for SGLT2 inhibitors (Figure S2B). The residues within 5Å that are involved in the interaction with the ligand were Glu206, Try547, Trp629, and His740 (Figure S2C).
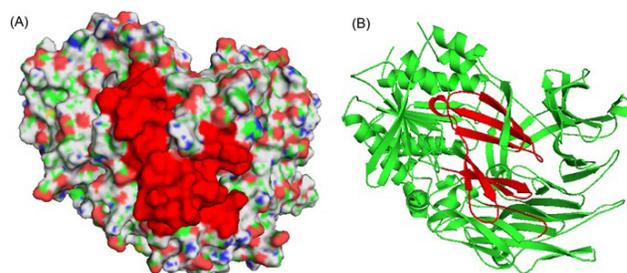

Figure 4: Prediction of Docking site on DPP4.
The docking site in DPP4 was selected by comparing the docking confirmation with the predicted-antigenic regions for DPP4 (A) Surface view and (B) Cartoon view of the most probable docking site identified (red).

Figure S1: Structure based multiple sequence alignment of DPP4 orthologs.

Structure-guided multiple sequence alignment for five species (*H. sapiens, M. musculus, R. norvegicus, G. gallus,* and *D. rerio*). The residues highlighted in red and marked with black bars defines the predicted-antigenic regions and are also selected as docking site for Human DPP4.
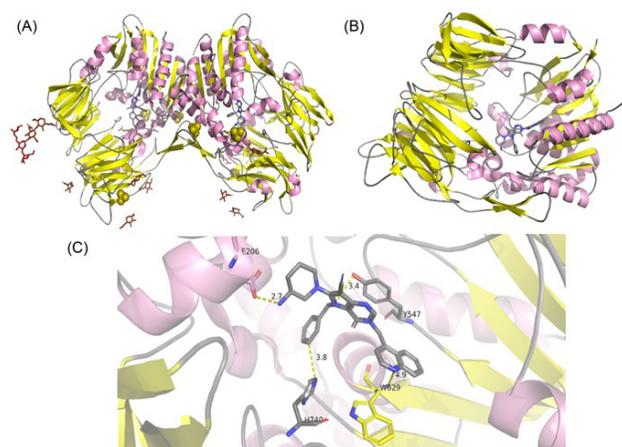

Figure S2: DDP4-N7F docking conformation as obtained from the crystal structure.

(A) The dimeric structure of DPP4 bound to ligand (N7F) and other cofactors (NAG, MAM, and SO4) depicted in cartoon representation. (B) Cartoon representation of DPP4 bound to ligand (N7F) (C) Critical residues of DPP4 involved in binding to the ligand within 5Å.

This binding pocket was not in close proximity to the predicted binding site of SGLT2 inhibitors. Thus, indicting a unique interaction of SGLT2 inhibitors with DPP4.

In this study, we have selected Phlorizin as the base compound, therefore, its binding in the predicted binding pocket of DPP4 was analyzed. As depicted in Figure 5 (A), Phlorizin binds to DPP4 in a cavity within the selected docking grid and exhibited binding energy of -10.75 kcal/mol (Table 1). A total of 10 amino acids of DPP4 were found to interact with Phlorizin, i.e. Pro362, His363, Thr365, Leu366, Ala409, Leu410, Thr411, Ser412, Lys463, and Ala465 and were found within 5Å distance (Figure 5B).

Further, approximately 500 compounds from Zinc database were found to have comparable binding energy to DPP4 with respect to Phlorizin in the high-throughput

screening. A total of 10 compounds which have highest binding affinity for DPP4 were selected as shown in Table 1.

