
# IDENTIFICATION OF NATURAL PRODUCTS AS ANTIDIABETIC AGENTS USING COMPUTER-AIDED DRUG DESIGN METHODS

**Laura Guasch Pàmies**

**Dipòsit Legal: T. 609-2013**

Laura Guasch Pàmies

# Identification of Natural Products as Antidiabetic Agents Using Computer-Aided Drug Design Methods

DOCTORAL THESIS

Directed by Dr. Santiago Garcia Vallvé,
Dr. Gerard Pujadas Anguiano and Dr. Miquel Mulero Abellán

Deparment of Biochemistry and Biotechnology

UNIVERSITAT ROVIRA I VIRGILI

Tarragona 2011

UNIVERSITAT ROVIRA I VIRGILI

Departament de Bioquímica i Biotecnologia
C/Marcel·lí Domingo s/n
Campus Sescelades
43007 Tarragona
Tel. : +34 977 559 521
Fax : +34 977 558 232

Dr.Santiago Garcia Vallvé, professor titular del Departament de Bioquímica i Biotecnologia de la Universitat Rovira i Virgili; Dr. Gerard Pujadas Anguiano, professor titular del Departament de Bioquímica i Biotecnologia de la Universitat Rovira i Virgili; i Dr. Miquel Mulero Abellán, professor lector del Departament de Bioquímica i Biotecnologia de la Universitat Rovira i Virgili.

CERTIFIQUEM

Que aquest treball, titulat "**Identification of natural products as antidiabetic agents using computer-aided drug design methods**", que presenta Laura Guasch Pàmies per a l'obtenció del títol de Doctora, ha estat realitzat sota la nostra direcció al Departament de Bioquímica i Biotecnologia d'aquesta universitat i que acompleix els requeriments per poder optar a Menció Europea.

Tarragona, 4 de novembre del 2011

| | | |
|---|---|---|
| Dr. Santiago Garcia Vallvé | Dr. Gerard Pujadas Anguiano | Dr. Miquel Mulero Abellán |

# AGRAÏMENTS

Aquests quatre anys han estat per mi com el camí cap a Ítaca, ric en experiències i en coneixement, i sobretot amb molts bons records. Vull agrair a tota la gent que ha compartit amb mi cadascun d'aquests moments, a tots i totes moltes gràcies.

La realització d'aquesta tesi no hauria estat possible sense els consells, el coneixement i la motivació per part del Dr. Santi Garcia Vallvé i del Dr. Gerard Pujadas Anguiano. Santi i Gerard, gràcies per la confiança i la llibertat que m'heu donat durant aquests anys.

També vull agrair als membres del Grup de Recerca de Nutrigenòmica que d'alguna manera o altra han estat partíceps en l'elaboració d'aquesta tesi: Prof. Lluís Arola, Prof. Cinta Bladé, Dra. Josepa Salvadó, Dr. Antón Romeu, Dra. Anna Ardèvol, Dra. Mayte Blay, Dr. Juan Fernández, Dr. Miquel Mulero, Dra. Begoña Muguerza, i especialment a la Dra. Montse Pinent. També agrair als tècnics i personal de secretaria del departament per la seva ajuda.

I would also like to thank Prof. Klaus Liedl for giving me the opportunity to stay in his group and all the members of the group for the warm welcome I received in Innsbruck. Many thanks to Dr. Gerhard Wolber for your implication in the project. I would like to specially thank Simo, Hannes and Patrick for funny moments. Laura y Leire, tampoco me puedo olvidar de vosotras, que bien nos lo pasamos en Innsbruck, cuántos recuerdos tengo, gracias por vuestra compañía y amistad.

Aquesta tesi no hagués estat la mateixa sense els cursos, seminaris, congressos en què he participat durant la meva trajectòria del doctorat. Ells m'han fet créixer, m'han donat l'oportunitat d'aprendre, de conèixer nous grups i per descomptat de descobrir nous indrets. Tinc un bon record de les persones que he anat coneixent i que plegades hem disfrutat d'aquestes reunions no només científiques.

Els companys de laboratori són un pilar fonamental per un becari pre-doctoral ja que amb ells comparteixes moltes hores tant amb maldecaps com amb alegries i són els que més et comprenen, a tots ells moltes gràcies. Pere, amb tu vaig començar amb arbres filogenètics i sempre has estat un exemple a seguir. Montserrat, gràcies per la teva ajuda en els meus primers farmacòfors i els ànims en tot moment. Albert, com enyoro els jocs de pilota al despatx i els entrenaments del *Recre*. Lídia, encara que no siguis de l'àmbit computacional també hi estàs benvinguda, m'han dit que fas la competència a la *Sirenita*. Marina, sembla ser que després de donar tants tombs has trobat el teu lloc a Barcelona, molta sort. Esther, sols dir-te que estic molt contenta d'haver-te conegut i d'haver viscut tants i tants moments plegades. Gràcies per la teva amistat, alegria i complicitat, i sobretot per donam suport en tot moment.

El temps passa, els companys es doctoren, i el despatx es va renovant amb aires de CTNS. Al Toni i al Josep que feu una parella científica fantàstica. A les Annes, Isa, Manuel i Mar, tots vosaltres m'heu acompanyat a la recta final de la tesi, us desitjo molta sort amb els vostres experiments.

Tampoc em puc oblidar dels doctorants que treballeu amb cèl·lules i rates. Anabelooo, gràcies per les teves visites al despatx i les nostres conyes, m'ho he passat genial amb tu. A la Cris, Ligia i Aleix, gràcies per haver compartit tantes converses al *Soteras* i haver fet els dinars més amens. A la Lídia i al Víctor, ànims amb les vostres tesis, ja us queda menys. I també a la Neus, Noe, Laura, Ester, Raquel, Husam i Mario. També tinc agraïments pels que ara ja són doctors i ja fan de les seves; a la Gemma, Ximena, Isa, Sabina i David.

Als amics de la carrera, que encara que passi el temps sempre trobem algun forat per reunir-nos i posar-nos al dia. Vane, moltes gràcies per ser una gran amiga, i per tots els moments que hem disfrutat. Tessa, gràcies per compartir amb mi descobertes musicals i viatges. Anna, gràcies per haver confiat en mi, i molt ànims que això ja ho tens. Núria, et desitjo molta sort en la teva nova etapa com a professora. Antonio, Xavi i Mike, encara recorodo l'habitació 101 i lo bé que ens ho vam passar tots plegats.

També vull agrair aquesta tesi a Tarragona i a tota la gent que hi he conegut, han estat uns anys genials que mai oblidaré. Idoia i Gis, segur que ens esperen molts festivals i concerts amb o sense ukelele. Irene, et dóno ànims per la teva etapa final. Helena, gràcies per la teva simpatia que espero poder continuar gaudint. A la Cris, Gerard, Laura, Joanet, Berta, Xavi, Joan Oriol, Noé, Chus, m'ha encantat compartir tants i tants sopars amb vosaltres.

Als meus consellers i conselleres de Montblanc i rodalies, jejeje, gràcies a vosaltres he pogut gaudir de tants moments memorables: carnavals, castanyades, calçotades, festes majors, viatges, sou únics!!! Escolà que des de la guardaria ens coneixem. Maria, per les rutes que encara ens queden per fer. Anna, per les cançons que versionem. Janet, pel teu estil i les *coreos* que ens fas. Joan, patentarem el *GuaschApp* aviat. Albert, crec que necessitaré un massatge després de tot plegat. Tampoc em puc deixar a la Lídia i al Raül, i a la Cristina i al Jordi, sou unes parelles fantàstiques.

A la família, que en tants dinars familiars m'heu preguntat que estava fent, gràcies pel vostre interès. A les padrines, Lola i Rosita, als tiets i tietes, als cosins i cosines i als més petits. *Kusi*, una menció especial per tu, gràcies pel teu caràcter i per tirar sempre endavant.

Gemma, encara que hem seguit camins molt diferents, sempre m'has ajudat i m'has comprès, i també has tingut molta paciència, ets la millor *teacher*!!! I amb el Ramon, us desitjo molta sort als dos.

Als meus pares, el més sincer agraïment, i a ells els hi dedico aquest llibre. Vosaltres sempre heu estat al meu costat i m'heu donat suport en totes les meves decisions. Gràcies per tots els valors que m'heu transmés; i encara que no us ho digui gaire sovint, us admiro i us estimo.

*Val més una gota de saber que un mar de fortuna.*

(Font del Grèvol - Montblanc)

***Als meus pares***

## CONTENTS

## 1. Introduction: diabetes – an emerging epidemic of the 21st century

Diabetes mellitus (DM) is a metabolic syndrome that constitutes a major health problem [1,2]. It is estimated that 246 million people worldwide have diabetes and that 380 million people will be afflicted with diabetes by 2025. In addition, 3.8 million people die each year from diabetes [3]. DM is characterized by abnormally high levels of plasma glucose, known as hyperglycemia, in the fasting state or after the administration of glucose during an oral glucose tolerance test. DM is caused by a relative or absolute deficiency in insulin secretion, a resistance to insulin secretion or both [4-6]. The World Health Organization recognizes two distinct clinical forms of diabetes (Figure 1), type 1 diabetes (T1DM) and type 2 diabetes (T2DM). T1DM, also referred to as the juvenile variety of DM, results from an absolute deficiency of insulin due to the destruction of insulin-producing pancreatic β-cells. T2DM is a multifactorial disease that is characterized by insulin resistance associated with not only hyperinsulinaemia and hyperglycemia but also atherosclerosis, hypertension and an abnormal lipid profile [7]. T2DM accounts for 90-95% of the diagnosed cases of DM [8]. Genetic and environmental factors, increased height and weight development, increased maternal age at delivery, and exposure to some viral infections have also been linked to the risk of developing T1DM. Several risk factors have been associated with T2DM, including obesity, changes in diet and physical activity, age, insulin resistance, a family history of diabetes and ethnicity [9,10]. Changes in diet and physical activity related to rapid development and urbanization have led to a sharp increase in the number of people developing diabetes.

T1DM and T2DM require careful monitoring and control. Without proper management, they can lead to very high blood sugar levels, which can result in long-term damage to various organs and tissues. The major chronic complications of diabetes are cardiovascular disease, which is the primary cause of death in people with diabetes [11,12]; nephropathy, which can result in total kidney failure and the need for dialysis or kidney transplant [13]; neuropathy, which can ultimately lead to ulceration and amputation of the toes, feet and lower limbs; and retinopathy, which is characterized by damage to the retina of the eye and can lead to a loss of vision.

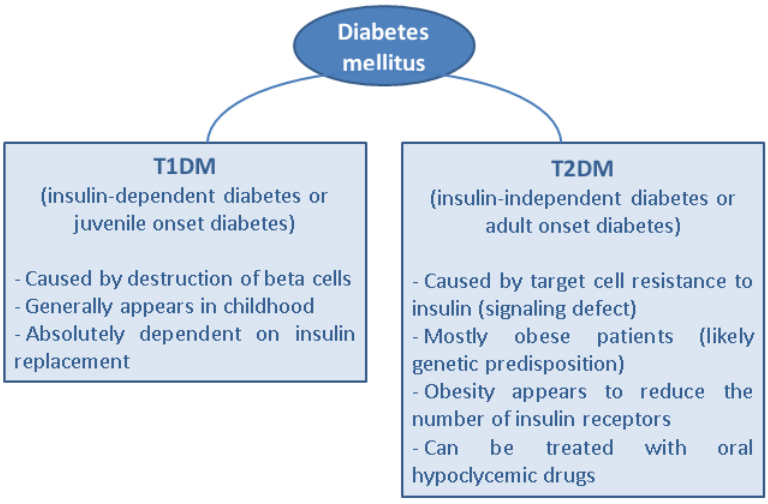



**Figure 1.** Classification of diabetes mellitus.

## 2. Targeting Type 2 Diabetes

Both T1DM and T2DM are chronic conditions that typically cannot be cured. However, all forms of diabetes have been treatable since the development of readily available insulin in 1921. The enhancement of insulin secretion by pancreatic islet β-cells is a major goal for the treatment of T2DM. Antidiabetic drugs or hypoglycemic agents are medications that work to lower blood glucose concentrations (i.e., the amount of sugar in the blood). There are different classes of antidiabetic drugs, and their selection depends on the nature of the diabetes and the age and situation of the person, as well as other factors. Antidiabetic drugs exert their useful effects through (1) increasing insulin levels in the body, (2) increasing the body's sensitivity (or decreasing its resistance) to insulin, or (3) decreasing glucose absorption in the intestines [14].

A list of these agents along with their molecular targets, mechanisms of action and side effects related to their use are summarized in Table 1 and are visualized in Figure 2. Because of their adverse side effects, most of these treatments are considered to be unsatisfactory in terms of the prevention of complications and preservation of quality of life. α-glucosidase inhibitors, such as acarbose and miglitol, while effective at decreasing the absorption of glucose by interfering with the action of α-glucosidases present in the small intestinal brush border, are often associated with abdominal bloating, diarrhea and flatulence. Conventional insulin secretagogues, such as sulfonylureas and the class of meglitinides, both result in the

induction of hypoglycemia. While metformin is the only therapeutic agent that has been demonstrated to reduce macrovascular events in T2DM, its use is not recommended in conditions in which a patient has decreased renal or hepatic function. Metformin is the first-line drug of choice for the treatment of T2DM, particularly in overweight and obese patients and those with normal kidney function [15]. Agonists of the peroxisome proliferator-activated nuclear receptor (PPAR), thiazolidinediones, are able to reduce insulin resistance but are under intense scrutiny because of concerns with their safety. In fact, the use of rosiglitazone has now been severely restricted in the US and has been completely suspended in Europe as a result of concerns regarding its cardiovascular safety [16,17]. Notably, insulin, which is used to treat T1DM patients (for whom the hormone is no longer produced internally), is also occasionally used for patients with T2DM when other medications fail to adequately control blood glucose levels. However, hypoglycemia and weight gain are common side effects. Thus, new approaches are needed to treat T2DM. One of the desirable approaches to achieve this goal would be to identify agents that promote/enhance glucose (nutrient)-dependent insulin secretion [18].

