Am J Transl Res 2023;15(7):4504-4520
www.ajtr.org /ISSN:1943-8141/AJTR0150570

Original Article

Computational identification of novel signature of T2DM-induced nephropathy and therapeutic bioactive compounds from *Azanza garckeana*

Bashir Lawal[1,2*], Yu-Cheng Kuo[3,4*], Sunday Amos Onikanni[5,6], Yi-Fong Chen[7], Tawakaltu Abdulrasheed-Adeleke[8], Adewale Oluwaseun Fadaka[9], Janet O Olugbodi[10], Halimat Yusuf Lukman[11], Femi Olawale[12], Mohamed H Mahmoud[13], Gaber El-Saber Batiha[14], Alexander TH Wu[15,16,17,18], Hsu-Shan Huang[19,20,21,22,23]

[1]UPMC Hillman Cancer Center, University of Pittsburgh, Pittsburgh, PA, USA; [2]Department of Pathology, University of Pittsburgh, Pittsburgh, PA 15213, USA; [3]Department of Pharmacology, School of Medicine, College of Medicine, Taipei Medical University, Taipei 11031, Taiwan; [4]School of Post-Baccalaureate Chinese Medicine, College of Chinese Medicine, China Medical University, Taichung 40402, Taiwan; [5]Department of Chemical Sciences, Biochemistry Unit, Afe-Babalola University, Ado-Ekiti, Ekiti State, Nigeria; [6]College of Medicine, Graduate Institute of Biomedical Sciences, China Medical University, Taiwan; [7]Division of Medicinal Products, Taiwan Food and Drug Administration, Ministry of Health and Welfare, No.161-2, Kunyang St., Nangang Dist., Taipei City 115209, Taiwan; [8]Department of Biochemistry, Federal University of Technology, P.M.B. 65, Minna 920001, Niger State, Nigeria; [9]Department of Science and Innovation (DSI)/Mintek Nanotechnology Innovation Centre (NIC) Biolabels Research Node, Department of Biotechnology, University of the Western Cape, Bellville 7535, South Africa; [10]Noble Life Sciences Inc. MD, USA; [11]Department of Chemical Sciences, Biochemistry Unit, College of Natural and Applied Sciences, Summit University, Offa, PMB 4412, Nigeria; [12]Department of Biochemistry, School of Life Science, University of KwaZulu Natal, Durban, South Africa; [13]Department of Biochemistry, College of Science, King Saud University, Riyadh, Saudi Arabia; [14]Department of Pharmacology and Therapeutics, Faculty of Veterinary Medicine, Damanhour University, Damanhour 22511, AlBeheira, Egypt; [15]TMU Research Center of Cancer Translational Medicine, Taipei Medical University, Taipei 11031, Taiwan; [16]The Ph.D. Program of Translational Medicine, College of Medical Science and Technology, Taipei Medical University, Taipei 11031, Taiwan; [17]Clinical Research Center, Taipei Medical University Hospital, Taipei Medical University, Taipei 11031, Taiwan; [18]Graduate Institute of Medical Sciences, National Defense Medical Center, Taipei 114, Taiwan; [19]Ph.D. Program for Cancer Molecular Biology and Drug Discovery, College of Medical Science and Technology, Taipei Medical University, and Academia Sinica, Taipei 11031, Taiwan; [20]Graduate Institute for Cancer Biology & Drug Discovery, College of Medical Science and Technology, Taipei Medical University, Taipei 11031, Taiwan; [21]Graduate Institute of Medical Sciences, National Defense Medical Centre, Taipei 11490, Taiwan; [22]School of Pharmacy, National Defense Medical Centre, Taipei 11490, Taiwan; [23]Ph.D. Program in Drug Discovery and Development Industry, College of Pharmacy, Taipei Medical University, Taipei 11031, Taiwan. [*]Equal contributors.

Received April 5, 2023; Accepted May 30, 2023; Epub July 15, 2023; Published July 30, 2023

Abstract: Objectives: Diabetic nephropathy (DN) is one of the most prevalent secondary complications associated with diabetes mellitus. Decades of research have implicated multiple pathways in the etiology and pathophysiology of diabetic nephropathy. There has been no reliable predictive biomarkers for the onset or progression of DN and no successful treatments are available. Methods: In the present study, we explored the datasets of RNA sequencing data from patients with Type II diabetes mellitus (T2DM)-induced nephropathy to identify a novel gene signature. We explored the target bioactive compounds identified from *Azanza garckeana*, a medicinal plant commonly used by the traditional treatment of diabetes nephropathy. Results: Our analysis identified lymphotoxin beta (LTB), SRY-box transcription factor 4 (SOX4), SOX9, and WAP four-disulfide core domain protein 2 (WFDC2) as novel signatures of T2DM-induced nephropathy. Additional analysis revealed the pathological involvement of the signature in cell-cell adhesion, immune, and inflammatory responses during diabetic nephropathy. Molecular docking and dynamic simulation at 100 ns conducted studies revealed that among the three compounds, Terpinen-4-ol exhibited higher binding efficacies (binding energies (ΔG) = -3.9~5.5 kcal/mol) against the targets. The targets, SOX4, and SOX9 demonstrated higher druggability towards the three compounds. WFDC2 was the least attractive target for the compounds. Conclusion: The present study was relevant in the diagnosis, prognosis, and treatment follow up of patients

with diabetes induced nephropathy. The study provided an insight into the therapeutic application of the bioactive principles from *Azanza garckeana*. Continued follow-up invitro validations study are ongoing in our laboratory.

Keywords: *Azanza garckeana*, dyslipidemia, hepatopathy, nephropathy, biochemical parameter

Introduction

Diabetes is one of the most common chronic metabolic diseases. Over 10.5% (536.6 million people) of the world's adult population are living with diabetes. It was estimated to rise to 12.2% (783.2 million) in 2045 [1]. The 2021 global expenditures of diabetes were estimated at 966 billion USD, and are projected to reach 1,054 billion USD by 2045 [1]. The global increase incidence of DM is concomitantly associated with increased diabetic complications, such as neuropathy, retinopathy, and nephropathy [2-5]. These complications demonstrate that diabetic nephropathy (DN) is the most destructive. It is the leading cause of end-stage renal disease in the developed world [6], and causes significant mortality and morbidity [4, 7]. Despite the substantial public health burden associated with DN and decades of deep research, a number of unmet needs remain. The molecular pathogenesis is poorly understood and there are no efficient therapeutic strategies to avoid, slow, or reverse the development and progression of diabetic nephropathy [8] other than regulation of the glucose levels, blood pressure, and lifestyle interventions which in most cases are less successful [9, 10]. Identifying a biomarker of genetic alteration that is associated with the susceptibility of DN in humans has been challenging [6, 11]. There is an urgent need for new biomarkers to stratify the risk of DN among patients with DM and to develop curative treatments strategies.

The recent increase in technology, integration of systems biology approaches, and evolving data from genomics studies, should enable proper biomarker identification [6]. In the present study, a unique NCI repository of clinical datasets of RNA sequencing data from tissue biopsies of type II diabetes patients with diabetes nephropathy were explored to identify a novel signature of diabetes nephropathy. We conducted subsequent bioinformatics analysis to identify genetic perturbation associated with this novel signature over the course of diabetes nephropathy.

Azanza garckeana a member of the Malvaceae family commonly known as Goron Tula has been extensively used as an herbal remedy for the treatment of diabetes and its associated complications [12]. Our previous studies have validated the antidiabetics, and attenuation effects of this plant extract against oxidative stress, inflammation, neuronal complications, and hepato-nephropathy in the experimental model of diabetes [13, 14]. Several other medicinal properties of this plant have been reported in the literature [15-18]. Considering the traditional reputation of this plant in the management of diabetes [19], and our experimental validation of its effectiveness, we evaluated the target bioactive compounds from extract of this plant against the novel signature using computational approaches including molecular docking and dynamics simulation. Molecular docking and simulation are structure-based modeling of binding interactions between a target protein molecule and a drug candidate [20-23]. Providing an insight into the behavior and mechanism of the drug activity in the binding site of target. This research provided an insight into novel signature of T2DM-induced nephropathy and its therapeutic potency of bioactive compounds from *Azanza garckeana*. The present study was relevant in the diagnosis, prognosis, and treatment follow up of patients with diabetes induced nephropathy. Follow-up invitro validations studies are ongoing in our laboratory.

Materials and methods

Acquisition of transcriptomic datasets and identification of differentially expressed genes (DEGs) associated with diabetic nephropathy

High-throughput sequencing and transcriptomic microarray-based datasets (GSE30122, GSE30529, GSE25724, and GSE30528) of diabetic hepatopathy/nephropathy and normal tissue samples (controls), were retrieved from the Gene Expression Omnibus database. The datasets respectively consisted of 19 and 50 (GSE30122); 10 and 12 (GSE30529); seven and six (GSE25724); and nine and 13 (GSE30528) samples from diabetic hepatopathic/nephropathic and normal tissues. The differentially expressed analysis package Lim-

ma was used for the analysis of DEGs between diabetic hepatopathic/nephropathic and normal tissues from each database. The threshold *value for DEGs was selected as a p value of <0.05 and |log2 fold change (FC)| of ≥1*.

Functional enrichment and pathways analysis of type 2 DM (T2DM) hub genes

A gene ontology (GO) function analysis was conducted to annotate DEGs from molecular functions (MFs) and biological processes (BPs). A Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis was conducted to annotate DEG pathways. Enrichment analyses were conducted using the Enrichr and STRING servers. All enriched terms were selected based on a false discovery rate (FDR)-adjusted *p* value of <0.05. Enrichment results were visualized using the enrichment plot module of ImageGP and are displayed in the form of dot plots.