The best binding energy of -11.3 kcal/mol was observed for the compound commonly known as Neocarthamin. Neocarthamin is an organic compound (flavonoid o-glycoside) extracted as a plant metabolite mainly from safflowers. The residue-based analysis of the docking conformation of Neocarthamin with DPP4 revealed that it has a better lock and key fit in the predicted ligand cavity (Figure 6A). A total of 13 amino acids of DPP4 were identified at a distance less than 5Å from the ligand and are involved in the binding viz, Glu 361, His363, Phe364, Thr365, Leu366, Ala409, Leu410, Thr411, Ser412, Phe461, Lys463, Glu464, and Ala465 (Figure 6B). Of these 13 amino acids, 6 residues were within 3Å distance from the ligand, thus suggesting that Neocarthamin has better binding to DPP4 as compared to Phlorizin.

Notably, the residues that are involved in interaction with Neocarthamin are hydrophobic, charged, and hydroxy amino acids, that resulted in optimal binding of ligand with the protein. Therefore, it is essential to understand the distribution of amino acids and their physicochemical properties for predicting the optimium docking site in the protein.
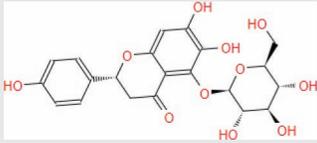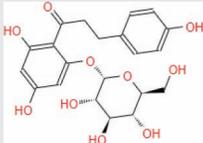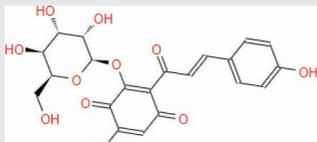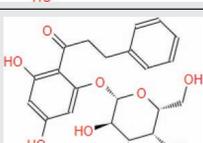
Notably, the amino acid of DPP4 which were found to interact with Neocarthamin are present in the DPPIV_N domain of the protein which is different from the known active site of this protein, which lies mainly in the Peptidase_S9 domain. The known active site of DPP4 lies in the C terminal (Peptidase_S9 domain) and has been shown to interact with a number of the drug-like compounds.[10]

Since the binding region of SGLT2 inhibitors to DPP4 is different than other peptidase inhibitors, the flavonoid compounds identified in this study depict a novel interaction which can be investigated further through experimental studies.

CONCLUSION

Clustering of sequences can provide essential information on the protein's structure, function, and modulation. While the evolutionary diversity adds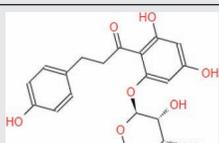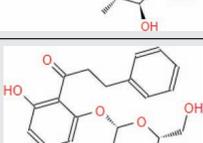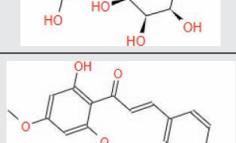 specific signatures to protein regions and domains, the key functional residues are conserved in paralogous and homologous proteins. Identification of such conserved and critical residues in catalytic proteins is important to identify the regions that can be targeted for identification for modulators, that in turn might have therapeutics potential against different diseases. Strategies to combat genetic diseases, such as diabetes, have relied