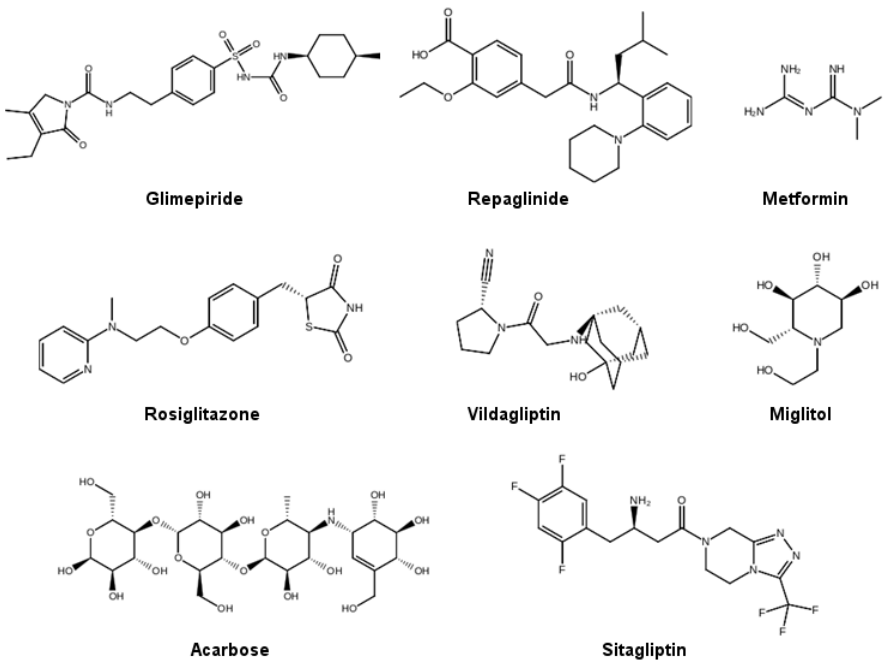



**Figure 2.** Some of the important marketed antidiabetic drugs.

**Table 1.** Drugs and their targets available nowadays for the treatment of diabetes.

| Drug class | Molecular Target | Mechanism/actions | Adverse events | Generic Name | Brand Name |
|---|---|---|---|---|---|
| Insulin | Insulin receptor | Correct insulin deficiency | Hypoglycemia, weight gain | Insulin glargine | Lantus® |
| | | | | Insulin lispro | Humalog® |
| Sulfonylureas | ATP-potassium channel | Stimulate insulin secretion | Hypoglycemia, weight gain | Glimepiride | Armaryl® |
| | | | | Glipizide | Glucotrol® |
| | | | | Glyburide | Diabeta® |
| Meglitinides | ATP-potassium channel | Stimulate insulin secretion | Hypoglycemia, weight gain | Repaglinide | Prandin® |
| | | | | Nateglinide | Starlix® |
| Biguanides | Unknown | Inhibition of hepatic glucose output | Gastrointestinal disturbances, lactic acidosis | Metformin | Glucophage® |
| Thiazolidinediones | PPARγ | Increase insulin sensitivity | Weight gain, edema, anemia | Pioglitazone | Actos® |
| | | | | Rosiglitazone | Avandia® |
| α-glucosidase inhibitors | alpha-glucosidase | Retard carbohydrate absorption | Gastrointestinal disturbances | Acarbose | Precose® |
| | | | | Miglitol | Glyset® |
| Glucagon-like peptide -1 analogues | GLP-1 receptor | Stimulate insulin secretion | Gastrointestinal disturbances, nausea, abdominal pain, weight loss | Exenatide | Byetta® |
| | | | | Liraglutide | Victoza® |
| Dipeptidyl peptidase-IV inhibitors | Dipeptidyl peptidase-IV | Increase blood concentration of the incretin GLP-1 | Increased risk for infection and headache | Vildagliptin | Galvus® |
| | | | | Sitagliptin | Januvia® |
| | | | | Saxagliptin | Onglyza® |
| Amylin analogues | Calcitonin receptor and RAMP1, RAMP2 or RAMP3 | Slow gastric emptying and supress glucagon | Nausea | Pramlintide | Symlin® |

The 2010 American Diabetes Association Standards of Medical Care in Diabetes added the criteria of glycated hemoglobin (HbA1c) levels ≥ 48 mmol/mol (≥6.5%) for the diagnosis of diabetes [19]. HbA1c is a form of hemoglobin that is measured primarily to identify the average plasma glucose concentration over prolonged periods of time. It is essential to monitor therapy with HbA1c and the levels of blood glucose and to adjust or advance therapy frequently (every 2 to 3 months) if the appropriate goal for each patient has not been achieved. The American Association of Clinical Endocrinologists/American College of Endocrinology (AACE/ACE) provides therapeutic pathways based on the current levels of HbA1c [20,21], which differ from the corresponding recommendation of the American Diabetes Association and European Association for the Study of Diabetes (ADA/EASD). In case of an initial HbA1c < 7.5%, lifestyle modification alone might be sufficient to achieve the goal of HbA1c levels below 6.5%. If this fails, then monotherapy is recommended with metformin as the preferred agent. In case of an initial HbA1c between 7.7 and 9.0%, pharmacotherapy should be started with a dual approach, because monotherapy may be insufficient to attain the 6.5% goal and thus inadequate to address the underlying pathophysiology (i.e., insulin resistance with advanced β-cell failure, inflammation and lipotoxicity). In addition to metformin, GLP-1 agonists/dipeptidyl peptidase-IV (DPP-IV) inhibitors are recommended as the first choice with an optional substitution of thiazolidinediones (TZDs) in the case of metabolic syndrome. If the initial HbA1c is above 9.0%, therapy should start with either a dual or triple approach. Triple therapy should include TZDs in addition to metformin plus GLP-1 agonists/DPP-IV inhibitors. In the event that the target of 6.5% is not reached on a previous regime or if symptomatic hyperglycemia develops, the algorithm recommends moving directly to insulin therapy [7,22].

Extensive research has been conducted on the molecular targets for T2DM, including PPARγ, protein tyrosine phosphatase-1B (PTP1B), DPP-IV, glycogen synthase kinase-3 (GSK-3), pyruvate dehydrogenase kinase (PDHK), cannabinoid receptors, fructose-bisphosphatases, and β3-adrenoceptor (β3-AR), in an attempt to develop newer antidiabetic agents [23,24]. These therapeutic targets are important, and most of them are suitable for an in silico analysis [8].

PTP1B is emerging as a strong target for the treatment of T2DM and obesity [25]. Genetic data and knock-out mouse model studies indicate a significant role of PTP1B in insulin signaling [26,27]. PTP1B knock-out mice have been shown to exhibit enhanced insulin sensitivity, as measured by improved glucose clearance [26], and are resistant to diet-induced obesity [27]. Many selective and potent inhibitors of PTP1B have been discovered. The various classes and current status of these molecules have been reviewed extensively elsewhere [28]. The β-ARs belong to the superfamily of G protein-coupled receptors (GPCR). Given that selective agonists of β3-AR are shown to have thermogenic and hypoglycemic effects in

mouse models, β3-AR is currently thought to be an important target for the treatment of obesity and T2DM [29,30]. Most of the reported β3-AR agonists possess either an arylethanolamine or aryloxypropanolamine substructure. PDHK is a group of highly specific enzymes that deactivate the pyruvate dehydrogenase complex (PDC), thus impairing carbohydrate metabolism by reducing the oxidation of pyruvate [31]. In diabetes, PDHK is activated and leads to the inactivation of PDC by ATP-dependent phosphorylation. A number of PDHK inhibitors are now available to enable this mechanism to be evaluated as a therapy for diabetes. Aicher et al. have reported a series of tetracyclic terpenes with an oxime functional group. They found that the oxime group forms hydrogen-bonding interactions with the substrate binding sites of PDHK and that these compounds have high potency at this target [32]. The endogenous cannabinoid system has been reported to play an important role in the regulation of food intake and lipid metabolism [33]. Recent reports suggest a role of the CB1 receptor in obesity, insulin resistance and related disorders [34]. Thus, there is hope for using CB1 antagonists as a new strategy for treating diabetes. In addition, CB1 antagonists are already being studied in relation to many other therapeutic areas; and, thus various medicinal chemistry strategies are being employed to discover new antagonists for this receptor [35]. PPARγ and DPP-IV are the therapeutic targets studied in this thesis, and they are described in more detail below.

## 2.1. Peroxisome proliferator-activated receptor gamma (PPARγ)

Peroxisome proliferator-activated receptors (PPARs) are members of the nuclear receptor superfamily that regulate the gene expression of proteins involved in energy, glucose and lipid metabolism, the proliferation and differentiation of adipocytes and the sensitivity of insulin [36]. They function as cellular sensors that activate transcription in response to the binding of natural or synthetic ligands. Three receptor subtypes, PPARα, PPARβ/δ and PPARγ, have been identified. Although the three subtypes share a high level of sequence and structural homology, they exhibit differences in tissue expression and physiological function [37]. PPARα is found in the liver, kidney, heart, and muscle. It is important for the uptake and oxidation of fatty acids and lipoprotein metabolism. PPARα is the target of lipid lowering fibrates. PPARγ is localized in fat, large intestine, and macrophages. It plays an important role in adipocyte differentiation. PPAR β/δ is expressed in most cell types. Agonists of PPARα and PPARγ are currently approved for use in treating dyslipidemia and T2DM, respectively [38]. PPARβ/δ agonists play important roles in dyslipidemia, cancer treatment, and cell differentiation within the central nervous system.

- **PPARγ agonists**

TZDs are an important class of synthetic PPARγ agonists. TZDs are antidiabetic agents that target adipose tissue and that improve insulin sensitivity. They are currently used in the treatment of T2DM. Despite the clinical benefit of these drugs, the use of TZDs has been associated with adverse effects, including weight gain, increased adipogenesis, renal fluid retention, and possible increased incidence of cardiovascular events [39,40]. Therefore, new PPARγ ligands with enhanced therapeutic efficacy and reduced adverse effects are needed. A promising new group of such ligands are selective PPARγ modulators (SPPARγMs) [39,40]. These compounds act as partial agonists of PPARγ and display different binding properties when compared with full agonists.

There is another type of synthetic PPAR agonists called dual PPARα/γ and pan PPARα/γ/β/δ ligands. They were developed in an attempt to achieve multiple therapeutic benefits; however, these compounds have encountered multiple safety issues that have thus far not been resolved [41].

- **PPARγ mechanism**

PPARs function through the formation of heterodimers with the retinoid X receptor (RXR) and dock to the promoter regions of genes, which regulates transcription in a ligand-dependent manner through the differential recruitment of co-activators and co-repressors [42]. PPARγ can considered a rheostat for insulin sensitivity that responds to an integrated nutritional status conveyed through multiple signals sensitive to the dietary and endocrine status [43].

Like other nuclear receptors, PPARs are modular in structure and contain the following functional domains: a N-terminal region, a DNA-binding domain (DBD), a flexible hinge region, a ligand binding domain (LBD) and a C-terminal region. The DBD contains two zinc finger motifs, which bind to specific sequences of DNA, known as hormone response elements, when the receptor is activated. The LBD has an extensive secondary structure that consists of 13 α-helices and a β-sheet (see Figure 3A) [44]. Natural and synthetic ligands bind to the LBD and either activate or repress the trans-activation activity of the receptor.

Because of their importance as pharmaceutical targets for regulating the fatty acid metabolism and antidiabetic drugs and because they provide an interesting example of receptors interacting with other molecular partners in a ligand-dependent manner, the structure of the PPAR LBD has been intensively studied at the atomic level. Since the first experimental X-ray structures of PPARγ were obtained in 1998 [42,45], numerous structures have been determined for PPARα, PPARγ and PPARδ

in both the liganded and apo forms, with or without a co-activator or a co-repressor, and in the presence or absence of RXR.

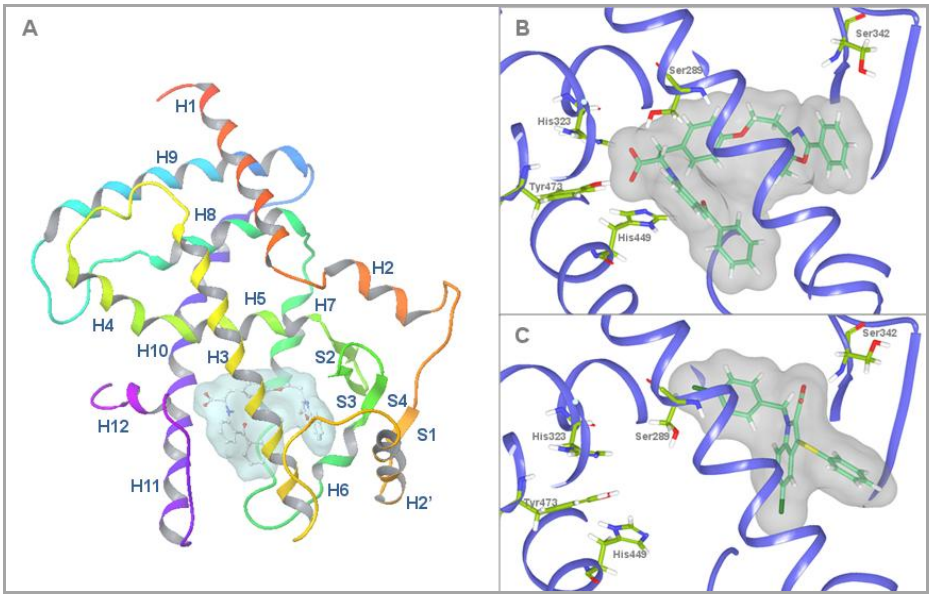



**Figure 3.** 3D structure of PPARγ along with the secondary structure elements (A). Binding models of (B) the PPARγ full agonist Farglitazar (crystal structure 1FM9) and (C) the PPARγ partial agonist nTZDpa (crystal structure 2Q5S). Important binding residues are depicted as wireframes. Oxygen, nitrogen, and hydrogen atoms are coloured red, blue and white, respectively.