Protein-protein interaction (PPI) analysis of T2DM hub genes

A PPI network of DEGs in T2DM was established using the Search Tool for the Retrieval of Interacting Genes (STRING) server version 11.5. DEGs were imported into multiple protein modules of the server, and the analysis was conducted based on software default parameters. The degree of connectivity of the DEGs at each node in the network was calculated by a connectivity analysis. Core genes in the interactions were defined as T2DM hub genes and were selected based on genes with a high node degree (≥20 node degrees).

Plant extraction and gas chromatography-mass spectrometry (GC-MS) analysis

The pulp of *Azanza garckeana* was collected in the month of June 2021 from Tula village, in Gombe State, Nigeria. It was authenticated at the Biological Science Department, Alex Ebonyi State, University Nigeria, and voucher identity was deposited. The pulp was air-dried, pulverized, and was extracted with methanol by maceration [18]. The methanol extracted was subsequently fractionated with hexane, chloroform, and ethyl-acetate. The ethyl-acetate fraction of *A. garckeana* was subjected to characterization using GCMS analysis [24]. The most abundant compounds were explored for therapeutic targeting the novel gene signature of T2DM-induced nephropathy in the silico model.

Ligand receptor interaction studies (molecular docking)

The crystal structure of lymphotoxin beta receptor (LTBR) was downloaded from PDB (ID: 4MXW). The grid box size was built of 200 × 150 × 126 points in the x, y, and z directions, and the center was located at x = -37.3, y = -22.6, and z = -248.7. The crystal structure of SRY box transcription factor 4 (SOX4) was downloaded from PDB (ID: 3U2B) [25]. The grid box size was built of 126 × 126 × 88 points in the x, y, and z directions, and the center was located at x = 14.4, y = -4.6, and z = -2.1. The crystal structure of SRY box transcription factor 9 (SOX9) was downloaded from PDB (ID: 4EUW). The grid box size was built of 126 × 70 × 120 points in the x, y, and z directions, and the center was located at x = -17.2, y = -20.9, and z = 14.4. WAP four-disulfide core domain protein 2 (WFDC2) was simulated by modeling. The three-dimensional (3D) structures of the ligands were obtained in mol2 format [26], and converted to PDB and subsequently PDBQT files using AutoDock Vina [27]. A blind docking protocol using AutoDock Vina (version 0.8) [27] default settings were adopted for molecular docking study. Protocols described in previous studies [22, 28-30] were adopted for ligand and receptor preparations prior to docking [30-32].

MD simulations and trajectory analysis

The molecular dynamic studies of the best docking posed complex (LTBR:Squalene, SOX4: Terpinen-4-ol, SOX9:Squalene, and WFDC2: 2,6-Di-tert-butyl-p-cresol) were conducted by using Schrödinger 2022 version 1 with Maestro version 13.1.137, MM share version 5.7.137, and Windows-x64 Platform as described in previous studies [33-37]. The MD analysis was done in replicate to avoid variation.

Results

Identification of a novel gene signature for diabetic nephropathy

Volcano plots of DEGs between diabetic nephropathic tissues and normal control tissues are shown in **Figure 1A**. Based on the cutoff criteria, totals of 181, 562, 73, and 167 upregulated DEGs in diabetic nephropathy were res-

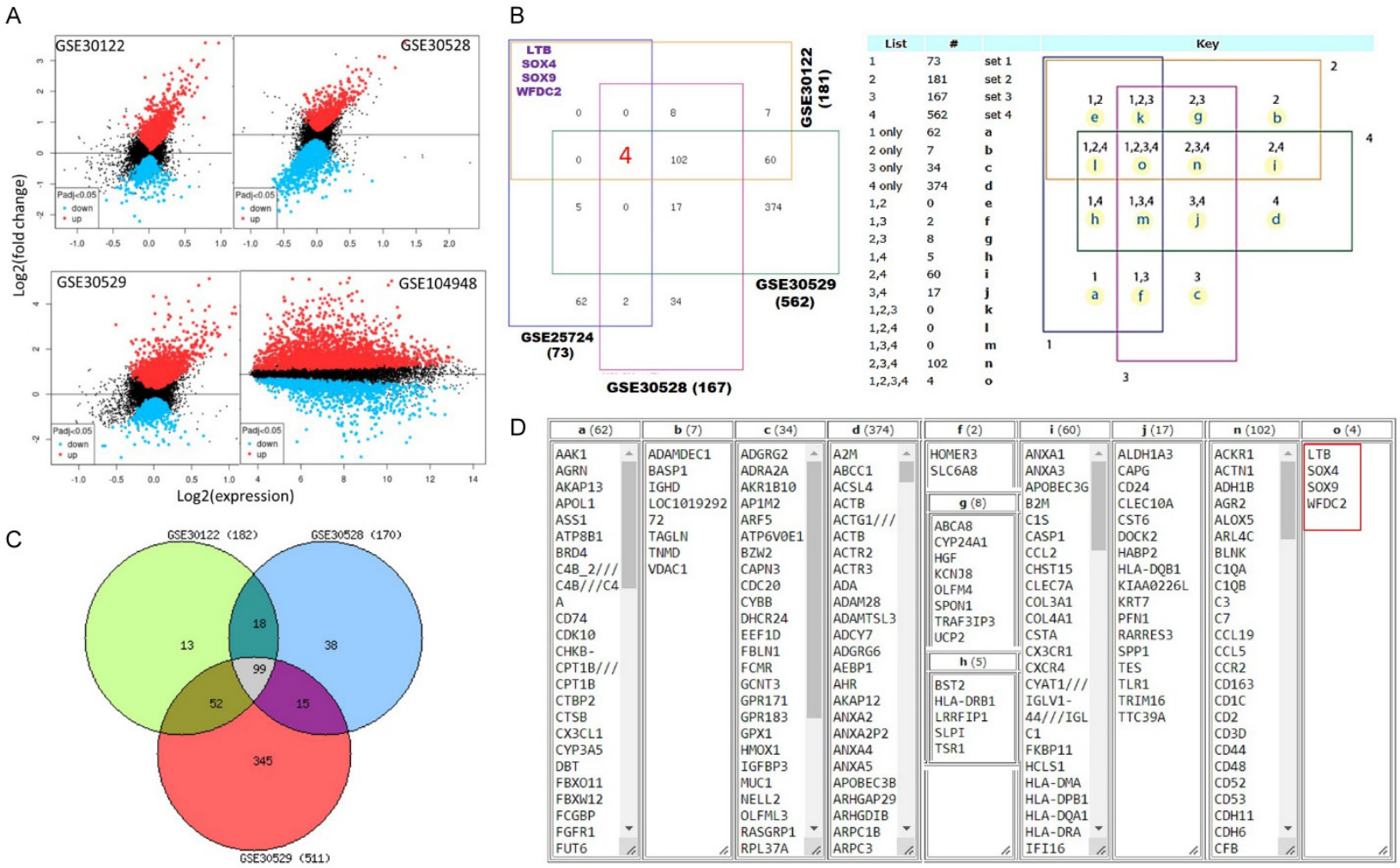

Figure 1. Identification of differentially expressed genes (DEGs) from transcriptomic datasets of diabetic hepatopathy/nephropathy. A. Volcano plots of DEGs between diabetic nephropathic tissues and normal control tissues. B. Venn diagram of DEGs integrated from the four datasets. C. Venn diagram of DEGs integrated from three datasets. D. Detailed gene list from each integration.

pectively identified from the GSE30122, GSE30529, GSE25724, and GSE30528 datasets. By integrating the DEGs from these datasets, four significant co-expressed DEGs (LTB, SOX4, SOX9, and WFDC2) were shared by the four datasets as shown in a Venn diagram (**Figure 1B**). The 99 DEGs were co-expressed in three of the datasets (GSE30122, GSE30528, and GSE30529; **Figure 1C**). The detailed gene list is provided in **Figure 1D**. These DEGs were subjected to subsequent network and enrichment analyses to identify T2DM hub genes and their roles in diabetes-induced hepatopathy/nephropathy.

PPI network and pathway enrichment of T2DM hub genes

The PPI network produced totals of 96 nodes, 481 edges, 10 average node degrees, 0.553 average local clustering coefficients, and a PPI enrichment *p* value of <1.0e-16. From PPI network interactions, 17 T2DM hub genes (≥20 node degrees) were identified, including *ITGB2, TYROBP, FN1, IL10RA, CCR2, C1QB, CD53, LY86, C1QA, PLEK, CD48, FYB, CD44, CCL5, CD2, IRF8*, and *LAPTM5* (**Figure 2A**). Enrichment analyses revealed that T2DM hub genes were enriched in KEGG pathways including complement and coagulation cascades, pertussis, extracellular matrix (ECM)-receptor interactions, phagosomes, focal adhesion, hematopoietic cell lineages, primary immunodeficiencies, systemic lupus erythematosus, cytokine-cytokine receptor interactions, cell adhesion molecules, rheumatoid arthritis, the Wnt signaling pathway, the nuclear factor (NF)-κB signaling pathway, the phosphatidylinositol 3-kinase (PI3K)-Akt signaling pathway, leukocyte transendothelial migration, and chemokine signaling pathway (**Figure 2B**). T2DM hub-gene enrichment processes included nephron tubule formation, kidney development and morphogenesis, and triglyceride metabolic processes associated with cell adhesion, immune, and inflammatory responses (**Figure 2C**, **2D**). Our analysis strongly suggested that pathological roles of T2DM hub genes in diabetic nephropathy were associated with cell-cell adhesion, immune, and inflammatory responses.