Zinc ID	Structural Representation	Binding Energy (kcal/mol)	Common name	Attributes
ZINC000257551108		-11.3	Neocarthamin	Endogenous human metabolite
ZINC000044021884		-10.75	Phlorizin	Known SGLT2 inhibitor through *in-vitro* studies
ZINC000100825248		-10.54	Carthamone	Endogenous human metabolite
ZINC000257357328		-10.4	--	Known SGLT2 inhibitor through *in-vitro* studies
ZINC000095579512		-10.37	Mahkoside A	Bioactive Natural Product
ZINC000014757230		-10.3	Kaempferol 5-Glucoside	Endogenous human metabolite
ZINC000004098321		-10.29	Glycyphyllin	Natural Product
ZINC000004252537		-10.22	--	Known SGLT2 inhibitor through *in-vitro* studies
ZINC000014726219		-10.2	Chalcono-sakuranetin	Endogenous human metabolite
ZINC000031164435		-10.11	Asebotin	Natural Product
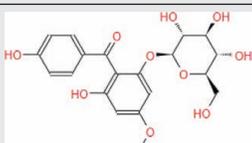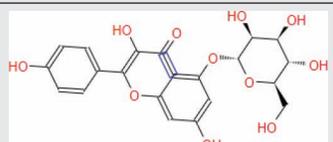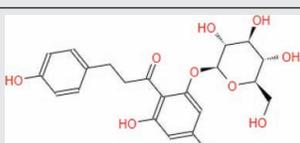
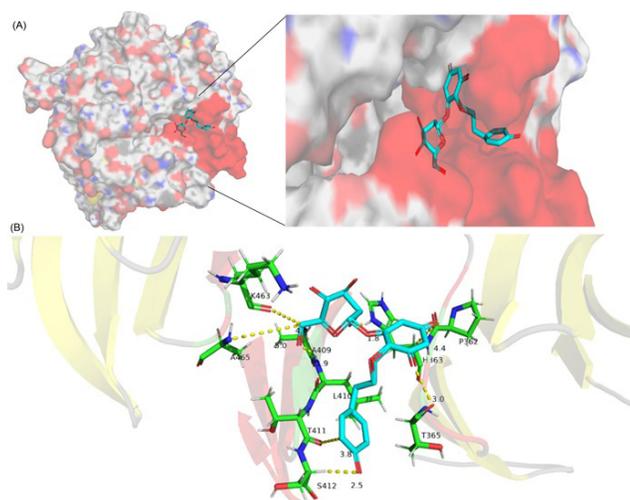Table 1: Top 10 ligands which have the highest binding affinity for DPP4.

Figure 5: Docking conformation of DDP4- ligand Phlorizin.
(A) Surface view of DPP4 with predicted docking site depicted in red and ligand (Phlorizin) in cyan. (B) Amino acid residues that are within 5Å of DPP4 (green) and are involved in binding to the ligand.
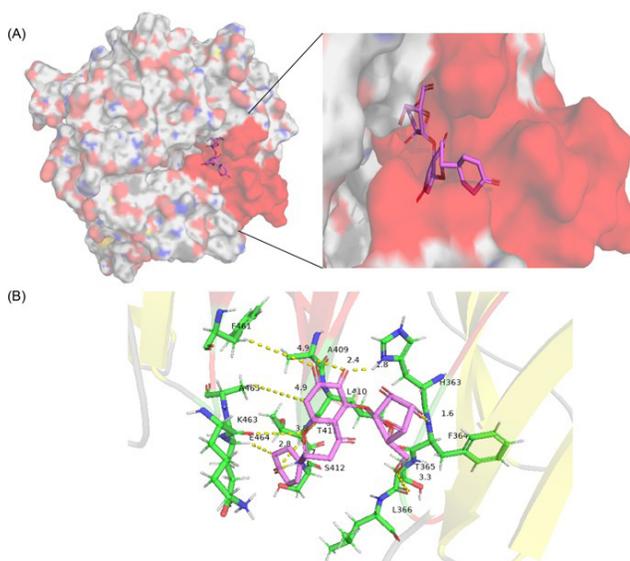

Figure 6: DDP4-Neocarthamin docking conformation.
(A) Surface view of DPP4 with predicted docking site depicted in red and ligand (Neocarthamin) in pink. (B) Critical residues of DPP4 (green) that are involved in binding to the ligand within 5Å distance.

upon identifying small molecule modulators against the proteins that are involved in sugar transport, reception, and metabolism pathways. Since the functional down-regulation of DPP4 is intrinsically associated with the lowering of blood glucose, it is tempting to analyze the sequences of DPP4 proteins from different organisms to comprehend its conserved regions that can be targeted by drugs or chemical modifiers.