PPARγ is thought to be activated by full agonists via a molecular switch in the most carboxy terminal helix, H12, of the LBD [44]. H12 forms part of the ligand-dependent activation domain AF-2 that closes on the ligand-binding site in response to ligand binding. The resulting active form can bind to several co-activator proteins that activate the cellular transcriptional machinery [44]. Full agonists occupy the large binding site of PPARγ in a U conformation and are generally formed by a polar head and a hydrophobic tail [46]. The polar head forms a net of hydrogen bonds with the Ser 289, His 323, His 449 and Tyr 473 PPARγ side chains (Figure 3B). This net of hydrogen bonds is responsible for the conformational change of H12 and the activation of PPARγ [46]. Partial agonists, however, activate PPARγ using a H12-independent mechanism [47,48]. The key interactions between partial agonists and the (LBD) of PPARγ are different, since partial agonists do not use the net of hydrogen bonds used by full agonists to bind to PPARγ. This causes a reduction in the degree of H12 stabilization that affects the recruitment of co-activators and that decreases the transcriptional activity of PPARγ [49,50]. With only minor differences, most of the currently described partial agonists interact with

the LBD of PPARγ through a hydrogen bond with Ser342 [46] and several hydrophobic interactions (Figure 3C). These hydrophobic interactions are similar to those used by full agonists. A new mechanism has been recently suggested by which partial and full PPARγ agonists may improve insulin sensitivity independent of receptor agonism. This mechanism consists in blocking the phosphorylation of PPARγ [51] and may explain how partial agonists can exhibit similar or higher antidiabetic effects than full agonists and the differing side-effect profiles of both types of agonists [52]. These partial agonists may then achieve comparable efficacy in insulin sensitization through a similar inhibitory effect on PPARγ phosphorylation whereas the differences in their agonist potency could be linked to differences in side effects [52].

## 2.2. Dipeptidyl peptidase-IV

The dipeptidyl peptidases (DPPs) are a subclass of the serine protease family. Members of this family include DPP I–IV, fibroblast activation protein-α (FAP), DPP-8 and DPP-9 [18]. Except DPP-IV all these enzymes remain poorly characterized and their natural substrates have not yet been identified [18]. DPP-IV is constitutively expressed on epithelial and endothelial cells of a variety of different tissues, for example, intestine, liver, lung, kidney and placenta. Recently, DPP-IV has emerged as a new treatment option of T2DM [53].

### ■ DPP-IV mechanisms

DPP-IV specifically removes N-terminal dipeptides from substrate containing proline or alanine at the penultimate position, transforming them into inactive or even antagonistic species. Researchers have found that the activity of two potent stimulators of insulin secretion, glucagon-like peptide-1 (GLP-1) and glucose-dependent insulinotropic polypeptide (gastric inhibitory polypeptide or GIP), is rapidly cleaved by DPP-IV. [54]. The structure of GIP, GLP-1 and GLP-2 reveals a highly conserved alanine at position 2, rendering these peptides ideal substrates for the DPP-IV.

Incretin hormones are defined as intestinal hormones released in response to nutrient ingestion, which potentiate the glucose-induced insulin response (the incretin effect). GLP-1 is an incretin hormone secreted by intestinal L-cells in response to meals. It stimulates insulin biosynthesis and secretion, reduces glucagon release, slows gastric emptying, reduces appetite, and stimulates regeneration and differentiation of islet β-cells [55]. On the other hand, the other most important incretin hormone GIP is produced by the duodenal K-cells and is extensively involved in glucose metabolism by enhancing insulin secretion [56]. Both peptides have very short half-lives because of their rapid degradation by DPP-IV. Therefore,

inhibiting DPP-IV prolongs the action of GLP-1 and GIP, which in turn improves glucose homeostasis with a low risk of hypoglycemia and potencial for disease modification (Figure 4).

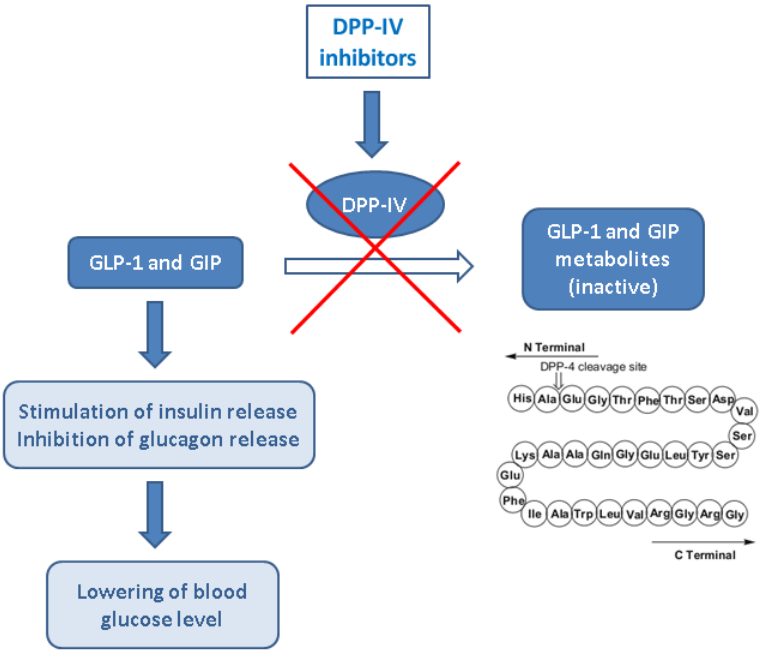



**Figure 4.** Diagram illustrating DPP-IV inhibition for controlling glucose levels.

## ■ DPP-IV characteristics: structure and binding site

DPP-IV is an acid transmembrane glycoprotein that consists of a cytoplasmic tail (residues 1-6), a transmembrane region (residues 7-28), and an extracellular region (29-766) (Figure 5A). The extracellular region can be further subdivided into two domains: a) the catalytic domain (residues 508-766), which shows an α/β hydrolase fold and contains the catalytic triad Ser630 – Asp708 – His740, and b) an eight-bladed β propeller domain (residues 56-497), which also contributes to the inhibitor binding site [57]. DPP-IV is enzymatically active as a homodimer.

The DPP-IV binding site is highly druggable in the sense that tight, specific binding to the enzyme can be achieved using small molecules with drug-like physicochemical properties [57,58]. The different interaction motifs used by DPP-IV ligands include the catalytic Ser630, the oxyanion hole Tyr631-Tyr547, the hydrophobic S1 pocket created by Tyr 631-Val 656-Trp 659-Tyr 662-Tyr 666-Val 711, the P2 region Arg 125-Asn 710, and the N-terminal recognition region Glu 205-Glu 206-Tyr 662 (Figure 5B).

The considerable number of DPP-IV crystal structures that have been published since 2003 provide a detailed picture of the structural characteristics of the binding site and the molecular recognition of small molecules. It is not surprising that a large number of diverse DPP-IV inhibitors have been discovered because the binding site offers a) a deep lipophilic pocket combined with several exposed aromatic side chains to achieve high affinity small molecule binding and b) significant solvent access, which allows for the tuning of the physicochemical properties of the inhibitors for improved pharmacokinetic behavior [57].

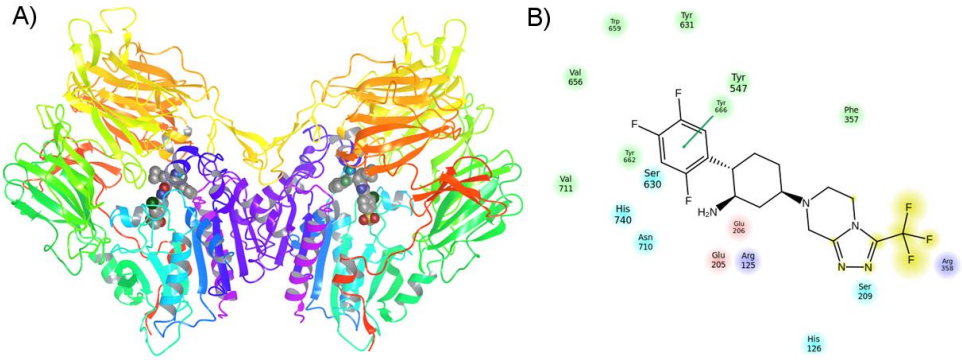



**Figure 5.** A) Human DPP-IV in complex with a fluoroolefin inhibitor (crystal structure 3C45 with two chains). B) Key interactions between DPP-IV and sitagliptin in two-dimensional representation. Residues colored in green are hydrophobic, residues colored in cyan are polar. Red residues are negative charged and could act as acceptors, whereas purple residues are positive charged and could act as donors. Ligand exposure to the solvent is colored in yellow

## 3. Natural Products - Alternative medicine

For thousands of years, medicine and natural products have been closely linked through the use of traditional medicines and natural poisons [59,60]. In the last 200 years, scientific developments have allowed progress in drug research. The first systematically studied drugs were plant constituents that are still used today, such as salicylic acid, digitoxin, morphine, quinine, and pilocarpine. The discovery, isolation, and biological studies of antibiotic compounds from microorganism cultures have revolutionized healthcare. Prominent examples of nature-derived antibiotics include streptomycin, chloramphenicol, chlortetracycline, cephalosporin C, erythromycin and vancomycin [61].

Natural products have gone through a long selection process to develop interactions with biological targets and are therefore a valuable source of ideas for novel chemical entities in drug development [61]. Owing to their diversity, target

affinity, and specificity, natural products have demonstrated enormous potential as modulators of biomolecular function, have been an essential source for drug discovery, and have provided design principles for combinatorial library development [62,63]. Natural products have proven to be the richest source of novel compound classes for biological studies and an essential source for the discovery of new drugs. In addition, natural products offer an advanced starting point in the search for highly specific and potent modulators of biomolecular function as well as novel drugs [63-65].

Natural products have played an important role in the traditional treatment of T2DM since ancient times [66,67]. The first recorded description of diabetes mellitus dates back to the Ebers papyrus in Egypt around 1500 B.C. [68]. Later, in India, the early Ayurvedic texts, such as the Sushruta Samhita and the Charaka Samhita, which were written in the 4th to 5th century B.C., described the use of approximately 760 and 500 species of medicinal plants, respectively. In China, the Ben Jing, which was written in about 104 B.C., provided detailed descriptions of 252 species with reference to those used to treat diabetes [69].

Plants are one of the most important sources of antidiabetic compounds. In many regions of the world, herbal remedies continue to be more accessible and affordable than conventional antidiabetic drugs. Additionally, in societies with well-developed, modern health care systems, the demand for herbal remedies to complement prescribed, modern therapies is growing for many diseases, including diabetes [67]. Each region of the world has developed a materia medica of antidiabetic remedies based on the local flora [70-72]. Climatic factors and cross-cultural communication also play a role. Generally, the use of a particular plant in a number of regions is strong evidence for its effectiveness. The families of plants with the most potent hypoglycemic effects include *Leguminosae* (11 species), *Lamiaceae* (8 species), *Liliaceae* (8 species), *Cucurbitaceae* (7 species), *Asteraceae* (6 species), *Moraceae* (6 species), *Rosaceae* (6 species), *Euphorbiaceae* (5 species) and *Araliaceae* (5 species). The most commonly studied species are *Opuntia streptacantha*, *Trigonella foenum-graecum*, *Momordica charantia*, *Ficus bengalensis*, *Polygala senega* and *Gymnema sylvestre* (see Table 2 for more detailed information) [73,74]. Moreover, a wide range of plant families and types of phytochemicals are associated with antidiabetic activity. At the same time, certain groups, such as alkaloids, saponins, xanthones and flavonoids, and nonstarch polysaccharides, appear to have effects with particular significance to diabetes treatment [69,73].