Characterization of previously validated anti-diabetic bioactive fraction (ethyl-acetate) fraction of A. garckeana

Based on our pilot *in vivo* study, the ethyl-acetate fraction of *A. garckeana* exhibited the most significant protection against diabetes-induced dyslipidemia, hepatopathy, and nephropathy. We characterized its bioactive constituents and revealed the presence of various bioactive constituents. The most abundant compounds that were present were 2,6-Di-tert-butyl-p-cresol (20.57%), spinacene (17.74%), and terpinen-4-ol (6.81%) and may have been responsible for the bioactivity of this fraction. Other compounds identified included; γ-Terpinene, linalool, limonene, caryophyllene, trans-α-Bergamotene, α-Farnesene, β-Bisabolene, 9,12-Octadecadienoic acid (Z,Z)-methyl ester, pentadecanoic acid, 14-methyl-, methyl ester, and methyl elaidate. The major compounds identified based on their molecular weight, molecular formula, retention time (RT), peak area percentage, and the class of bio activities is presented in **Table 1**. The chromatogram is shown in **Figure 3**. The compounds were subjected to molecular docking studies to analyze their targeting the novel T2DM hub-gene signature of diabetic nephropathy.

Exploring the therapeutic bioactive compounds for targeting the novel T2DM hub-gene signature of diabetic nephropathy

Our molecular docking analysis revealed that 2,6-Di-tert-butyl-p-cresol, spinacene and terpinen-4-ol docked to different interacting amino acid residues of the target proteins and at different binding affinities (**Figures 4**-**7**; **Tables 2**, **3**). The protein-ligand interaction profile of the spinacene, 2,6-Di-tert-butyl-p-cresol, and terpinene-4-ol against the target proteins are presented in **Table 3**. The targets, SOX4 and SOX9 demonstrated higher druggability towards the three compounds. WFDC2 was the least attractive target for the compounds (**Table 2**). The interaction of LTB involved several Van der Waals forces, conventional hydrogen bond, alkyl, Pi-pi stacked, and Pi-sigma interaction with spinacene (ARG61, GLU90, HIS91, TRP92, ASN93, THR96, GLN99, TYR94, LEU95, ILE97), 2,6-Di-tert-butyl-p-cresol (CYS43, GLN46, TYR50, TYR94, TYR51, PRO53), and Terpinen-4-ol (GLN46, GLU49, TYR50, ARG61, TYR51, TYR94, TYR51). The higher Van der wall forces can be attributed to the higher binding affinity of LTB to spinacene (-4.7 kcal/mol), and 2,6-Ditert-butyl-p-cresol (-4.6 kcal/mol) when compared to Terpinen-4-ol which has the least binding affinity (-3.9 kcal/mol).

SOX4 interacted by its AA residues; MET7, ASN8, LYS49, GLU57, ARG60, ALA9, PHE10,

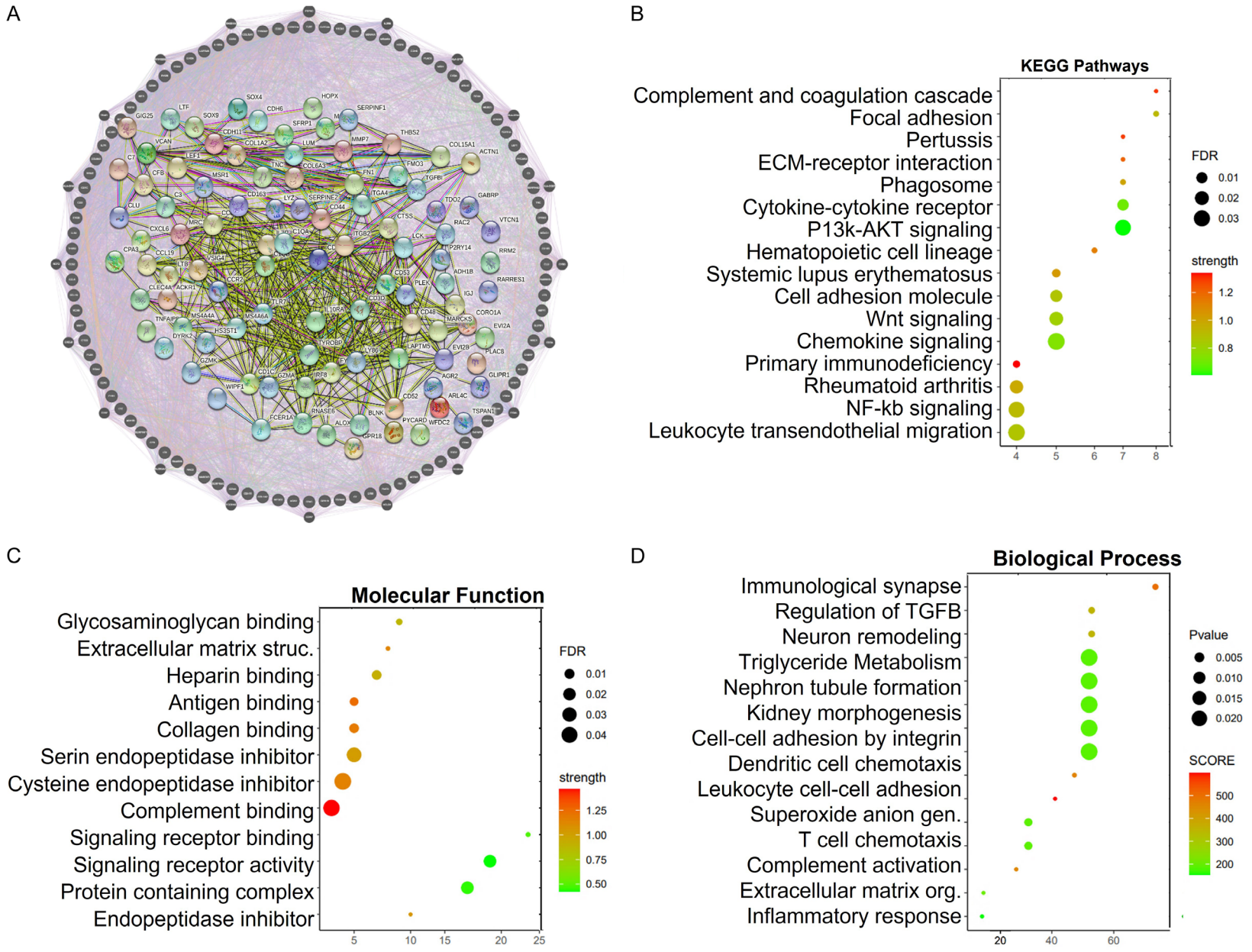
A
B
KEGG Pathways
Complement and coagulation cascade
Focal adhesion
Pertussis
ECM-receptor interaction
Phagosome
Cytokine-cytokine receptor
P13k-AKT signaling
Hematopoietic cell lineage
Systemic lupus erythematosus
Cell adhesion molecule
Wnt signaling
Chemokine signaling
Primary immunodeficiency
Rheumatoid arthritis
NF-kb signaling
Leukocyte transendothelial migration
4
5
6
7
8
FDR
0.01
0.02
0.03
strength
1.2
1.0
0.8
C
Molecular Function
Glycosaminoglycan binding
Extracellular matrix struc.
Heparin binding
Antigen binding
Collagen binding
Serin endopeptidase inhibitor
Cysteine endopeptidase inhibitor
Complement binding
Signaling receptor binding
Signaling receptor activity
Protein containing complex
Endopeptidase inhibitor
5
10
15
20
25
FDR
0.01
0.02
0.03
0.04
strength
1.25
1.00
0.75
0.50
D
Biological Process
Immunological synapse
Regulation of TGFB
Neuron remodeling
Triglyceride Metabolism
Nephron tubule formation
Kidney morphogenesis
Cell-cell adhesion by integrin
Dendritic cell chemotaxis
Leukocyte cell-cell adhesion
Superoxide anion gen.
T cell chemotaxis
Complement activation
Extracellular matrix org.
Inflammatory response
20
40
60
Pvalue
0.005
0.010
0.015
0.020
SCORE
500
400
300
200

Figure 2. Functional enrichment and network pathways analysis of type 2 diabetes mellitus (T2DM) hub genes. (A) The constructed protein-protein interaction (PPI) network. (B) Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment. (C) Gene ontology (GO) molecular functions and (D) biological processes of T2DM hub genes.

Table 1. Compounds identified from the gas chromatography-mass spectrometry (GC-MS) analysis of the ethyl-acetate fraction of *Azanza garckeana*

RT	Area%	Compound	MF	MW (g/mol)	Classes of compounds
3.663	0.27	Limonene	$C_{10}H_{16}$	136.23	monoterpene
3.995	0.12	γ-Terpinene	$C_{10}H_{16}$	136.23	monoterpene
4.545	0.65	Linalool	$C_{10}H_{18}O$	154.25	monoterpene
5.723	6.81	4-Carvomenthenol	$C_{10}H_{18}O$	154.25	monoterpenes
8.192	0.64	Eugenol	$C_{10}H_{12}O_2$	164.2	terpenes
9.049	0.82	Caryophyllene	$C_{15}H_{24}$	204.35	sesquiterpene
9.215	0.75	trans-α-Bergamotene	$C_{15}H_{24}$	204.35	sesquiterpene
9.915	0.56	α-Farnesene	$C_{15}H_{24}$	204.35	sesquiterpenes
10.117	0.55	β-Bisabolene	$C_{15}H_{24}$	204.35	sesquiterpenes
10.185	20.57	2,6-Di-tert-butyl-p-cresol	$C_{15}H_{24}O$	220.35	phenol
14.652	3.69	Pentadecanoic acid, 14-methyl-, methyl ester	$C_{17}H_{34}O_2$	270.5	fatty acid ester
15.005	0.06	n-Hexadecanoic acid	$C_{16}H_{32}O_2$	256.42	fatty acid
16.271	3.22	9,12-Octadecadienoic acid (Z,Z)-methyl ester	$C_{19}H_{34}O_2$	294.479	fatty acid ester
16.323	2.96	Methyl elaidate	$C_{19}H_{36}O_2$	296.5	fatty acid ester
16.551	0.92	Methyl stearate	$C_{19}H_{36}O_2$	298.504	fatty acid
18.409	17.74	Spinacene	$C_{30}H_{50}$	410.7	triterpene

RT = retention time; MF = molecular formula; MW = molecular weight; PA = peak area.