In this study, we took a broad-spectrum approach to evaluate the sequence complexity of DPP4 proteins in different organisms, and theoretically rationalize the

potential inhibitors against this protein. We leveraged a simple process of dynamic analysis of DPP4 sequences in the comparative graphical format. Our comprehensive analysis of the potential antigenic regions and domains in DPP4 proteins has invaluable significance in predicting the structural determinants that can be targeted for drug development. Our study identifies the sequence-based structural characteristics, such as transmembrane regions, hydrophobic patches, and electrostatic charge interactions in DPP4, that provide precise information of the protein that can be targeted while conceiving its inhibition by antagonistic molecules. Moreover, the results of the phylogenic assessment of DPP4 from different organisms will help in comparative functional and metabolomic analysis. In principle, this study is aimed to accelerate the identification of modulators/ inhibitors against DPP4 through a structure-based docking approach. Given the fact that several of the SGLT2 inhibitors are FDA approved for the treatment of type-II diabetes mellitus (T2DM), we analyzed the qualitative and quantitative potential of SGLT2 inhibitors as the potential modulators of DPP4. We took advantage of a high-throughput virtual screening method to identify the most potential modulator of DPP4. We have screened a large number of gliflozins against DPP4 and identified ubenimex has the best binding ligand with the highest free energy of binding and lowest entropic fluctuation in the binding cavity of DPP4. Neocarthamin is a flavonoid plant metabolite that is extracted from safflowers. This organic compound is an analog of Phlorizin, which is a known SGLT2 inhibitor. Both *in-vitro* and *in-vivo* studies have determined that flavonoids such as Phlorizin are effective in decreasing the uptake of glucose through renal and intestinal cells. The inactivation of DPP4 by small modulators like Neocarthamin can amplify the inhibitory signals in glucose metabolism. Overall, this study will not only improve the focused structural analysis of human DPP4 protein, but it will also broaden the understanding of the interactions of SGLT2 inhibitors with human DPP4 protein.

In summary, we present a naïve set of results that combine the sequence-based analysis of structural attributes and evolutionary convergence of DPP4 domains in a systematic approach that helped to identify the SGLT2 inhibitor, Neocarthaminas, a strong inhibitor of DPP4.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

ABBREVIATIONS

SGLT2: Sodium-Glucose Cotransporter-2; **DPP4:** Dipeptidyl peptidase-4; **GLP-1:** Glucagon-like peptide 1; **T2DM:** type-II diabetes mellitus; **FDA:** Food and Drug Administration.

REFERENCES

1. Finn RD, Clements J, Arndt W, Miller BL, Wheeler TJ, Schreiber F, Bateman A, Eddy SR. HMMER web server: 2015 Update. Nucleic Acids Res. 2015;43(W1):W30-8. doi: 10.1093/nar/gkv397, PMID 25943547.

2. Pei, Pei J, Kim BH, Grishin NV. PROMALS3D: A tool for multiple protein sequence and structure alignments. Nucleic Acids Res. 2008;36(7):2295-300. doi: 10.1093/nar/gkn072, PMID 18287115.

3. Hall BG. Building phylogenetic trees from molecular data with MEGA. Mol Biol Evol. 2013;30(5):1229-35. doi: 10.1093/molbev/mst012, PMID 23486614.

4. Buchan DWA, Minneci F, Nugent TCO, Bryson K, Jones DT. Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Res. 2013;41(Web Server issue);Web Server Issue:W349-57. doi: 10.1093/nar/gkt381, PMID 23748958.

5. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol. 1999;292(2):195-202. doi: 10.1006/jmbi.1999.3091, PMID 10493868.

6. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S. InterProScan 5: Genome-scale protein function classification. Bioinformatics. 2014;30(9):1236-40. doi: 10.1093/bioinformatics/btu031, PMID 24451626.

7. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001;305(3):567-80. doi: 10.1006/jmbi.2000.4315, PMID 11152613.

8. Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. F.E.B.S. Lett. 1990;276(1-2):172-4. doi: 10.1016/0014-5793(90)80535-Q, PMID 1702393.