**Table 2.** Species of plants reported to be used traditionally to treat diabetes.

| Species | Family | Active Principle for antidiabetic activity | Part Used | Area Traditionally Used |
|---|---|---|---|---|
| *Opuntia streptacantha* | Cactaceae | Isoquinoline alkaloid, cyanogenetic alkaloids | Aerial parts | C. America |
| *Trigonella foenum-graecum* | Fabaceae | Saponins, 4-hydroxyisoleucine | Seed, leaf | Africa, India, Middle East |
| *Momordica charantia* | Cucurbitaceae | Charantin, polypeptide (p-insulin), sterols | Fruit, leaf | Africa, India, C. America, Australia, Middle East |
| *Ficus bengalensis* | Moraceae | Bengalinoside, Phytosterolin, flavonoid, glycoside, glycosidal fraction | Bark | S.E. Asia |
| *Polygala senega* | Polygalaceae | Triterpenoid glycosides, senegins II and III | Root | Asia |
| *Gymnema sylvestre* | Asclepiadaceae | Gymnemic acids III, IV, V, VII, and gymnemoside B | All | Tropics |

Numerous mechanisms of actions have been proposed for these plant extracts. Some hypotheses relate to their effects on the activity of pancreatic β-cells (synthesis, release, cell regeneration/revitalization), an increase in the protective/inhibitory effect against insulinase, an increase in insulin sensitivity or the insulin-like activity of the plant extracts. Other mechanisms may involve improved glucose homeostasis, such as an increase in the peripheral utilization of glucose, increased synthesis of hepatic glycogen and/or a decrease in the glycogenolysis acting on enzymes, the inhibition of intestinal glucose absorption, a reduction in the glycaemic index of carbohydrates and a reduction in the effect of glutathione. All of these actions may be responsible for the reduction and or abolition of diabetic complications [75]. Plants hold definite promise in the management of DM. The isolation and identification of the active constituents of these plants and the preparation of standardized doses and dosing regimens may be important for improving the hypoglycemic action of these plant products.

## 4. Functional Food in Diabetes

Foods can be considered functional if they are proven to beneficially affect one or more of the target functions in the body, beyond adequate nutritional effects, in a way that is relevant to an improved state of health and well-being, a reduction in the risk of diseases, or both [76]. Nutraceutical substances with commercial value can be obtained from functional foods that have demonstrated a physiological benefit or that are capable of providing some sort of protection against a chronic or infectious disease [77]. Several natural compounds are described as nutraceutical. The initial step in the research and development of a functional food is the identification of a specific interaction between one or a few components of this food and a function in the organism that is potentially beneficial for health [76]. Often, this interaction results from natural products that act as a functional food component.

Functional foods might have a particularly high impact on the prevention or treatment of excessive weight gain and diabetes, for which the link between nutrition, biological responses and diseases is clearly established [78]. Many functional foods and supplements are promoted as being beneficial for the management of diabetes or for reducing the risk of developing diabetes and its complications [79]. However, the available evidence on functional foods identified in this field is incomplete, primarily because of the lack of diet-based intervention trials that are of sufficient duration to be relevant to the natural history of diseases such as obesity and diabetes [78].

Weight control is an effective technique for the management of diabetes, and functional foods promoting weight loss could be developed for those with T2DM [80]. Studies have shown that a modest weight loss of 5–10% of body weight is associated with improvements in cholesterol, blood pressure and insulin sensitivity, which are known risk factors for CVD and T2DM [15,81,82]. However, it may also be possible to incorporate functional foods that affect insulin action independently of weight loss into the diet. To lower the glycemic index, nuts and peanuts can be potentially included in a healthy diet. However, more long-term studies are needed to demonstrate the effects of nuts and peanuts on glycemia [80,83]. Given the potential benefits of omega-3 FAs on CVD risk, the regular consumption of fish is recommended [80,84]. Finally, studies of patients with T2DM also show that cinnamon may have the potential to lower glucose levels [85,86]. However, more research on the proposed health benefits of cinnamon supplementation is necessary before unambiguous recommendations can be made. In conclusion, a growing number of individuals are adding functional foods and natural health products to their diet to help control blood glucose; however, a large amount of research is still needed before the benefits of these supplements can be confirmed and before these foods can be recommended routinely for glycemic control [79,80]. A new era of

nutrition science is just beginning, and there is the potential for exciting developments regarding the role of food in achieving optimal health and in the prevention and management of diseases [78].

## 5. Computer-aided drug design methods in the discovery of antidiabetic drugs

Computer-aided drug design (CADD) methodologies have made great advances and have contributed significantly to the discovery and/or optimization of many clinically used drugs in recent years [87]. Drug discovery and development is a time-consuming and expensive process. On average, it takes 10–15 y and $500–800 million to introduce a drug into the market [88,89]. Accordingly, CADD approaches have been widely used in the pharmaceutical industry to accelerate the process [90,91]. CADD helps scientists focus on the most promising compounds so that they can minimize the synthetic and biological testing efforts. In practice, the choice of employing CADD approaches is usually determined by the availability of experimentally determined 3D structures of the target proteins. Thus, there are two major types of drug design: ligand-based drug design and structure-based drug design. If protein structures are unknown, various methods of ligand-based drug design can be employed, such as quantitative structure activity relationship (QSAR) and pharmacophore analysis. If the target structures are known, structure-based approaches can be used, such as molecular docking, which employs the 3D structures of the targets to design novel active compounds with improved potency. As more structures are becoming available, the prediction accuracy will likely improve [90].

### 5.1. ADMET properties

To exert a pharmacological effect in tissues, a compound has to penetrate various physiological barriers, such as the gastrointestinal barrier, the blood-brain barrier and the microcirculatory barrier, to reach the blood circulation. Once in circulation, the compound is subsequently transported to its effector site for distribution into tissues and organs, degraded by specialized enzymes, and finally removed from the body via excretion. Accordingly, the absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties of a compound directly affect its usefulness and safety [87]. Thus, a reliable filter for the selection of good candidate drugs for development would greatly reduce the time and cost of R&D. Therefore, pharmaceutical companies are trying to move ADMET evaluations into the early stages of drug discovery [92]. The huge libraries of compounds are typically subjected to pre-filtering with the ADMET properties.

Lipinski [93] studied the physicochemical properties of 2245 drugs from the World Drug Index (WDI) and found that poor absorption and permeation are more likely to occur when the molecular weight < 500 g/mol, Clog P < 5, hydrogen bond donors < 5 and hydrogen bond acceptors < 10. A "rule of five" was subsequently proposed with respect to drug-likeness. However, these rules may only serve as the minimal criteria for evaluating drug-likeness. The general rules for assessing ADMET properties have been extended to more complex computational and mathematical methods [94].

## 5.2. Virtual screening

Virtual screening (VS) is a computational method for identifying lead compounds from a large and chemically diverse compound library. This computational method is valuable for discovering lead compounds in a faster, more cost-efficient, and less resource-intensive manner compared with experimental methods, such as high-throughput screening. However, the generic definition of VS is significantly wider and may encompass many different methods [95]. VS techniques can be divided into ligand-based and structure-based approaches (see Figure 6). Actually, VS is a combination of several techniques that are applied one after another, similar to a funnel, so it is defined as VS workflow.

Successful VS relies on having a scoring method that assigns good scores to interesting molecules (usually defined as active against a target protein of interest) and worse scores to uninteresting (inactive) molecules. Accordingly, a successful virtual screen will provide a set of compounds for experimental screening that is highly enriched in active molecules. There are a number of approaches to quantifying the success of a particular tool for VS [96,97]. The enrichment factor (EF) represents one of the most prominent performance descriptors in VS. EF is defined as (TP/n) / (A/N); where TP is the number of hits found at x % of the database, n is the number of compounds screened at x % of the database, A is the number of actives in the entire database, and N is the number of compounds in the entire database. Sensitivity and specificity are also descriptors that assess the enrichment of active molecules from a database. Sensitivity (Se, or true positive rate) describes the ratio of the number of active molecules found by the VS method to the number of all active database compounds. Specificity (Sp, or true negative rate) represents the ratio of the number of inactive compounds that were not selected by the VS protocol to the number of all inactive molecules included in the database [98].

The best VS workflow is selected for the prospective VS of large compound libraries. This workflow produces a score ordered hit list of database compounds

that is usually subjected to post-filtering. As a result, a selection of VS hits is obtained that can be tested for activity in biological assays.

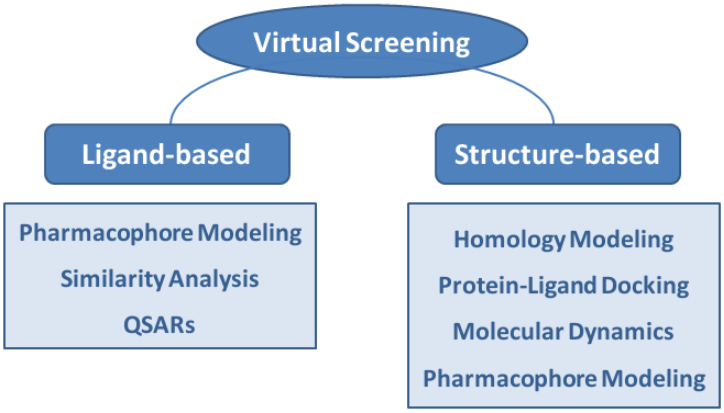



**Figure 6.** Virtual screening methods applied in drug discovery.

## 5.3. Ligand-based approaches

Ligand-based drug design (or indirect drug design) relies on knowledge of other molecules that bind to the biological target of interest. Common ligand-based VS methods are pharmacophore modeling, similarity analysis and QSARs.

### 5.3.1. Ligand-Based Pharmacophore Modeling

According to the definition of Wermuth et al., a pharmacophore describes the 3D arrangement of steric and electronic features that are necessary to trigger or block a biological response [99]. Ligand-based pharmacophore model generation relies on information regarding the known biological activity of ligands without any structural information for the macromolecular target. The elucidation of a shared feature pharmacophore is based on the 3D alignment of the conformational models from active compounds. A molecular superimposition algorithm arranges the 3D structures of the training compounds in such a way that equal chemical functionalities are located in similar positions. Pharmacophoric features are then placed on the positions where all compounds share a chemical functionality. To refine a shared feature pharmacophore, it is also possible to include information from inactive compounds in the model generation process.

Computational models representing pharmacophores have shown unique potential in scaffold hopping. Therefore, pharmacophore models are frequently applied to generate novel starting points for drug design campaigns [100]. A variety

of pharmacophore modeling approaches has been developed, such as Catalyst/Discovery Studio, Phase [101], MOE and LigandScout [102].

### 5.3.2. Similarity Analysis

Similarity search algorithms use techniques such as 2D fingerprints descriptors (Fingerprint Similarity Analysis) or 3D shape descriptors (Electrostatic/Shape Similarity Analysis) to compare a biological active query molecule with a database molecule. 2D fingerprints are bit strings that encode the presence or absence of chemical substructures. Originally developed for chemical substructure search, 2D fingerprints soon became a popular technique for determining molecular similarity between chemical compounds [103]. For the similarity search, the fingerprint of a query molecule is compared with the fingerprint of a database molecule. This comparison is performed using a metric (e.g., Tanimoto coefficient), which expresses the similarity as a score [104]. Popular 2D fingerprint algorithms include Scitegic's Extended Connectivity Fingerprints (ECFPs), MDL's Molecular ACCess System (MACCS), Daylight fingerprints and Molprint2D [105].

One of the most popular algorithms for 3D shape-based similarity searches is Openeye's ROCS. This similarity search algorithm not only compares the molecular shape of two molecules but also identifies similarities in their chemical feature patterns (hydrogen bonds, hydrophobic atoms, anions, cations, and ring moieties). Database molecules aligned by ROCS can be re-scored by the EON algorithm, which determines the electrostatic similarity between query and database molecules.

### 5.3.3. Quantitative Structure-Activity Relationships

QSAR describes the mathematical relationships between the structural attributes and target properties of a set of chemicals [106,107]. QSARs are applied to predict the biological activities or ADMET properties of database molecules with similar chemical structures. This method is only fruitful if the dataset contains compounds that are structurally related to the molecules used to construct the model. Therefore, in contrast to lead discovery techniques, such as similarity analysis and pharmacophore modeling, QSARs are frequently used in the optimization phases of drug design [108]. Many different 1D, 2D, 3D and multidimensional QSAR approaches have been developed during the past several decades [107,109]. The major differences in these methods include the chemical descriptors and mathematical approaches that are used to establish the correlation between the target properties and the descriptors. QSAR models are typically created using a training set of ligands, and the models are then tested against the test set of ligands.

1D-QSARs explain biological activity by correlating it with a single value for a specific physicochemical property (e.g., log P value) of the ligand. 2D-QSARs also

take the structural properties of compounds into account, and the affinity is correlated with structural patterns (connectivity, 2D pharmacophore, etc.) without considering an explicit 3D representation of these properties [109]. In 3D-QSARs, affinity is correlated with the 3D structure of the ligands. Comparative molecular field analysis (CoMFA) [110] is perhaps the most popular example of 3D-QSAR. It describes the steric and electrostatic fields of ligands aligned in their putative bioactive conformation. CoMFA models allow for the prediction of biological activity, as well as 3D visualizations of the steric and electrostatic contributions to protein-ligand binding. The comparative molecular similarity indices analysis (CoMSIA) method calculates three additional molecular field properties (hydrophobicity, hydrogen bond acceptors, and hydrogen bond donors) to generate 3D-QSAR models [111].

An accurate representation of the bioactive conformation of ligands is crucial in 3D-QSAR to obtain the correct ligand alignment. If no experimentally determined bioactive conformation is available, the conformation has to be predicted using protein-ligand docking. mQSAR approaches provide a promising alternative to classic 3D-QSAR for drug-discovery purposes . Such ligands are represented as an ensemble of configurations using 4D-QSAR techniques. 5D-QSAR and 6D-QSAR simulate ligand-induced changes of the binding site or different solvation states, respectively, by calculating different models for each possible scenario [109].

## 5.4. Structure-based approaches

If 3D structural data for a pharmacological target protein is accessible, several structure-based VS techniques can be applied for drug design. In general, such structure-based methods are computationally more expensive than ligand-based VS. However, they provide unique details about protein-ligand interactions and thus are valuable tools for lead discovery and optimization [112].