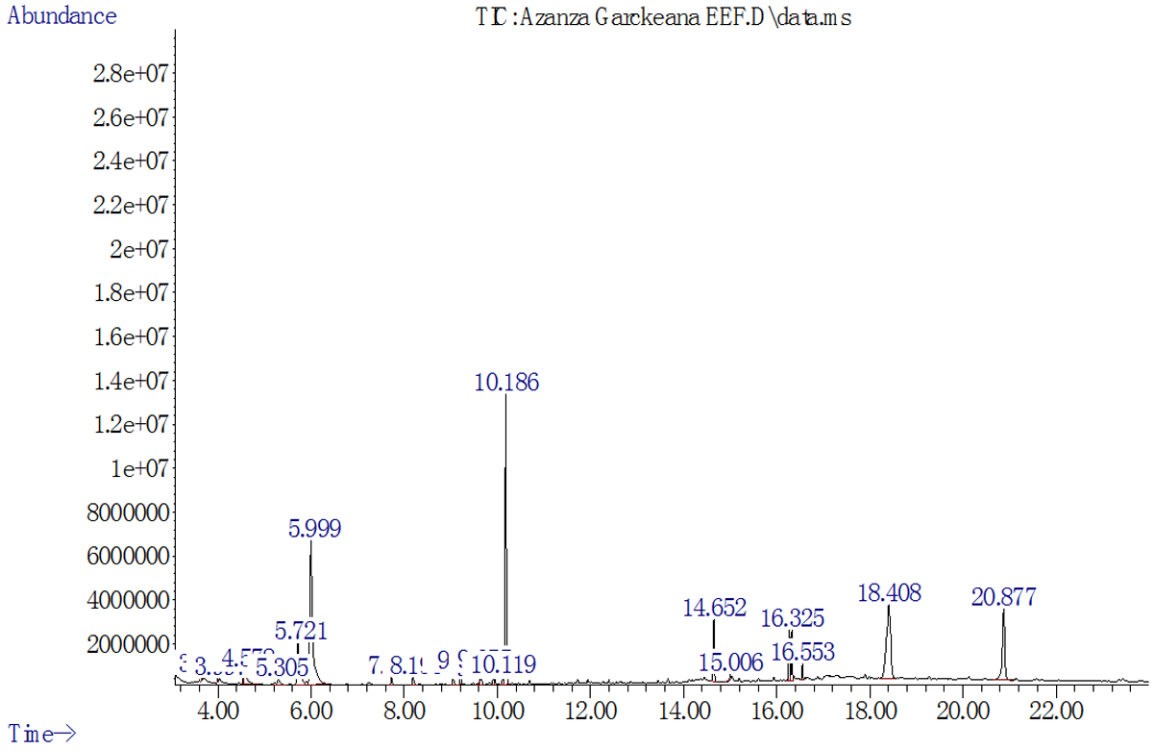

Figure 3. Gas chromatography-mass spectrometry (GC-MS) chromatogram of the ethyl-acetate fraction of *Azanza garckeana*.

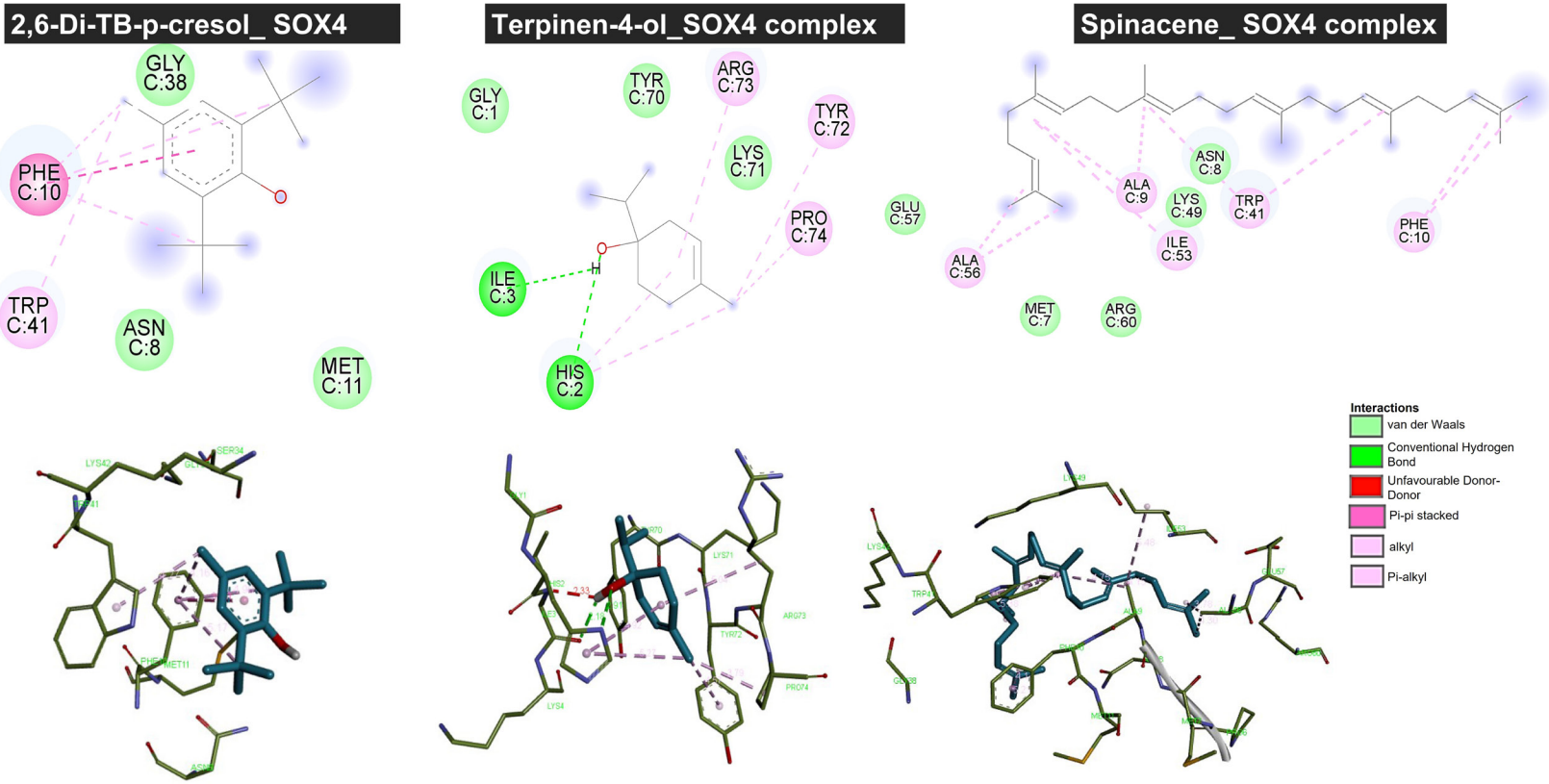

Figure 4. Three-dimensional (3D) and two-dimensional (2D) views of receptor-ligand interactions between the protein target SRY-box transcription factor 4 (SOX4) and ligands: terpinen-4-ol, 2,6-Di-tert-butyl-p-cresol, and spinacene.