9. Liu W, Xie Y, Ma J, Luo X, Nie P, Zuo Z, Lahrmann U, Zhao Q, Zheng Y, Zhao Y, Xue Y, Ren J. IBS: an illustrator for the presentation and visualization of biological sequences. Bioinformatics. 2015;31(20):3359-61. doi: 10.1093/bioinformatics/btv362, PMID 26069263.

10. Sutton JM, Clark DE, Dunsdon SJ, Fenton G, Fillmore A, Harris NV, Higgs C, Hurley CA, Krintel SL, MacKenzie RE, Duttaroy A, Gangl E, Maniara W, Sedrani R, Namoto K, Ostermann N, Gerhartz B, Sirockin F, Trappe J, Hassiepen U, Baeschlin DK. Novel heterocyclic DPP-4 inhibitors for the treatment of type 2 diabetes. Bioorg Med Chem Lett. 2012;22(3):1464-8. doi: 10.1016/j.bmcl.2011.11.054, PMID 22177783.

11. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci U S A. 2001;98(18):10037-41. doi: 10.1073/pnas.181342398, PMID 11517324.

12. Land H, Humble MS. YASARA: A tool to obtain structural guidance in biocatalytic investigations. Methods Mol Biol. 2018;1685:43-67. doi: 10.1007/978-1-4939-7366-8_4, PMID 29086303.

13. Labbé CM, Rey J, Lagorce D, Vavruša M, Becot J, Sperandio O, Villoutreix BO, Tufféry P, Miteva MA. MTiOpenScreen: A web server for structure-based virtual screening. Nucleic Acids Res. 2015;43(W1):W448-54. doi: 10.1093/nar/gkv306, PMID 25855812.

14. White JR. Apple trees to sodium glucose Co-transporter inhibitors: a review of SGLT2 inhibition. Clin Diabetes. 2010;28(1):5-10. doi: 10.2337/diaclin.28.1.5.

15. Ehrenkranz JRL, Lewis NG, Kahn CR, Roth J. Phlorizin: a review. Diabetes Metab Res Rev. 2005;21(1):31-8. doi: 10.1002/dmrr.532, PMID 15624123.

16. Hummel CS, Lu C, Liu J, Ghezzi C, Hirayama BA, Loo DDF, Kepe V, Barrio JR, Wright EM. Structural selectivity of human SGLT inhibitors. Am J Physiol Cell Physiol. 2012;302(2):C373-82. doi: 10.1152/ajpcell.00328.2011, PMID 21940664.

PICTORIAL ABSTRACT
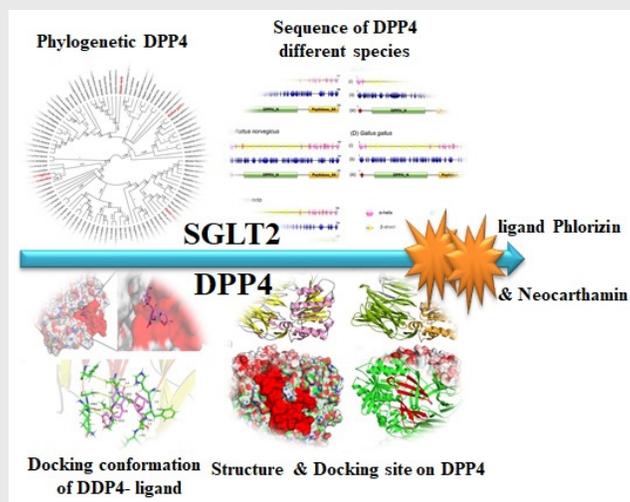
Cite this article: Rafeeq MM, Haque S, Sain ZM, Alzamami A, Alturki NA, Mashraqi MM, Patel A. Sodium-glucose Cotransporter-2 Inhibitors as Modulator of Dipeptidyl Peptidase-4 in Diabetes. Indian J of Pharmaceutical Education and Research. 2021;55(4):1008-16.