### 5.4.1. Homology Modeling

The large gap between the number of available sequences and the number of experimentally solved protein structures, which is limited by the cost, time, and experimental challenges inherent to the process of structural determination, could possibly be resolved using homology modeling [113]. In the absence of experimental structures, homology modeling plays an important role in the structure-based drug discovery process. Homology or comparative modeling is a process for predicting protein structure from the general observation that proteins with similar sequences have similar structures. Given an experimentally established protein structure (template), models can be generated for a homologous sequence (target) that either shares a significant sequence (30% or more) or structural similarity (e.g.,

class A GPCRs share a common seven trans-membrane helical structure, despite low sequence homology between family members) with the template. The process of protein homology modeling consists of the following steps: (1) identification of known 3D structure(s) of a related protein that can serve as a template; (2) sequence alignment of the target and template proteins; (3) model building for the target based on the 3D structure of the template and the alignment; and (4) refining/validation/evaluation of the models. These steps may be repeated until a satisfactory model is built [114]. Although homology models are simplifications of the real 3D protein structure and therefore contain errors, their suitability for VS campaigns has been proven [114,115].

### 5.4.2. Protein-Ligand Docking

One of the most common structure-based VS approaches is protein-ligand docking. Molecular docking is used for computational schemes that attempt to find the best matches between a receptor and a ligand. It involves the prediction of ligand conformations and orientation (or posing) within a binding site and attempts to place the ligand into the binding site in configurations and conformations appropriate for interacting with the receptor [116]. The protein-ligand docking process is divided into two major steps: the correct placement of the ligand at the protein binding-site and the estimation of ligand affinity using a scoring function [98].

In theory, the search space consists of all possible orientations and conformations of the protein paired with the ligand. However, in practice, it is impossible to exhaustively explore the search space with current computational resources. Most docking programs account for ligand flexibility, and several attempt to model a flexible protein receptor. Each "snapshot" of the pair is referred to as a pose. A variety of conformational search strategies have been applied to the ligand and to the receptor. These strategies include systematic or stochastic torsional searches about rotatable bonds, molecular dynamics simulations, and genetic algorithms to "evolve" new low energy conformations [117].

The evaluation and ranking of the ligand conformations predicted on the basis of the search algorithm is a critical aspect of every docking protocol [118]. The ability to generate the correct conformation is not sufficient. It is also necessary to be able to recognize it. The scoring function should enable the distinction between the true binding modes and all of the other alternative modes explored, or between active and random compounds. However, a very rigorous scoring function would be computationally too expensive and would thus render the analysis of the several binding modes unfeasible. Hence, a number of assumptions and simplifications are used to reduce the complexity of the scoring functions, with a natural cost in terms

of accuracy. For this reason, the lack of a suitable scoring function, both in terms of speed and accuracy, is the major bottleneck factor in docking simulations [116].

In summary, molecular docking is useful for discriminating active molecules from inactive compounds and to identify ligand conformations similar to the ones observed in the crystal structures of protein-ligand complexes. However, the ranking of compounds in terms of their binding affinities is challenging [119]. Some popular docking software programs are AutoDock, DOCK, eHiTS, FlexX, Fred, GOLD, Glide, MOEDock, and Surflex [118].

### 5.4.3. Molecular Dynamics Simulations

Molecular dynamics (MD) simulations have become increasingly useful in studying biological systems relevant to drug discovery [120,121]. In some cases, the experimentally derived protein structure may not be suitable for structure-based VS. For example, the structure could represent a closed conformation of the protein in which the motion of a hinge region blocks the entrance to the ligand-binding pocket. For docking-based VS, the open conformation of the target protein has to be predicted. Such a prediction of protein conformations can be performed using MD simulation [122]. In addition to determining the open conformation of proteins, conformations induced by co-factor binding can be predicted by MD simulations [123]. With regard to structure-based VS, MD simulations play a pivotal role in understanding the features that are important for ligand-binding affinity. This information could be employed to select higher-affinity ligands from screening processes.

### 5.4.4. Structure-Based Pharmacophore Modeling

Structure-based pharmacophore modeling uses the spatial information regarding the target protein to generate a topological description of ligand-receptor interactions [100]. 3D structural information on the protein is usually obtained from X-ray crystallography or multidimensional nuclear magnet resonance spectroscopy. Starting from the 3D coordinates of a ligand bound to a macromolecular target, possible interactions between the two binding partners are evaluated. It is essential to ensure the reliability of the binding-site residues and ligand coordinates by visually inspecting their degree of fitness to the corresponding electron density map available, for instance, at the Uppsala Electron Density Server [124]. The next step is the manual or automatic analysis of chemical interactions between the ligand and the macromolecule. On the basis of opposing chemical functionalities and their geometric arrangement toward each other, pharmacophore features are placed on the ligand side where interactions are observed. Excluded volume spheres can be placed on binding site atoms to indicate sterically unfavorable regions for a mapped ligand conformation. Examples for the generation and optimization of structure-based

pharmacophore models can be found in the literature [125,126]. Software programs that allow for the manual construction of pharmacophores from protein-ligand complexes include Schrodinger's Phase, Accelrys' Discovery Studio, MOE by the Chemical Computing Group and Inte:Ligand's LigandScout.

Salam et al. described a novel method for generating structure-based pharmacophores using energetic analysis [127]. This method combines pharmacophore perception and database screening with protein-ligand energetic terms computed with a docking scoring function (i.e., Glide XP) to rank the importance of pharmacophore features. The combination of energy terms from a structure-based analysis and the speed of a ligand-based pharmacophore search results in a method that leverages the strengths of both approaches to produce high enrichments with a good diversity of active molecules.

## References

1. Meetoo D, McGovern Peter, Safadi R. (2007) An epidemiological overview of diabetes across the world. British Journal of Nursing 16(16): 1002-1007.

2. Wild S, Roglic G, Green A, Sicree R, King H. (2004) Global prevalence of diabetes. Diabetes Care 27(5): 1047.

3. International diabetes federation (2011) www.idf.org .

4. A DR, Muhammad A. (2011) Type 2 diabetes can be prevented with early pharmacological intervention. Diabetes Care 34 Suppl 2: S202-209.

5. Cernea S, Raz I. (2011) Therapy in the early stage: Incretins. Diabetes Care 34 Suppl 2: S264-271.

6. Marchetti P, Lupi R, Del Guerra S, Bugliani M, D'Aleo Valentina, et al. (2009) Goals of treatment for type 2 diabetes: Beta-cell preservation for glycemic control. Diabetes Care 32 Suppl 2: S178-183.

7. Schwanstecher C, Schwanstecher M. (2011) Targeting type 2 diabetes. In: Schwanstecher M, editor. Diabetes - Perspectives in Drug Therapy. Berlin, Heidelberg: Springer Berlin Heidelberg. pp. 1-33.

8. Bharatam PV, Patel DS, Adane L, Mittal A, Sundriyal S. (2007) Modeling and informatics in designing anti-diabetic agents. Current Pharmaceutical Design 13(34): 3518.

9. Viljoen A, Sinclair AJ. (2011) Diabetes and insulin resistance in older people. Med Clin North Am 95(3): 615-629.

10. Aekplakorn W, Chariyalertsak S, Kessomboon P, Sangthong R, Inthawong R, et al. (2011) Prevalence and management of diabetes and metabolic risk factors in thai adults: The thai national health examination survey IV 2009. Diabetes Care .

11. Holt P. (2011) Taking hypoglycaemia seriously: Diabetes, dementia and heart disease. Br J Community Nurs 16(5): 246-249.

12. Voors AA, van der Horst,Iwan C C. (2011) Diabetes: A driver for heart failure. Heart {(British} Cardiac Society) 97(9): 774-780.

13. Ritz E. (2011) Limitations and future treatment options in type 2 diabetes with renal impairment. Diabetes Care 34 Suppl 2: S330-334.

14. Distefano JK, Watanabe RM. (2010) Pharmacogenetics of anti-diabetes drugs. Pharmaceuticals {(Basel Switzerland) 3(8): 2610-2646.

15. Association AD. (2009) Standards of medical care in diabetes--2009. Diabetes Care 32(Supplement_1): S13-S61.

16. Grether U, Klaus W, Kuhn B, Maerki HP, Mohr P, et al. (2010) New insights on the mechanism of PPAR-targeted drugs. ChemMedChem 5(12):1973-1976.

17. Nissen SE, Wolski K. (2010) Rosiglitazone revisited: An updated meta-analysis of risk for myocardial infarction and cardiovascular mortality. Arch Intern Med 170(14): 1191-1201.

18. Havale SH, Pal M. (2009) Medicinal chemistry approaches to the inhibition of dipeptidyl peptidase-4 for the treatment of type 2 diabetes. Bioorganic & Medicinal Chemistry 17(5): 1783-1802.

19. Executive summary: Standards of medical care in diabetes--2010. (2009) Diabetes Care 33(Supplement_1): S4-S10.

20. Rodbard HW, Jellinger PS. (2010) The american association of clinical Endocrinologists/American college of endocrinology (AACE/ACE) algorithm for managing glycaemia in patients with type 2 diabetes mellitus: Comparison with the ADA/EASD algorithm. Diabetologia 53(11): 2458-2460.

21. Rodbard HW, Jellinger PS, Davidson JA, Einhorn D, Garber AJ, et al. (2009) Statement by an american association of clinical Endocrinologists/American college of endocrinology consensus panel on type 2 diabetes mellitus: An algorithm for glycemic control. Endocrine Practice: Official Journal of the American College of Endocrinology and the American Association of Clinical Endocrinologists 15(6): 540-559.

22. Garber AJ. (2010) Incretin-based therapies in the management of type 2 diabetes: Rationale and reality in a managed care setting. Am J Manag Care 16(7 Suppl): S187-194.

23. Mohler ML, He Y, Wu Z, Hwang DJ, Miller DD. (2009) Recent and emerging anti-diabetes targets. Med Res Rev 29(1): 125-195.

24. Carpino PA, Goodwin B. (2010) Diabetes area participation analysis: A review of companies and targets described in the 2008 - 2010 patent literature. Expert Opinion on Therapeutic Patents 20(12): 1627-1651.

25. Evans JL, Jallal B. (1999) Protein tyrosine phosphatases: Their role in insulin action and potential as drug targets. Expert Opin Investig Drugs 8(2): 139-160.

26. Elchebly M, Payette P, Michaliszyn E, Cromlish W, Collins S, et al. (1999) Increased insulin sensitivity and obesity resistance in mice lacking the protein tyrosine phosphatase-1B gene. Science {(New} York, {N.Y.)} 283(5407): 1544-1548.

27. Klaman LD, Boss O, Peroni OD, Kim JK, Martino JL, et al. (2000) Increased energy expenditure, decreased adiposity, and tissue-specific insulin sensitivity in protein-tyrosine phosphatase 1B-deficient mice. Mol Cell Biol 20(15): 5479-5489.

28. Liu G. (2003) Protein tyrosine phosphatase 1B inhibition: Opportunities and challenges. Curr Med Chem 10(15): 1407-1421.

29. Danforth E, Jr, Himms-Hagen J H. (1997) Obesity and diabetes and the beta-3 adrenergic receptor. Eur J Endocrinol 136(4): 362-365.

30. de Souza CJ, Burkey BF. (2001) Beta 3-adrenoceptor agonists as anti-diabetic and anti-obesity drugs in humans. Curr Pharm Des 7(14): 1433-1449.

31. Mayers RM, Leighton B, Kilgour E. (2005) PDH kinase inhibitors: A novel therapy for type II diabetes? Biochem Soc Trans 33(Pt 2): 367-370.

32. Aicher TD, Damon RE, Koletar J, Vinluan CC, Brand LJ, et al. (1999) Triterpene and diterpene inhibitors of pyruvate dehydrogenase kinase (PDK). Bioorganic & Medicinal Chemistry Letters 9(15): 2223-2228.

33. Gomez R, Navarro M, Ferrer B, Trigo JM, Bilbao A, et al. (2002) A peripheral mechanism for CB1 cannabinoid receptor-dependent modulation of feeding. J Neurosci 22(21): 9612-9617.

34. Juan-Pico Pablo, Fuentes E, Bermadez-Silva F Javier, Javier Diaz-Molina F, Ripoll C, et al. (2006) Cannabinoid receptors regulate ca(2+) signals and insulin secretion in pancreatic beta-cell. Cell Calcium 39(2): 155-162.

35. Lange JHM, Kruse CG. (2005) Keynote review: Medicinal chemistry strategies to CB1 cannabinoid receptor antagonists. Drug Discov Today 10(10): 693-702.

36. Francis GA, Fayard E, Picard F, Auwerx J. (2003) Nuclear receptors and the control of metabolism. Annu Rev Physiol 65: 261-311.

37. Berger J, Moller DE. (2002) The mechanisms of action of PPARs. Annu Rev Med 53: 409-435.

38. Shearer BG, Billin AN. (2007) The next generation of PPAR drugs: Do we have the tools to find them? Biochim Biophys Acta 1771(8): 1082-1093.

39. Feldman PL, Lambert MH, Henke BR. (2008) PPAR modulators and PPAR pan agonists for metabolic diseases: The next generation of drugs targeting peroxisome proliferator-activated receptors? Current Topics in Medicinal Chemistry 8(9): 728-749.

40. Pourcet B, Fruchart JC, Staels B, Glineur C. (2006) Selective PPAR modulators, dual and pan PPAR agonists: Multimodal drugs for the treatment of type 2 diabetes and atherosclerosis. Expert Opin Emerg Drugs 11(3): 379-401.