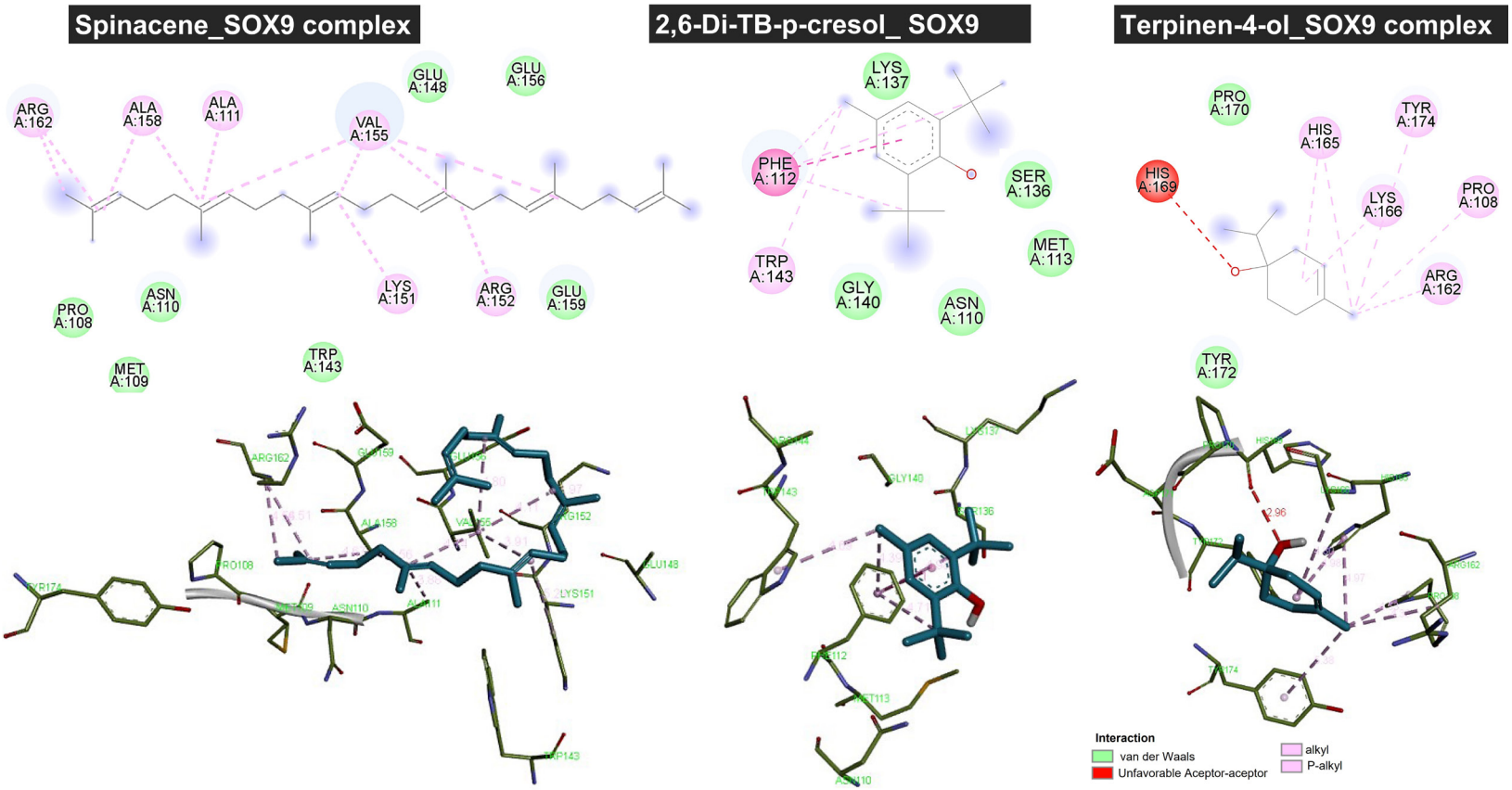

Figure 5. The three (3)-dimensional (3D) and two (2)-dimensional view of receptor-ligand interaction between the protein target; SOX9 (SRY-Box Transcription Factor 9), and ligands; terpinen-4-ol, 2,6-Di-tert-butyl-p-cresol, and spinacene.

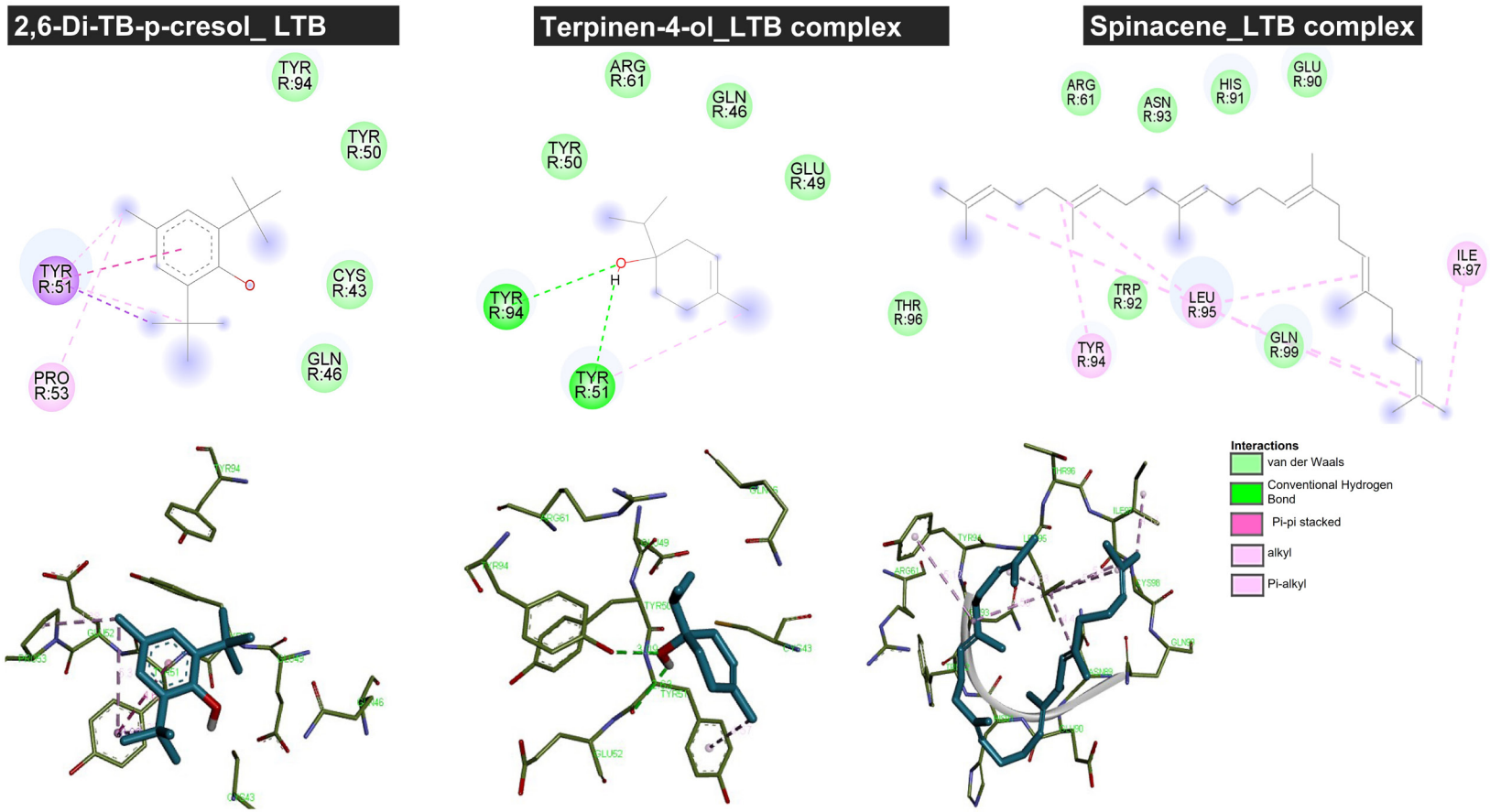

Figure 6. Three-dimensional (3D) and two-dimensional (2D) views of receptor-ligand interactions between the protein target lymphotoxin beta (LTB) and ligands: terpinen-4-ol, 2,6-Di-tert-butyl-p-cresol, and spinacene.

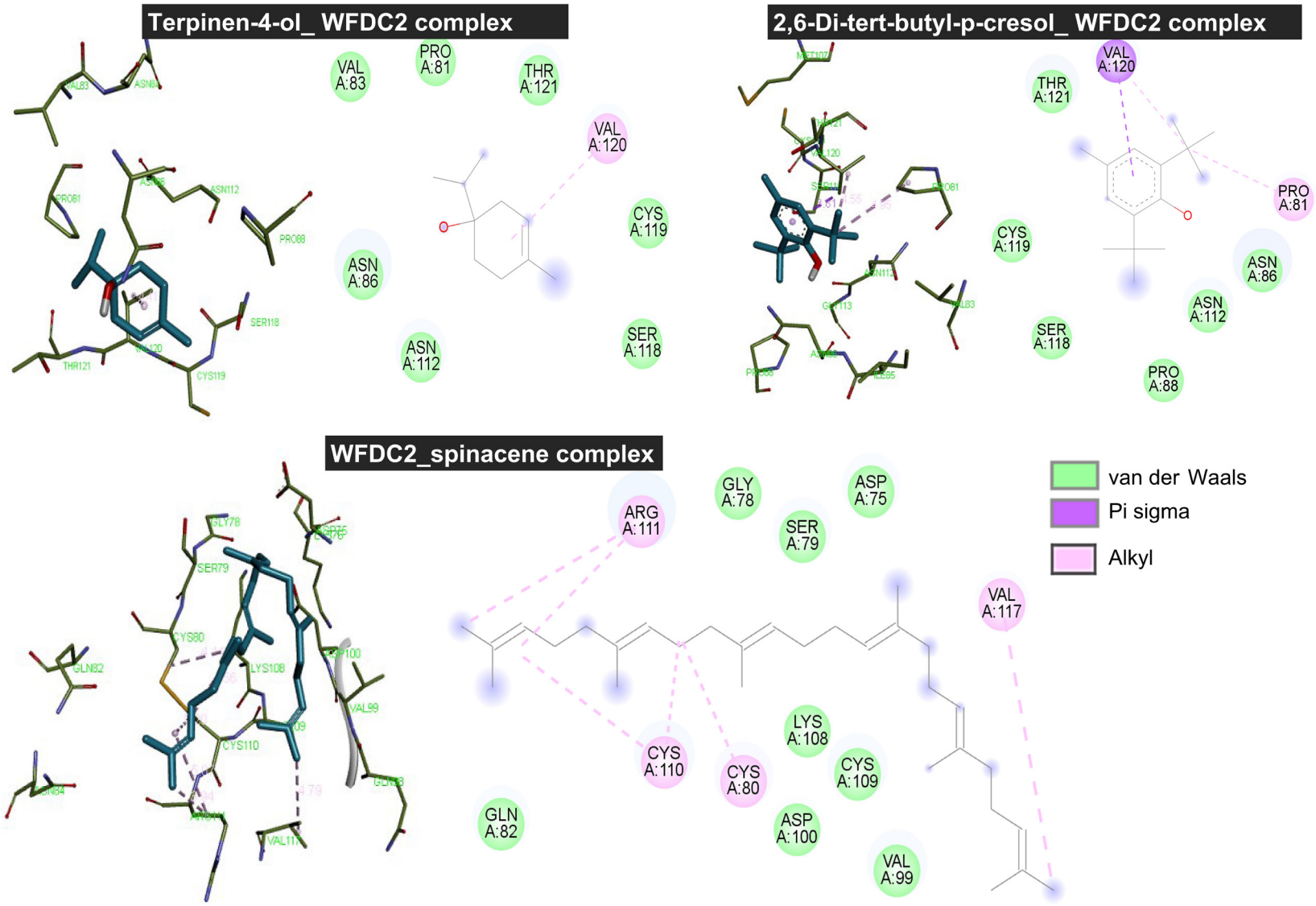

Figure 7. Three-dimensional (3D) and two-dimensional (2D) views of receptor-ligand interactions between the protein target WAP four-disulfide core domain 2 (WFDC2) and ligands: terpinen-4-ol, 2,6-Di-tert-butyl-p-cresol, and spinacene.

Table 2. Molecular docking-based binding affinities (ΔG) of bioactive compounds from the ethyl-acetate fraction of *Azanza garckeana* for targeting the novel type 2 diabetes mellitus hub-gene signature of diabetic nephropathy

Compound	Compound CID	MF	MW (g/mol)	Abundance (%)	ΔG (kcal/mol)			
					SOX4	SOX9	LTB	WFDC2
2,6-Di-tert-butyl-p-cresol	CID: 31404	$C_{15}H_{24}O$	220.35	20.57	-4.7	-5.0	-4.6	-4.8
Spinacene	CID: 638072	$C_{30}H_{50}$	410.7	17.74	-4.5	-5.2	-4.7	-3.6
Terpinen-4-ol	CID: 11230	$C_{10}H_{18}O$	154.25	6.81	-5.5	-4.5	-3.9	-4.2

MF, molecular formula; MW, molecular weight; SOX4, SYR boxtranscription factor 4; LTB, lymphotoxin beta; WFDC2, WAP four-disulfide core domain 2. The lowest free binding energies were calculated by the AutoDock Vina program.

Table 3. The protein-ligand interaction profile of analysis of the spinacene, 2,6-Di-tert-butyl-p-cresol, and terpinene-4-ol against the target proteins

Targets/ Ligands	Interacting amino acid residues		
	Squalene	2,6-Di-tert-butyl-p-cresol	Terpinen-4-ol
LTBR	[a](ARG61, GLU90, HIS91, TRP92, ASN93, THR96, GLN99), [c](TYR94, LEU95, ILE97)	[a](CYS43, GLN46, TYR50, TYR94), [c](TYR51, PRO53)	[a](GLN46, GLU49, TYR50, ARG61), [b](TYR51, TYR94), [c](TYR51)
SOX4	[a](MET7, ASN8, LYS49, GLU57, ARG60), [c](ALA9, PHE10, TRP41, ILE53, ALA56)	[a](ASN8, MET11, GLY38), [c](PHE10, TRP41)	[a](GLY1, TYR70, LYS71), [b](HIS2, ILE3), [c](TYR72, ARG73, PRO74), [d](ILE3)
SOX9	[a](PRO108, MET109, ASN110, TRP143, GLU148, GLU156, GLU159), [c](ALA111, LYS151, ARG152, VAL155, ALA158, ARG162)	[a](ASN110, MET113, SER136, LYS137, GLY140), [c](PHE112, TRP143)	[a](PRO170, TYR172), [c](PRO108, ARG162, HIS165, LYS166, TYR174), [d](HIS169)
WFDC2	[a](ASP75, GLY78, SER79, GLN82, VAL99, ASP100, LYS108, CYS109), [c](CYS80, CYS110, ARG111, VAL117)	[a](ASN86, PRO88, ASN112, SER118, CYS119, THR121), [c](PRO81, VAL120)	[a](PRO81, VAL83, ASN86, ASN112, SER118, CYS119, THR121), [c](VAL120)

[a]Van der Waals; [b]Conventional hydrogen bond; [c]Alkyl or Pi-alkyl, Pi-pi stacked, Pi-sigma; [d]Others.

TRP41, ILE53, and ALA56 (spinacene), ASN8, MET11, GLY38, PHE10, TRP41 (2,6-Di-tert-butyl-p-cresol), and terpinene-4-ol (GLY1, TYR-70, LYS71, HIS2, ILE3, TYR72, ARG73, PRO74, and ILE3). These interactions, resulted in better binding efficacy of terpinene-4-ol (-5.5 kcal/mol) to SOX4 when compared to other ligands (**Tables 2**, **3**). The interacting residues of WFDC2 and SOX9 are presented in **Table 3**. SOX4 demonstrated a higher binding sensitivity to the ligands (-4.5 to -5.5 kcal/mol). WFDC2 (-3.6 to -4.8 kcal/mol) exhibited the least interaction affinities (**Table 2**).

Molecular dynamic simulation of ligand-target complexes

The molecular dynamic simulation of ligand-target complex was conducted. In this simulation study, the RMSD from **Figure 8** illustrated that the SOX4-Terpinen-4-ol complex had a higher RMSD. It displayed large fluctuations in RMSD trend, indicating complex flexibility. SOX9-Squalene and WFDC2-2,6-Di-tert-butyl-p-cresol complexes show a similar trend of RMSD with little fluctuation patterns that tend to increase at 0-30 ns and between 60-100 ns. LTBR-Squalene complex demonstrated higher stability with very little deviation at the early stage of simulation (0-20 ns) after which attained equilibrium until the end of the simulation period (**Figure 7**). This desirable RMSD suggests the stability of the ligand in the target protein cavity. The degree of mobility in a biological system can be indicated by the Rg profile. The WFDC2-2,6-Di-tert-butyl-p-cresol, and SOX4-Terpinen-4-ol had the smallest rGyr mean value of 2.91, and 2.44 respectively and were stable from the onset and throughout the 100-ns simulation period. LTBR-Squalene complex and SOX9-Squalene complex had a high mean Rg pattern which fluctuated throughout the simulation period. Our SASA analysis showed that LTBR-Squalene complex exhibited the

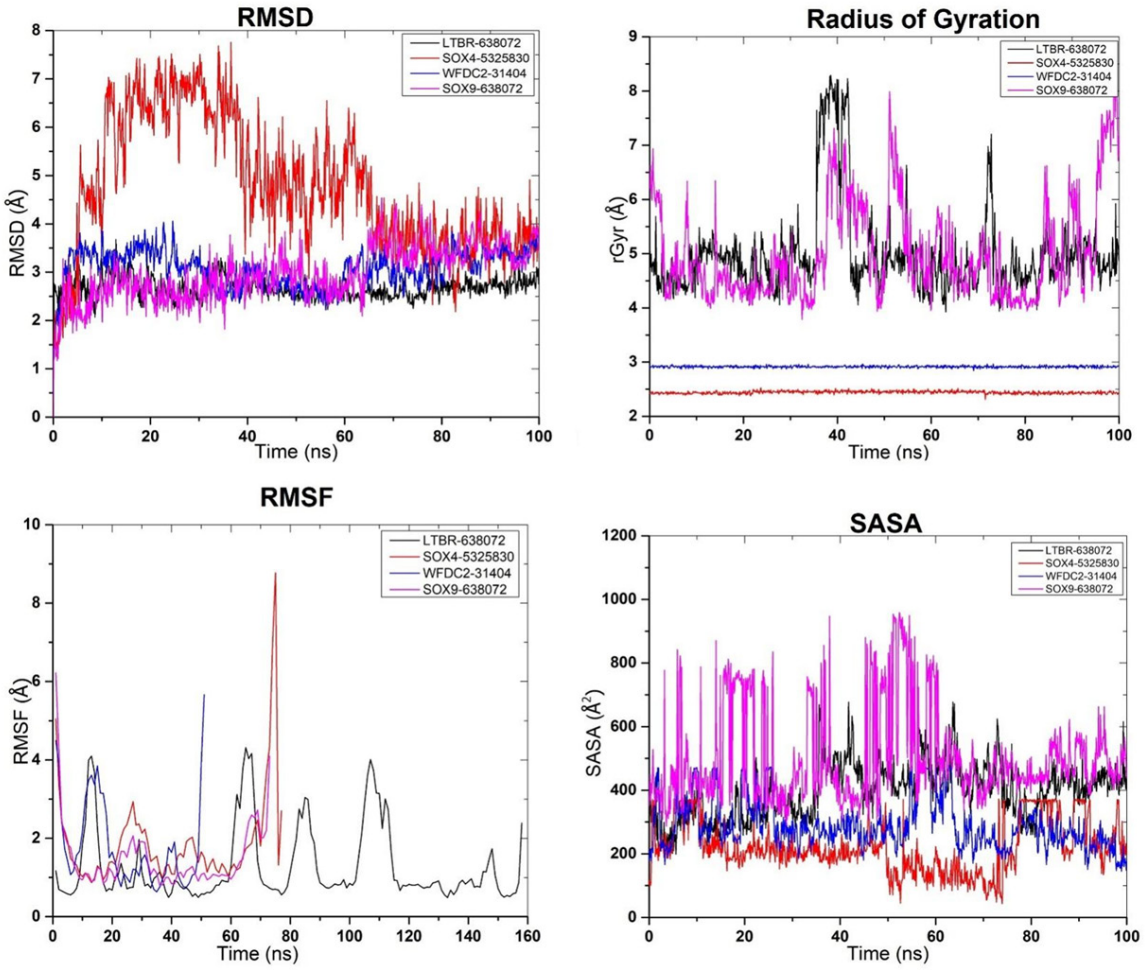

Figure 8. MD analyses of LTBR-638072 (LTBR-Squalene complex), SOX4-5325830 (SOX4-Terpinen-4-ol complex), SOX9-638072 (SOX9-Squalene) and WFDC2-31404 (WFDC2-2,6-Di-tert-butyl-p-cresol complex) showing the RMSD graphical illustration, RMSF plot, rGyr representation, and SASA diagram. All MD simulations were performed by using Schrödinger version 2022_1.

most stable SASA profiles with mild deviation demonstrating a rigid and stable profile. The high fluctuating trend in SASA observed in SOX9-Squalene and SOX4-Terpinen-4-ol complex suggested protein expansion and loose binding of receptor to the ligand. LTBR-Squalene complex had the least RMSF value of 1.27 (**Table 1**) suggesting its rigidity. The higher mean RMSF of SOX4-Terpinen-4-ol complex imply more flexibility throughout the MD simulation.

Discussion

Diabetic nephropathy (DN) is one of the most prevalent and most destructive secondary complications associated with diabetes mellitus. It is a leading cause of end-stage renal disease in the developed world [6] causing significant mortality and morbidity [4, 7]. Decades of research have implicated multiple pathways in the etiology and pathophysiology of diabetic nephropathy. There has been no reliable predictive biomarkers for the onset or progression of DN and no successful treatments are available. This unmet clinical need calls for an urgent need for new biomarkers to stratify risk of DN among patients with DM and to develop appropriate therapeutic strategies.