41. Ahmed I, Furlong K, Flood J, Treat VP, Goldstein BJ. (2007) Dual PPAR alpha/gamma agonists: Promises and pitfalls in type 2 diabetes. Am J Ther 14(1): 49-62.

42. Nolte RT, Wisely GB, Westin S, Cobb JE, Lambert MH, et al. (1998) Ligand binding and co-activator assembly of the peroxisome proliferator-activated receptor-gamma. Nature 395(6698): 137-143.

43. Higgins LS, Depaoli AM. (2010) Selective peroxisome proliferator-activated receptor gamma (PPARgamma) modulation as a strategy for safer therapeutic PPARgamma activation. Am J Clin Nutr 91(1): 267S-272S.

44. Zoete V, Grosdidier A, Michielin O. (2007) Peroxisome proliferator-activated receptor structures: Ligand specificity, molecular switch and interactions with regulators. Biochim Biophys Acta 1771(8): 915-925.

45. Uppenberg J, Svensson C, Jaki M, Bertilsson G, Jendeberg L, et al. (1998) Crystal structure of the ligand binding domain of the human nuclear receptor PPARgamma. The Journal of Biological Chemistry 273(47): 31108-31112.

46. Farce A, Renault N, Chavatte P. (2009) Structural insight into PPARgamma ligands binding. Curr Med Chem 16(14): 1768-1789.

47. Bruning JB, Chalmers MJ, Prasad S, Busby SA, Kamenecka TM, et al. (2007) Partial agonists activate PPARgamma using a helix 12 independent mechanism. Structure 15(10): 1258-1271.

48. Pochetti G, Godio C, Mitro N, Caruso D, Galmozzi A, et al. (2007) Insights into the mechanism of partial agonism: Crystal structures of the peroxisome proliferator-activated receptor gamma ligand-binding domain in the complex with two enantiomeric ligands. J Biol Chem 282(23): 17314-17324.

49. Lu IL, Huang CF, Peng YH, Lin YT, Hsieh HP, et al. (2006) Structure-based drug design of a novel family of PPARgamma partial agonists: Virtual screening, X-ray crystallography, and in vitro/in vivo biological activities. J Med Chem 49(9): 2703-2712.

50. Gelman L, Feige JN, Desvergne B. (2007) Molecular basis of selective PPARgamma modulation for the treatment of type 2 diabetes. Biochim Biophys Acta 1771(8): 1094-1107.

51. Choi JH, Banks AS, Estall JL, Kajimura S, Boström P, et al. (2010) Anti-diabetic drugs inhibit obesity-linked phosphorylation of PPARgamma by Cdk5. Nature 466(7305): 451-456.

52. Jones D. (2010) Potential remains for PPAR-targeted drugs. Nature Reviews.Drug Discovery 9(9): 668-669.

53. Yazbeck R, Howarth GS, Abbott CA. (2009) Dipeptidyl peptidase inhibitors, an emerging drug class for inflammatory disease? Trends Pharmacol Sci 30(11): 600-607.

54. Mentlein R, Gallwitz B, Schmidt WE. (1993) Dipeptidyl-peptidase IV hydrolyses gastric inhibitory polypeptide, glucagon-like peptide-1(7-36)amide, peptide histidine methionine and is responsible for their degradation in human serum. European Journal of Biochemistry {FEBS} 214(3): 829-835.

55. Brubaker PL, Drucker DJ. (2004) Minireview: Glucagon-like peptides regulate cell proliferation and apoptosis in the pancreas, gut, and central nervous system. Endocrinology 145(6): 2653-2659.

56. Meier JJ, Nauck MA, Schmidt WE, Gallwitz B. (2002) Gastric inhibitory polypeptide: The neglected incretin revisited. Regul Pept 107(1-3): 1-13.

57. Kuhn B, Hennig M, Mattei P. (2007) Molecular recognition of ligands in dipeptidyl peptidase IV. Current Topics in Medicinal Chemistry 7(6): 609-619.

58. Zettl H, Schubert-Zsilavecz Manfred, Steinhilber D. (2010) Medicinal chemistry of incretin mimetics and DPP-4 inhibitors. ChemMedChem 5(2): 179-185.

59. Butler MS. (2004) The role of natural product chemistry in drug discovery. J Nat Prod 67(12): 2141-2153.

60. Drews J. (2000) Drug discovery: A historical perspective. Science {(New} York, {N.Y.)} 287(5460): 1960-1964.

61. Schuster D, Wolber G. (2010) Identification of bioactive natural products by pharmacophore-based virtual screening. Curr Pharm Des 16(15): 1666-1681.

62. Clardy J, Walsh C. (2004) Lessons from natural molecules. Nature 432(7019): 829-837.

63. Hong J. (2011) Role of natural product diversity in chemical biology. Curr Opin Chem Biol 15(3): 350-354.

64. Rollinger JM, Stuppner H, Langer T. (2008) Virtual screening for the discovery of bioactive natural products. Progress in Drug Research.Fortschritte Der Arzneimittelforschung.ProgrÃ¨s Des Recherches Pharmaceutiques 65: 211, 213-249.

65. Rollinger JM, Langer T, Stuppner H. (2006) Strategies for efficient lead structure discovery from natural products. Curr Med Chem 13(13): 1491-1507.

66. Akhilesh K. Tripathi , Pravin K. Bhoyar, Jagdish R. Baheti , Dinesh M. Biyani , M. Khalique , Mayuresh S. Kothmire , Yogesh M. Amgaonkar,Anand B.Bhanarkar. (2011) Herbal antidiabetics: A review. International Journal of Research in Pharmaceutical Sciences 2(1): 30-37.

67. Soumyanath A. (2005) Traditional medicines for modern times: Antidiabetic plants. Press.

68. Bailey CJ, Day C. (1989) Traditional plant medicines as treatments for diabetes. Diabetes Care 12(8): 553-564.

69. Howes M, Simmonds M. (2005) Plants used in the treatment of diabetes. In: Soumyanath A, editor. Traditional Medicines for Modern Times. : CRC} Press.

70. Modak M, Dixit P, Londhe J, Ghaskadbi S, Paul A Devasagayam T. (2007) Indian herbs and herbal drugs used for the treatment of diabetes. Journal of Clinical Biochemistry and Nutrition 40(3): 163-173.

71. Xie W, Zhao Y, Zhang Y. (2011) Traditional chinese medicines in treatment of patients with type 2 diabetes mellitus. Evidence-Based Complementary and Alternative Medicine: {eCAM} 2011: 726723.

72. Andrade-Cetto Adolfo, Heinrich M. (2005) Mexican plants with hypoglycaemic effect used in the treatment of diabetes. J Ethnopharmacol 99(3): 325-348.

73. Jung SH, Seol HJ, Jeon SJ, Son KH, Lee JR. (2009) Insulin-sensitizing activities of tanshinones, diterpene compounds of the root of salvia miltiorrhiza bunge. Phytomedicine 16(4): 327-335.

74. Bnouham M, Ziyyat A, Mekhfi H, Tahri A, Legssyer A. (2006) Medicinal plants with potential antidiabetic activity - A review of ten years of herbal medicine research (1990-2000). Int J Diabetes & Metabolism 14(1).

75. Romila Y, P. B. Mazumder,M.Dutta Choudhury. (2010) A review on antidiabetic plants used by the people of manipur charactirized by hypoglycemic activity. .

76. Roberfroid MB. (1999) What is beneficial for health? the concept of functional food. Food and Chemical Toxicology: An International Journal Published for the British Industrial Biological Research Association 37(9-10): 1039-1041.

77. Lockwood GB. (2011) The quality of commercially available nutraceutical supplements and food sources. J Pharm Pharmacol 63(1): 3-10.

78. Riccardi G, Capaldo B, Vaccaro O. (2005) Functional foods in the management of obesity and type 2 diabetes. Curr Opin Clin Nutr Metab Care 8(6): 630-635.

79. Dämon S, SchÃ¤tzer M, Häfler J, Tomasec G, Hoppichler F. (2011) Nutrition and diabetes mellitus: An overview of the current evidence. Wien Med Wochenschr 161(11-12): 282-288.

80. Rudkowska I. (2009) Functional foods for health: Focus on diabetes. Maturitas 62(3): 263-269.

81. Knowler WC, Barrett-Connor Elizabeth, Fowler SE, Hamman RF, Lachin JM, et al. (2002) Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. N Engl J Med 346(6): 393-403.

82. Tuomilehto J, Lindström J, Eriksson JG, Valle TT, Hämäläinen H, et al. (2001) Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. N Engl J Med 344(18): 1343-1350.

83. Jiang R, Manson JE, Stampfer MJ, Liu S, Willett WC, et al. (2002) Nut and peanut butter consumption and risk of type 2 diabetes in women. JAMA the Journal of the American Medical Association 288(20): 2554-2560.

84. Giacco R, Cuomo V, Vessby B, Uusitupa M, Hermansen K, et al. (2007) Fish oil, insulin sensitivity, insulin secretion and glucose tolerance in healthy people: Is there any effect of fish oil supplementation in relation to the type of background diet and habitual dietary intake of n-6 and n-3 fatty acids? Nutrition, Metabolism, and Cardiovascular Diseases: {NMCD} 17(8): 572-580.

85. Khan A, Safdar M, Ali Khan MM, Khattak KN, Anderson RA. (2003) Cinnamon improves glucose and lipids of people with type 2 diabetes. Diabetes Care 26(12): 3215-3218.

86. Pham AQ, Kourlas H, Pham DQ. (2007) Cinnamon supplementation in patients with type 2 diabetes mellitus. Pharmacotherapy 27(4): 595-599.

87. Song CM, Lim SJ, Tong JC. (2009) Recent advances in computer-aided drug design. Briefings in Bioinformatics 10(5): 579-591.

88. Workman P. (2003) How much gets there and what does it do?: The need for better pharmacokinetic and pharmacodynamic endpoints in contemporary drug discovery and development. Curr Pharm Des 9(11): 891-902.

89. Brown D, Superti-Furga Giulio. (2003) Rediscovering the sweet spot in drug discovery. Drug Discov Today 8(23): 1067-1077.

90. Zhang S, Lu W, Liu X, Diao Y, Bai F, et al. (2011) Fast and effective identification of the bioactive compounds and their targets from medicinal plants via computational chemical biology approach. Med.Chem.Commun. 2(6): 471-477.

91. Veselovsky AV, Ivanov AS. (2003) Strategy of computer-aided drug design. Curr Drug Targets Infect Disord 3(1): 33-40.

92. Di L, Kerns EH, Carter GT. (2009) Drug-like property concepts in pharmaceutical design. Curr Pharm Des 15(19): 2184-2194.

93. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 46(1-3): 3-26.

94. Lagorce D, Sperandio O, Galons H, Miteva MA, Villoutreix BO. (2008) FAF-Drugs2: Free ADME/tox filtering tool to assist drug discovery and chemical biology projects. BMC Bioinformatics 9: 396.

95. Sousa SF, Cerqueira,Nuno M F S A., Fernandes PA, Ramos MJ. (2010) Virtual screening in drug design and development. Combinatorial Chemistry \& High Throughput Screening 13(5): 442-453.

96. Truchon J, Bayly CI. (2007) Evaluating virtual screening methods: Good and bad metrics for the early recognition problem. Journal of Chemical Information and Modeling 47(2): 488-508.

97. Hawkins PCD, Warren GL, Skillman AG, Nicholls A. (2008) How to do an evaluation: Pitfalls and traps. Journal of {Computer-Aided} Molecular Design 22(3-4): 179-190.

98. Kirchmair J, Markt P, Distinto S, Wolber G, Langer T. (2008) Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment

assessments, and decoy selection--what can we learn from earlier mistakes? Journal of {Computer-Aided} Molecular Design 22(3-4): 213-228.

99. Wermuth C, Ganellin CR, Lindberg P, Mitscher LA, James A. Bristol. (1998) Chapter 36. glossary of terms used in medicinal chemistry (IUPAC recommendations 1997. In: Anonymous : Academic Press. pp. 385-395.

100. Leach AR, Gillet VJ, Lewis RA, Taylor R. (2010) Three-dimensional pharmacophore methods in drug discovery. J Med Chem 53(2): 539-558.

101. Dixon SL, Smondyrev AM, Knoll EH, Rao SN, Shaw DE, et al. (2006) PHASE: A new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. methodology and preliminary results. J Comput Aided Mol Des 20(10-11): 647-671.

102. Wolber G, Langer T. (2005) LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. Journal of Chemical Information and Modeling 45(1): 160-169.

103. Leach AR, Gillet VJ. (2003) An introduction to chemoinformatics. Springer.

104. Willett P. (2006) Similarity-based virtual screening using 2D fingerprints. Drug Discov Today 11(23-24): 1046-1053.

105. Duan J, Dixon SL, Lowrie JF, Sherman W. (2010) Analysis and comparison of 2D fingerprints: Insights into database screening performance using eight fingerprint methods. Journal of Molecular Graphics & Modelling 29(2): 157-170.

106. Perkins R, Fang H, Tong W, Welsh WJ. (2003) Quantitative structure-activity relationship methods: Perspectives on drug discovery and toxicology. Environmental Toxicology and Chemistry 22(8): 1666-1679.

107. Verma J, Khedkar VM, Coutinho EC. (2010) 3D-QSAR in drug design--a review. Current Topics in Medicinal Chemistry 10(1): 95-115.

108. Fischer PM. (2008) Computational chemistry approaches to drug discovery in signal transduction. Biotechnology Journal 3(4): 452-470.