The present study takes a crucial step towards achieving this goal by identifying a novel gene signature whose deregulatory expression precisely stratifies patient samples with diabetic nephropathy progression and by analyzing their pathways interactions [9]. Identifying this novel

signature by integrating deregulatory genes from multiple RNA sequencing datasets from patients with T2DM induced nephropathy will lead to results providing novel insight into the pathogenesis of DN, and identify biomarkers of DN as a therapeutic target. Our initial analyses of these datasets identified a total of 181, 562, 73, and 167 upregulated DEGs that can account for the differences in their clinical course of the different datasets. Integration of these DEGs resulted in a four significant co-expressed DEGs (LTB, SOX4, SOX9, and WFDC2) by the four datasets of DN.

The novel gene signature in DN were enriched with signaling pathways involved in 'cell-cell adhesion', 'immune response' and 'inflammatory response'. The enriched cytokine-cytokine receptor interaction observed with the hub genes have been previously implicated in the initiation and progression processes of diabetic nephropathy [38]. These immune response and inflammation genes, WFDC2, LTB, and SOX9 are of particular interest. WFDC2 contributes to EMT by activating AKT signaling pathway and regulating MMP-2 expression [39]. It has been reported to regulate immune suppression [40]. Nakagawa et al. [41] identified SOX9 and WFDC2 as a molecular markers of tubulointerstitial fibrosis and tubular cell damage in patients with chronic kidney disease. A microarray analysis revealed elevated kidney expression levels of WFDC2 in patients who undergo renal transplant [42]. Increased expression of WFDC2 in myofibroblasts from mouse kidneys has been reported to mediate renal fibrosis [43]. It has been reported that the increased expression of SOX9 correlated with the extent of histopathology detected in renal biopsies [41]. LTβR can activate NF-κB pathways that promote renal inflammation and has been regarded as a therapeutic target in renal inflammation [44]. These previous findings strengthen our findings that WFDC2 and SOX9 were associated with the progression of DN and that target inhibition can provide attractive strategies for the treatment of diabetes nephropathy. The results of the present study provide vital information for the development of novel diagnostic tools and therapeutic agents to predict and prevent DN. More experimental studies are needed to determine how the novel gene signature influences the initiation and progression of DN.

We explored the target of bioactive compounds identified from *Azanza garckeana*, a medicinal plant commonly used by the traditional treatment of diabetes. Molecular docking is a structure-based modeling of binding interactions between a target protein molecule and a ligand small molecule for targeted inhibition [20-23]. It provides an estimate of the binding efficacy of the drug candidate to the target molecule, and hints at the behavior of small-molecule drugs in the binding site of target proteins and possible mechanistic roles of drug candidates [45-49].

Our molecular docking analysis revealed that among the three compounds, Terpinen-4-ol exhibited higher binding efficacies (binding energies (ΔG) = -3.9~5.5 kcal/mol) against the targets. The targets, SOX4 and SOX9, demonstrated higher druggability towards the three compounds. WFDC2 was the least attractive target for the compounds (**Table 2**). Analysis of the ligand interaction complex revealed that the 2,6-Di-tert-butyl-p-cresol, spinacene, and terpinen-4-ol were bound to the targets by alkyl, van der Waals, hydrophobic, and pi interactions. The molecular docking and dynamic analysis suggested the therapeutic bioactive compounds particularly Terpinen-4-ol from the ethyl-acetate fraction of *A. garckeana* for targeting the novel T2DM hub-gene signature of diabetic nephropathy. The results of the present study strongly suggest the application of *A. garckeana* for the effective treatment of diabetic nephropathy. The ethyl-acetate fraction of this plant represents a reserve of candidates for developing new drugs.

The present study was relevant in the diagnosis, prognosis, and treatment follow up of patients with diabetes induced nephropathy. The study provided an insight into the therapeutic application of the bioactive principles from *Azanza garckeana.* Follow-up invitro validations studies are on-going in our laboratory.

Conclusions

In conclusion, the present study identified a novel signature of T2DM-induced nephropathy which was relevant in the diagnosis, prognosis, and treatment follow up of patients with diabetes induced nephropathy. Molecular docking and dynamics simulation revealed the affinities and therapeutic application of bioactive compounds from *Azanza garckeana*. Follow-up invitro validations studies are on-going in our laboratory.

Acknowledgements

The National Science and Technology Council, Taiwan, grant number NSTC111-2314-B-038-017, the Ministry of Science and Technology, Taiwan, grant number MOST111-2314-B-038-122, and the Shin Kong Wu Ho-Su Memorial Hospital SKH-TMU-112-02 awarded to H.-S. Huang. ATH Wu is also funded by The National Science and Technology Council, Taiwan, grant numbers 111-2314-B-038-098 and 111-2314-B-038-142. The authors would like to extend their gratitude to King Saud University (Riyadh, Saudi Arabia) for the partial funding of this research through Researchers Supporting Project number (RSP-2023-R406).

Disclosure of conflict of interest

None.

Address correspondence to: Hsu-Shan Huang, Ph.D Program for Cancer Molecular Biology and Drug Discovery, College of Medical Science and Technology, Taipei Medical University and Academia Sinica, Taipei 11031, Taiwan. E-mail: huanghs99@tmu.edu.tw; Alexander TH Wu, The Ph.D Program of Translational Medicine, College of Medical Science and Technology, Taipei Medical University, Taipei 11031, Taiwan. E-mail: chaw1211@tmu.edu.tw

References

[1] Sun H, Saeedi P, Karuranga S, Pinkepank M, Ogurtsova K, Duncan BB, Stein C, Basit A, Chan JCN, Mbanya JC, Pavkov ME, Ramachandaran A, Wild SH, James S, Herman WH, Zhang P, Bommer C, Kuo S, Boyko EJ and Magliano DJ. IDF diabetes atlas: global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes Res Clin Pract 2022; 183: 109119.

[2] Retnakaran R, Cull CA, Thorne KI, Adler AI and Holman RR; UKPDS Study Group. Risk factors for renal dysfunction in type 2 diabetes: U.K. prospective diabetes study 74. Diabetes 2006; 55: 1832-1839.

[3] Tantigegn S, Ewunetie AA, Agazhe M, Aschale A, Gebrie M, Diress G and Alamneh BE. Time to diabetic neuropathy and its predictors among adult type 2 diabetes mellitus patients in Amhara regional state comprehensive specialized hospitals, Northwest Ethiopia, 2022: a retrospective follow up study. PLoS One 2023; 18: e0284568.

[4] Sulaiman MK. Diabetic nephropathy: recent advances in pathophysiology and challenges in dietary management. Diabetol Metab Syndr 2019; 11: 7.

[5] Valencia WM and Florez H. How to prevent the microvascular complications of type 2 diabetes beyond glucose control. BMJ 2017; 356: i6505.

[6] Azushima K, Gurley SB and Coffman TM. Modelling diabetic nephropathy in mice. Nat Rev Nephrol 2018; 14: 48-56.

[7] Zhang J, Liu J and Qin X. Advances in early biomarkers of diabetic nephropathy. Rev Assoc Med Bras (1992) 2018; 64: 85-92.

[8] Johnson SA and Spurney RF. Twenty years after ACEIs and ARBs: emerging treatment strategies for diabetic nephropathy. Am J Physiol Renal Physiol 2015; 309: F807-820.

[9] Hur J, Sullivan KA, Pande M, Hong Y, Sima AA, Jagadish HV, Kretzler M and Feldman EL. The identification of gene expression profiles associated with progression of human diabetic neuropathy. Brain 2011; 134: 3222-3235.

[10] Thomas MC, Cooper ME and Zimmet P. Changing epidemiology of type 2 diabetes mellitus and associated chronic kidney disease. Nat Rev Nephrol 2016; 12: 73-81.

[11] Bowden DW and Freedman BI. The challenging search for diabetic nephropathy genes. Diabetes 2012; 61: 1923-1924.

[12] Bioltif YE, Edward NB and Tyeng TD. A chemical overview of Azanza garckeana. Biology Medicine & Natural Product Chemistry 2020; 9: 91-95.

[13] Lawal B, Sani S, Onikanni AS, Ibrahim YO, Agboola AR, Lukman HY, Olawale F, Jigam AA, Batiha GE, Babalola SB, Mostafa-Hedeab G, Lima CMG, Wu ATH, Huang HS and Conte-Junior CA. Preclinical anti-inflammatory and antioxidant effects of Azanza garckeana in STZ-induced glycemic-impaired rats, and pharmacoinformatics of it major phytoconstituents. Biomed Pharmacother 2022; 152: 113196.

[14] Alozieuwa UB, Lawal B, Sani S, Onikanni AS, Osuji O, Ibrahim YO, Babalola SB, Mostafa-Hedeab G, Alsayegh AA, Albogami S, Batiha GE, Wu ATH, Huang HS and Conte-Junior CA. Luteolin-rich extract of thespesia garckeana F. Hoffm. (Snot Apple) contains potential drug-like candidates and modulates glycemic and oxidoinflammatory aberrations in experimental animals. Oxid Med Cell Longev 2022; 2022: 1215097.

[15] Yusuf AA, Lawal B, Sani S, Garba R, Mohammed BA, Oshevire DB and Adesina DA. Pharmacological activities of Azanza garckeana (Goron Tula) grown in Nigeria. Clinical Phytoscience 2020; 6: 1-8.

[16] Dikko Y, Khan M, Tor-Anyiin T, Anyam J and Linus U. In vitro antimicrobial activity of fruit pulp extracts of Azanza garckeana (F. hoffm.)

Exell & Hillc. and isolation of one of its active principles, betulinic acid. Methodology 2016.
[17] Michael K, Onyia L and Jidauna S. Evaluation of phytochemicals in Azanza garckeana (Gorontula) seed. Journal of Agriculture and Veterinary Science 2015; 8: 71-74.
[18] Itodo JI, Aluwong T, Uchendu C, Ogbuagu NE, Bugau JS, Adewuyi BA, Samuel FU, Abah KO, Shinkut M and Ogbuagu PK. The comparative effects of azanza garckeana fruit pulp and melatonin on serum and testicular oxidative stress changes, live sperm cells and spermatozoa abnormalities evoked by chronic administration of bisphenol A in rabbit bucks. Veterinary Integrative Sciences 2023; 21: 523-43.
[19] Mwove JK. Potential for development of novel food products from Azanza garckeana tree fruit: a review. International Journal of Food Studies 2021; 10: 121-132.
[20] Molecular docking of bioactive compounds against BRCA and COX proteins. Prog Drug Res 2016; 71: 181-183.
[21] Yeh YC, Lawal B, Hsiao M, Huang TH and Huang CF. Identification of NSP3 (SH2D3C) as a prognostic biomarker of tumor progression and immune evasion for lung cancer and evaluation of organosulfur compounds from Allium sativum L. as therapeutic candidates. Biomedicines 2021; 9: 1582.
[22] Wu SY, Lin KC, Lawal B, Wu ATH and Wu CZ. MXD3 as an onco-immunological biomarker encompassing the tumor microenvironment, disease staging, prognoses, and therapeutic responses in multiple cancer types. Comput Struct Biotechnol J 2021; 19: 4970-4983.
[23] Lawal B, Wang YC, Wu ATH and Huang HS. Prooncogenic c-Met/EGFR, biomarker signatures of the tumor microenvironment are clinical and therapy response prognosticators in colorectal cancer, and therapeutic targets of 3-phenyl-2H-benzo[e][1,3]-oxazine-2,4(3H)-dione derivatives. Front Pharmacol 2021; 12: 691234.
[24] Ndako M, Jigam AA, Kabiru AY, Umar SI and Lawal B. Polar extracts from Gymnosporia senegalensis (syn. Maytenus senegalensis) root bark, its effects on nociception, edema, and malarial infection. Phytomed Plus 2021; 1: 100113.
[25] Jauch R, Ng CK, Narasimhan K and Kolatkar PR. The crystal structure of the Sox4 HMG domain-DNA complex suggests a mechanism for positional interdependence in DNA recognition. Biochem J 2012; 443: 39-47.
[26] Hanwell MD, Curtis DE, Lonie DC, Vandermeersch T, Zurek E and Hutchison GR. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. J Cheminform 2012; 4: 17.
[27] Trott O and Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 2010; 31: 455-461.
[28] Lawal B, Liu YL, Mokgautsi N, Khedkar H, Sumitra MR, Wu ATH and Huang HS. Pharmacoinformatics and preclinical studies of nsc765690 and nsc765599, potential stat3/cdk2/4/6 inhibitors with antitumor activities against nci60 human tumor cell lines. Biomedicines 2021; 9: 92.
[29] Lawal B, Lee CY, Mokgautsi N, Sumitra MR, Khedkar H, Wu ATH and Huang HS. mTOR/EGFR/iNOS/MAP2K1/FGFR/TGFB1 are druggable candidates for N-(2, 4-difluorophenyl)-2', 4'-difluoro-4-hydroxybiphenyl-3-carboxamide (NSC765598), with consequent anticancer implications. Front Oncol 2021; 11: 656738.
[30] Abdulrasheed-Adeleke T, Lawal B, Agwupuye EI, Kuo Y, Eni AM, Ekoh OF, Lukman HY, Onikanni AS, Olawale F, Saidu S and Ibrahim YO. Apigetrin-enriched Pulmeria alba extract prevents assault of STZ on pancreatic β-cells and neuronal oxidative stress with concomitant attenuation of tissue damage and suppression of inflammation in the brain of diabetic rats. Biomed Pharmacother 2023; 162: 114582.
[31] Lawal B, Wu AT, Chen CH, George TA and Wu SY. Identification of INFG/STAT1/NOTCH3 as γ-Mangostin's potential targets for overcoming doxorubicin resistance and reducing cancer-associated fibroblasts in triple-negative breast cancer. Biomed Pharmacother 2023; 163: 114800.
[32] Olugbodi JO, Samaila K, Lawal B, Anunobi OO, Baty RS, Ilesanmi OB and Batiha GE. Computational and preclinical evidence of anti-ischemic properties of L-carnitine-rich supplement via stimulation of anti-inflammatory and antioxidant events in testicular torsed rats. Oxid Med Cell Longev 2021; 2021: 5543340.
[33] De Vita S, Chini MG, Bifulco G and Lauro G. Insights into the ligand binding to bromodomain-containing protein 9 (BRD9): a guide to the selection of potential binders by computational methods. Molecules 2021; 26: 7192.
[34] Linani A, Benarous K, Bou-Salah L, Yousfi M and Goumri-Said S. Exploring structural mechanism of COVID-19 treatment with glutathione as a potential peptide inhibitor to the main protease: molecular dynamics simulation and MM/PBSA free energy calculations study. Int J Pept Res Ther 2022; 28: 55.
[35] Mangat HK, Rani M, Pathak RK, Yadav IS, Utreja D, Chhuneja PK and Chhuneja P. Virtual screening, molecular dynamics and binding energy-MM-PBSA studies of natural compounds to identify potential EcR inhibitors

against Bemisia tabaci Gennadius. PLoS One 2022; 17: e0261545.
[36] Gogoi B, Chowdhury P, Goswami N, Gogoi N, Naiya T, Chetia P, Mahanta S, Chetia D, Tanti B, Borah P and Handique PJ. Identification of potential plant-based inhibitor against viral proteases of SARS-CoV-2 through molecular docking, MM-PBSA binding energy calculations and molecular dynamics simulation. Mol Divers 2021; 25: 1963-1977.
[37] Onikanni SA, Lawal B, Fadaka AO, Bakare O, Adewole E, Taher M, Khotib J, Susanti D, Oyinloye BE, Ajiboye BO, Ojo OA and Sibuyi NRS. Computational and preclinical prediction of the antimicrobial properties of an agent isolated from monodora myristica: a novel DNA gyrase inhibitor. Molecules 2023; 28: 1593.
[38] Rivero A, Mora C, Muros M, García J, Herrera H and Navarro-González JF. Pathogenic perspectives for the role of inflammation in diabetic nephropathy. Clin Sci (Lond) 2009; 116: 479-492.
[39] Chen Y, Huang L, Wang S, Li JL, Li M, Wu Y and Liu T. WFDC2 contributes to epithelial-mesenchymal transition (EMT) by activating AKT signaling pathway and regulating MMP-2 expression. Cancer Manag Res 2019; 11: 2415-2424.
[40] James NE, Emerson JB, Borgstadt AD, Beffa L, Oliver MT, Hovanesian V, Urh A, Singh RK, Rowswell-Turner R, DiSilvestro PA, Ou J, Moore RG and Ribeiro JR. The biomarker HE4 (WFDC2) promotes a pro-angiogenic and immunosuppressive tumor microenvironment via regulation of STAT3 target genes. Sci Rep 2020; 10: 8558.
[41] Nakagawa S, Nishihara K, Miyata H, Shinke H, Tomita E, Kajiwara M, Matsubara T, Iehara N, Igarashi Y, Yamada H, Fukatsu A, Yanagita M, Matsubara K and Masuda S. Molecular markers of tubulointerstitial fibrosis and tubular cell damage in patients with chronic kidney disease. PLoS One 2015; 10: e0136994.
[42] Bunnag S, Einecke G, Reeve J, Jhangri GS, Mueller TF, Sis B, Hidalgo LG, Mengel M, Kayser D, Kaplan B and Halloran PF. Molecular correlates of renal function in kidney transplant biopsies. J Am Soc Nephrol 2009; 20: 1149-1160.
[43] LeBleu VS, Teng Y, O'Connell JT, Charytan D, Müller GA, Müller CA, Sugimoto H and Kalluri R. Identification of human epididymis protein-4 as a fibroblast-derived mediator of fibrosis. Nat Med 2013; 19: 227-231.
[44] Seleznik G, Seeger H, Bauer J, Fu K, Czerkowicz J, Papandile A, Poreci U, Rabah D, Ranger A, Cohen CD, Lindenmeyer M, Chen J, Edenhofer I, Anders HJ, Lech M, Wüthrich RP, Ruddle NH, Moeller MJ, Kozakowski N, Regele H, Browning JL, Heikenwalder M and Segerer S. The lymphotoxin β receptor is a potential therapeutic target in renal inflammation. Kidney Int 2016; 89: 113-126.
[45] Meng XY, Zhang HX, Mezei M and Cui M. Molecular docking: a powerful approach for structure-based drug discovery. Curr Comput Aided Drug Des 2011; 7: 146-157.
[46] Kitchen DB, Decornez H, Furr JR and Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov 2004; 3: 935-949.
[47] Chen JH, Wu ATH, Lawal B, Tzeng DTW, Lee JC, Ho CL and Chao TY. Identification of cancer hub gene signatures associated with immunesuppressive tumor microenvironment and ovatodiolide as a potential cancer immunotherapeutic agent. Cancers (Basel) 2021; 13: 3847.
[48] Lawal B, Kuo YC, Tang SL, Liu FC, Wu ATH, Lin HY and Huang HS. Transcriptomic-based identification of the immuno-oncogenic signature of cholangiocarcinoma for HLC-018 multi-target therapy exploration. Cells 2021; 10: 2873.
[49] Lawal B, Kuo YC, Sumitra MR, Wu AT and Huang HS. In vivo pharmacokinetic and anticancer studies of HH-N25, a selective inhibitor of topoisomerase I, and hormonal signaling for treating breast cancer. J Inflamm Res 2021; 14: 4901-4913.