109. Lill MA. (2007) Multi-dimensional QSAR in drug discovery. Drug Discov Today 12(23-24): 1013-1017.

110. Cramer RD,3rd, Patterson DE, Bunce JD. (1989) Recent advances in comparative molecular field analysis (CoMFA). Prog Clin Biol Res 291: 161-165.

111. Klebe G, Abraham U. (1999) Comparative molecular similarity index analysis (CoMSIA) to study hydrogen-bonding properties and to score combinatorial libraries. Journal of {Computer-Aided} Molecular Design 13(1): 1-10.

112. Waszkowycz B. (2008) Towards improving compound selection in structure-based virtual screening. Drug Discov Today 13(5-6): 219-226.

113. Eswar N, Eramian D, Webb B, Shen M, Sali A. (2008) Protein structure modeling with MODELLER. Methods in Molecular Biology 426: 145-159.

114. Cavasotto CN, Phatak SS. (2009) Homology modeling in drug discovery: Current trends and applications. Drug Discov Today 14(13-14): 676-683.

115. Miguet L, Zhang Z, Barbier M, Grigorov MG. (2006) Comparison of a homology model and the crystallographic structure of human 11beta-hydroxysteroid dehydrogenase type 1 (11betaHSD1) in a structure-based identification of inhibitors. Journal of {Computer-Aided} Molecular Design 20(2): 67-81.

116. Kitchen DB, Decornez H, Furr JR, Bajorath J. (2004) Docking and scoring in virtual screening for drug discovery: Methods and applications. Nature Reviews.Drug Discovery 3(11): 935-949.

117. Dias R, de Azevedo WF. (2008) Molecular docking algorithms. Curr Drug Targets 9(12): 1040-1047.

118. Huang S, Grinter SZ, Zou X. (2010) Scoring functions and their evaluation methods for protein-ligand docking: Recent advances and future directions. Physical Chemistry Chemical Physics: {PCCP} 12(40): 12899-12908.

119. Warren GL, Andrews CW, Capelli A, Clarke B, LaLonde Judith, et al. (2006) A critical assessment of docking programs and scoring functions. J Med Chem 49(20): 5912-5931.

120. Salsbury FR. (2010) Molecular dynamics simulations of protein dynamics and their relevance to drug discovery. Current Opinion in Pharmacology 10(6): 738-744.

121. Perez-Sanchez Horacio, Wenzel W. (2011) Optimization methods for virtual screening on novel computational architectures. Current {Computer-Aided} Drug Design 7(1): 44-52.

122. Marco E, Gago F. (2007) Overcoming the inadequacies or limitations of experimental structures as drug targets by using computational modeling tools and molecular dynamics simulations. ChemMedChem} 2(10): 1388-1401.

123. Amaro RE, Li WW. (2010) Emerging methods for ensemble-based virtual screening. Current Topics in Medicinal Chemistry 10(1): 3-13.

124. Kleywegt GJ, Harris MR, Zou JY, Taylor TC, Wahlby A, et al. (2004) The uppsala electron-density server Acta Crystallogr D Biol Crystallogr 60(Pt 12 Pt 1): 2240-2249.

125. Markt P, Schuster D, Kirchmair J, Laggner C, Langer T. (2007) Pharmacophore modeling and parallel screening for PPAR ligands. J Comput Aided Mol Des 21(10-11): 575-590.

126. Barreca ML, De Luca L, Iraci N, Rao A, Ferro S, et al. (2007) Structure-based pharmacophore identification of new chemical scaffolds as non-nucleoside reverse transcriptase inhibitors. Journal of Chemical Information and Modeling 47(2): 557-562.

127. Salam NK, Nuti R, Sherman W. (2009) Novel method for generating structure-based pharmacophores using energetic analysis. Journal of Chemical Information and Modeling 49(10): 2356-2368.

## CONTEXT AND GOALS

In the past decade, there has been increased interest in nutrition as a natural way to improve quality of life and ameliorate different pathological states. Functional foods and nutraceuticals provide an opportunity to obtain a wide variety of health benefits from bioactive compounds.

Type 2 diabetes mellitus (T2DM) is considered to be the "epidemic of the 21st century" and, consequently, is one of the main challenges in drug discovery today. T2DM is a multifactorial disease that is characterized by insulin resistance associated not only with hyperinsulinaemia and hyperglycemia but also with artherosclerosis, hypertension and an abnormal lipid profile. While current T2DM therapies that increase insulin secretion have been shown to have therapeutically beneficial effects, these often suffer from undesirable side effects, such as hypoglycemia and weight gain. Thus, there is a substantial unmet medical need for better drugs to treat T2DM. In recent years, computer-aided drug design (CADD) methodologies have contributed significantly to the discovery and/or optimization of many clinically used drugs. Drug discovery and development is a time-consuming and expensive process. CADD helps scientists focus on the most promising compounds so that they can minimize the synthetic and biological testing efforts. Natural products have played an important role in pharmaceutical research, not only from the point of view of the discovery of active principles, but also in the research of substances that could be used as lead compounds during the development of new drugs. Thus, their potential use in the development of new functional foods for specific population sectors is very promising. Unfortunately, the identification of novel bioactivities for natural extracts via in vitro or in vivo approaches is a complex and expensive process. Thus, virtual screening workflows may play an essential role in significantly lowering the research and development expenses associated with this identification.

This doctoral thesis has been developed in the Chemoinformatics Unit of the Nutrigenomic Research Group of the Biochemistry and Biotechnology Department of the 'Rovira i Virgili' University of Tarragona. The research interests of this unit are to study the bioactivity of new potential ingredients for functional food design and to describe the mechanisms of action for these new bioactive compounds in order to prevent/reduce the metabolic risk factors associated with metabolic syndromes. In particular, this thesis focuses on the identification of new natural compounds as antidiabetic drugs. Nuclear receptor peroxisome proliferator-activating receptor γ (PPARγ) and the enzyme inhibitor dipeptidyl peptidase-IV (DPP-IV) have been shown to be appropriate targets for antidiabetic drugs, and targeting them appears to have good prospects for a successful therapeutic approach for treating T2DM.

Our specific objectives were:

1. **to contribute to the knowledge of the structural-activity relationships of PPARγ full agonists**
   To achieve this objective, it will be necessary to develop and validate two 3D-QSAR models that correlate the structures of known PPARγ full agonists with their binding affinities and transactivation activities.

2. **to explore the different binding features of full and partial PPARγ agonists**
   This will be necessary to predict molecules that could act as PPARγ partial agonists but not as PPARγ full agonists.

3. **to develop and validate a virtual screening (VS) workflow to predict natural molecules that can act as PPARγ partial agonists**
   The predicted molecules will be candidates for use in functional foods or will be lead-hopping candidates for the design of new antidiabetic compounds with fewer side effects than PPARγ full agonists.

4. **to select some of the natural molecules that are predicted to be PPARγ partial agonists and to validate their bioactivities in vitro**
   The clustering of the VS hits with known PPARγ partial agonists will allow for the selection of new chemical scaffolds that are unlike the known PPARγ partial agonists and the validation of their bioactivity in vitro.

5. **to identify some natural extracts with known antidiabetic activity that contain at least one molecule that we predict to be a PPARγ partial agonist**
   The application of the previously developed VS workflow to a natural product database that contains the natural source(s) of each molecule and the identification of which VS hits are present in extracts with known antidiabetic activity will allow for the achievement of this objective.

6. **to develop and validate a VS workflow to predict natural molecules that can act as DPP-IV inhibitors**
   The predicted molecules will be used as lead compounds in drug-design projects or in functional foods with antidiabetic properties.

7. **to select some natural molecules that are predicted to be DPP-IV inhibitors and to validate their bioactivities in vitro**
   These molecules will allow us to verify the reliability of our prediction using an in vitro test of the inhibitory effect of some of the selected VS hits on DPP-IV.

8. **to identify some natural extracts with known antidiabetic activity that contain at least one molecule that we predict to be a DPP-IV inhibitor**
   The application of the previously developed VS workflow to a natural product database that contains the natural source(s) of each molecule and the identification of which VS hits are present in extracts with known antidiabetic activity will allow us to achieve this objective.

CHAPTER 1

# DEVELOPMENT OF DOCKING-BASED 3D-QSAR MODELS FOR PPARγ FULL AGONISTS

## ABSTRACT

Peroxisome proliferator-activated receptor gamma (PPARγ) has become an attractive molecular target for drugs that aim to treat diabetes mellitus type II, and its therapeutic potency against skin cancer and other skin diseases is also currently being explored. To study the relationship between the structure of several PPARγ full agonists and the trans-activation activity of PPARγ, we have performed a three-dimensional quantitative structure-activity relationship (3D-QSAR) study of tyrosine-based derivatives, based on the 3D alignment of conformations obtained by docking. Highly predictive 3D-QSAR models, with Pearson-R values of 0.86 and 0.90, were obtained for the transactivation activity and binding affinity of PPARγ, respectively. These models are in good agreement with the structural characteristics of the binding pocket of PPARγ. These results may be useful for the prediction of the trans-activation activities of new PPARγ full agonists, and they may also help derive insights to improve the bioactivities of the currently known PPARγ agonists.

## Introduction

Peroxisome proliferator-activated receptors (PPAR) are fatty acid-activated transcription factors that belong to the nuclear hormone receptor family [1,2]. Three PPAR isotypes, PPARα, PPARβ/δ and PPARγ, have previously been identified. Each of these subtypes appears to be differentiated in a tissue-specific manner and plays a pivotal role in glucose and lipid homeostasis [3,4]. PPARγ constitutes a primary target for the development of drug candidates for the treatment of type II diabetes. Thiazolidinediones (TZDs) represent the first known PPARγ agonists used as oral antidiabetic agents [4,5]. In addition, several studies have suggested that oral PPARγ full agonists not only exert an antidiabetic effect but also may act as a promising therapeutic target for a broad variety of skin disorders, including inflammatory skin diseases, such as psoriasis and atopic dermatitis, melanoma and other skin malignancies [6-9]. Furthermore, PPARγ full agonists may even induce cell growth arrest, apoptosis and terminal differentiation in various human malignant tumors [7]. There are several synthetic PPARγ full agonists besides TZDs with high potency and selectivity [10-14].

Over the past decade, a number of protein structures of the PPARγ ligand-binding domain (LBD), co-crystallized with ligands or in the apo-form, have been resolved by X-ray crystallography [4,15]. The binding pocket of PPARγ is very large and has a Y-shaped form, consisting of an entrance (arm III) that branches off into two pockets [16]. Arm I is extended toward H12, and arm II is situated between helix H3 and a β-sheet [16]. Arm I is the only substantially polar cavity of the PPARγ LBD, whereas arms II and III are mainly hydrophobic. To show biological activity, only two arms need to interact with the ligand; therefore, PPARγ full agonists occupy arms I and II [17].

It is expected that the use of quantitative structure-activity relationship (QSAR) approaches could correlate the observed biological activities with structural changes of the ligands [18]. A number of QSAR studies on PPARγ agonists have been performed [19-21], some of which have been applied to TZD sets [22,23] and others have been performed with non-TZD sets such as tyrosine-based structures [19-21]. Although some 3D-QSAR studies of non-TZDs have been reported, none of them have used molecular docking to align the molecules and generate the 3D-QSAR model. In addition, none of them have analyzed, in parallel, the binding affinity and transactivation activity of the compounds analyzed. The present study aims to expand the knowledge of structure-activity relationships of PPARγ full agonists by using tyrosine-based derivatives for developing two 3D-QSAR models that: a) correlate the binding affinity and transactivation activity with the structures of the agonists used to develop the models; and b) are able to predict the $pIC_{50}$ and $pEC_{50}$ of a set of other

PPARγ full agonists not used during the development of the models. The resulting models also provide some structural insights for the improvement of PPARγ full agonist bioactivities.

## Materials and Methods

**Datasets.** A dataset of 49 tyrosine-based compounds with measured $pK_i$ (*i.e.*, binding affinity) and $pEC_{50}$ (*i.e.*, transactivation activity) values obtained from the same laboratory [10-12] was used to generate two 3D-QSAR models (see Supporting Information Table 1 and Figure 1A). The chemical structures of these 49 compounds are unequivocally known (*i.e.*, there are either no chiral atoms in their structure or the chirality of the molecules is defined), their $pEC_{50}$ and pKi values span six and five orders of magnitude, respectively, and each order of magnitude is represented by several compounds. Of the 49 molecules, 25 were randomly assigned to the training set, whereas the remaining 24 molecules were assigned to the test set. An additional set of 6 thiazolidinediones [10] and 68 indanyacetic acid derivates (for which only $pEC_{50}$ values were available) [14] were used as an external validation set (see Supporting Information Tables 2 and 3, Figure 1B and 1C).

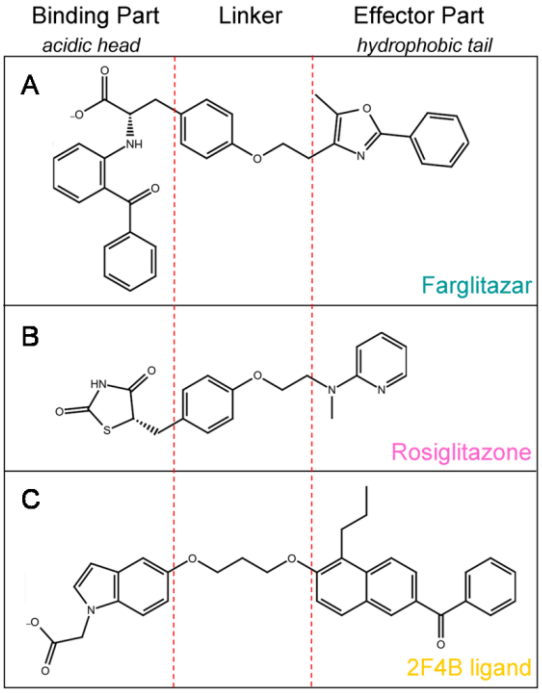



**Figure 1.** Schematic representation of the common parts of PPARγ full agonists.

All compound structures were built with ChemDraw Ultra v11.0 (CambridgeSoft Corporation, Cambrigde, MA, USA; http://www.cambridgesoft.com), and their 3D structures were further minimized with the LigPrep v2.4 program (Schrödinger LLC., Portland, USA; http://www.schrodinger.com), using the OPLS_2005 force field at pH 7.0 and the rest of the parameter values by default.

**Molecular Alignments.** The most crucial step for a 3D-QSAR construction model is the alignment of the molecules. We chose a structure-based docking strategy that was carried out using the poses predicted by docking using the Glide v5.6 program (Schrödinger LLC., Portland, USA; http://www.schrodinger.com). The 49 tyrosine-based PPARγ full agonists were docked within the binding site of the 1FM9 structure (see Figure 2). Meanwhile the 6 thiazolidinediones and the 68 indanyacetic acid derivates used as an external set were docked within the binding sites of the 1FM6 and 2F4B structures, respectively. The binding site was defined using the *Receptor Grid Generation* panel with the default options. Standard-precision (SP) docking was selected for screening the ligands. We selected the flexible docking mode, in which the Glide program generates conformations internally during the docking process. We did not request any constraint for docking. Each docking run generated at most twenty poses per ligand that survived the post-docking minimization process. The GlideScore was used as a function of fitness. The best scoring pose was selected for each ligand and used as an input structure for the subsequent 3D-QSAR analysis.

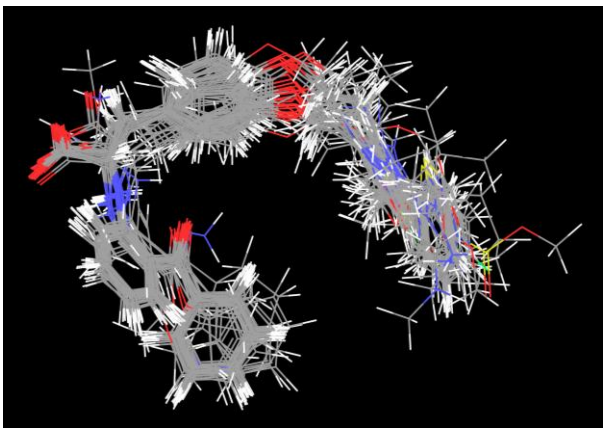


**Figure 2.** Structural alignment of the selected docking poses of the 49 tyrosine-based PPARγ full agonists used to develop the 3D-QSAR models.

**Generation of the 3D-QSAR models.** The selected conformations of the ligands, obtained with the previously described alignment protocol, were used for the generation of two 3D-QSAR models, one for $pIC_{50}$ and the other for $pEC_{50}$. The Phase v3.2 program (Schrödinger LLC., Portland, USA; http://www.schrodinger.com) was

utilized to carry out the calculations using the *Atom-Based QSAR* panel. In atom-based QSAR, a molecule is treated as a set of overlapping van der Waals spheres. To encode the basic characteristics of the local chemical structure, each atom (and hence each sphere) is placed into one of six categories according to a simple set of rules: a) hydrogens attached to polar atoms are classified as hydrogen bond donors (D); b) carbons, halogens and C–H hydrogens are classified as hydrophobic/non-polar (H); c) atoms with an explicit negative ionic charge are classified as negative ionic (N); d) atoms with an explicit positive ionic charge are classified as positive ionic (P); e) non-ionic nitrogens and oxygens are classified as electron-withdrawing (W); and f) all other types of atoms are classified as miscellaneous (X) [24]. The QSAR model partitions the space occupied by the ligands into a cubic grid. Any structural component can occupy part of one or more cubes. The cube size that was selected was 1Å. The independent variables in the regression are given by the binary-valued occupancies ("bits") of the cubes (by structural components), while the dependent variables are the transactivation activity or the binding affinity. The regression is performed by constructing a series of models with an increasing number of partial least square (PLS) factors. The accuracy of the models increases when the number of PLS factors increases, until over-fitting begins to occur.

**Statistical validations of the QSAR models.** The performance of the QSAR models was evaluated by measuring the accuracy of the predictions. The statistical parameters that were used to evaluate the predictions for the training set were as follows: a) the coefficient of determination ($R^2$); b) the standard deviation of regression (SD); c) the F statistic that measures the overall significance of the model; d) the statistical significance (P) that measures the probability that the correlation could occur by chance; and e) a s*tability* value that has a maximum value of 1 and measures the stability of the model predictions to changes in the training set composition. The parameters used to evaluate the predictions for the test set were as follows: a) the $Q^2$ (*i.e.*, the equivalent of the $R^2$ for the test set); b) the root-mean-square error (RMSE); and c) the Pearson correlation coefficient (r).

## Results and discussion

**Datasets.** Although different SAR studies of PPARγ full agonists have been reported, we only selected the studies that contain compounds with a wide range of both transactivation and binding activity values. The dataset that we selected contained 49 tyrosine-based PPARγ full agonists [10-12], and Figure 3 shows the correlation ($R^2$=0.6448) between the experimental transactivation activity ($pEC_{50}$) and the experimental binding activity ($pIC_{50}$) of these 49 tyrosine-based PPARγ full agonists. A relationship between the two variables is expected, because when the binding affinity increases, the AF-2 domain will be better stabilized, thus increasing the transactivation activity of PPARγ. However, some compounds, such as the compounds

**3**, **4**, **6** and **21_2**, are slightly away from the line of best fit of Figure 3. Because not all the molecular interactions between PPARγ agonists and PPARγ contribute to the stabilization of the AF-2 domain and to the trans-activation activity, there are different profiles of binding, such as those presented by partial agonists [17]. For this reason, it is interesting to study in more detail the structural differences between the features of PPARγ full agonists used for binding and those needed to activate the trans-activation activity of PPARγ. To do so, two 3D-QSAR models, one for the binding affinity and the other for the trans-activation activity were developed and compared.

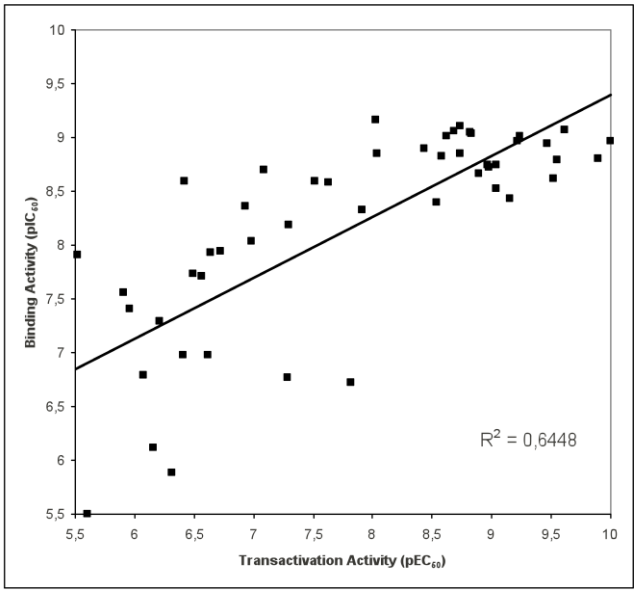



**Figure 3.** Correlation between the experimental transactivation activity and the experimental binding activity of the 49 tyrosine-based PPARγ full agonists used for the construction of the 3D-QSAR models.

**Molecular Alignment.** A structure-based docking strategy was adopted for aligning the 49 tyrosine-based PPARγ full agonists analyzed. This set of PPARγ full agonists follows a common structural pattern (see Figure 1) that consists of a three-module structure, comprising an acidic head linked to an aromatic center and a hydrophobic tail. The 49 molecules were docked into the crystal structure of PPARγ, and the highest scoring pose was selected for each of the molecules. These selected poses are predicted to be the most stable conformation of each molecule for binding to the PPARγ active site. All of the selected poses of the 49 analyzed molecules were visually inspected to demonstrate that they were able to establish the molecular interactions that can establish other PPARγ full agonists. These interactions include several hydrogen bonds with residues Ser289, His323, His449 and Tyr473 from arm I

of the LBD of PPARγ, and several hydrophobic interactions with residues Phe360 and Ile456 from arm I and Ile281, Leu330, Ile341, Leu353 and Met364 from arm II of the LBD of PPARγ [17] (Figure 4). Figure 2 represents the final alignment of this dataset. The chemical similarity between the compounds analyzed, their excellent alignment and the evidence that they can interact with the LBD of PPARγ similarly to other PPARγ full agonists, ensuring that a 3D-QSAR model can be obtained from all the selected compounds.

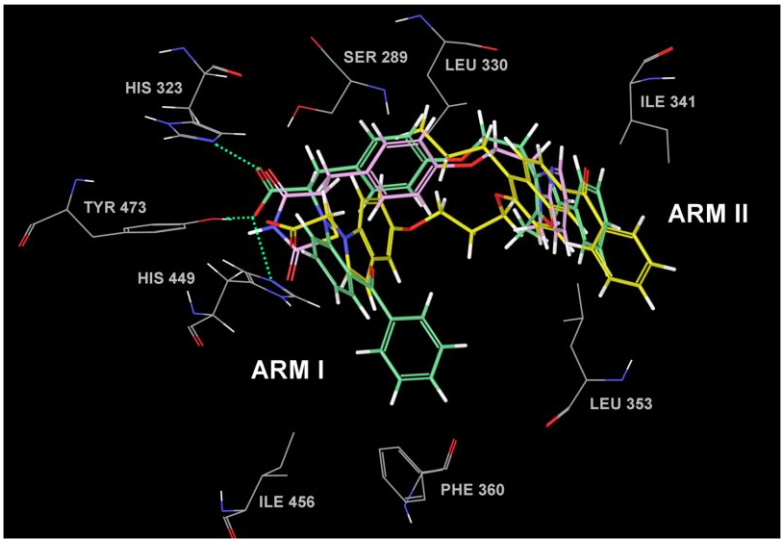



**Figure 4.** The main interactions between PPARγ full agonists (farglitazar colored in green, rosiglitazone colored in pink and 2F4B ligand colored in yellow) and the LBD of PPARγ. The acidic head of the ligands is involved in a hydrogen bond network with residues of arm I (*i.e.*, Ser289, His323, His449 and Tyr473) whereas the remaining residues make hydrophobic interactions. Hydrogen bonds of the carboxylic acid from farglitazar are shown by dashed green lines.

**3D-QSAR Models.** The aligned compounds were used to generate two 3D-QSAR models, one for analyzing the binding affinity between the ligands and PPARγ (namely the $pIC_{50}$ model) and the other for analyzing the transactivation activity of PPARγ (namely the $pEC_{50}$ model). Table 1 and Figure 5 show the statistic of the constructed 3D-QSAR models. In both models, to avoid an over-fitting effect, two PLS factors were chosen. The Pearson correlation coefficient of the $pEC_{50}$ model was 0.8625 with an $R^2$ of 0.9049 for the training set and a $Q^2$ of 0.6966 for the test set. For the $pIC_{50}$ model, the Pearson correlation coefficient was 0.9035 with an $R^2$ of 0.9223 for the training set and a $Q^2$ of 0.6385 for the test set.

**Table 1.** Statistics of the best 3D-QSAR models for analyzing the transactivation activity of PPARγ ($pEC_{50}$ model) and the binding affinity ($pIC_{50}$ model) derived from an 50% randomly selected training set. To avoid an over-fitting effect, two factor models (marked in bold) were chosen.

| Model | # | SD | $R^2$ | F | P | Stability | RMSE | $Q^2$ | PearsonR |
|---|---|---|---|---|---|---|---|---|---|
| **$pEC_{50}$** | 1 | 0.92 | 0.63 | 39.0 | 2.28e-06 | 0.726 | 0.91 | 0.59 | 0.8381 |
| | **2** | **0.47** | **0.90** | **104.6** | **5.78e-12** | **0.240** | **0.78** | **0.70** | **0.8625** |
| | 3 | 0.27 | 0.97 | 233.6 | 2.77e-16 | 0.161 | 0.84 | 0.68 | 0.8383 |
| | 4 | 0.17 | 0.99 | 434.3 | 3.97e-19 | 0.082 | 0.90 | 0.59 | 0.8042 |
| | 5 | 0.11 | 0.99 | 818.0 | 1.73e-21 | 0.063 | 0.91 | 0.59 | 0.7982 |
| **$pIC_{50}$** | 1 | 0.56 | 0.63 | 39.0 | 2.28e-06 | 0.723 | 0.74 | 0.53 | 0.9078 |
| | **2** | **0.26** | **0.92** | **130.6** | **6.2e-13** | **0.164** | **0.65** | **0.64** | **0.9035** |
| | 3 | 0.12 | 0.99 | 471.8 | 2.03e-19 | 0.070 | 0.66 | 0.63 | 0.8998 |
| | 4 | 0.07 | 0.99 | 1023.8 | 8.05e-23 | 0.082 | 0.69 | 0.58 | 0.8834 |
| | 5 | 0.04 | 0.99 | 2249.6 | 1.2e-25 | 0.086 | 0.71 | 0.57 | 0.8721 |

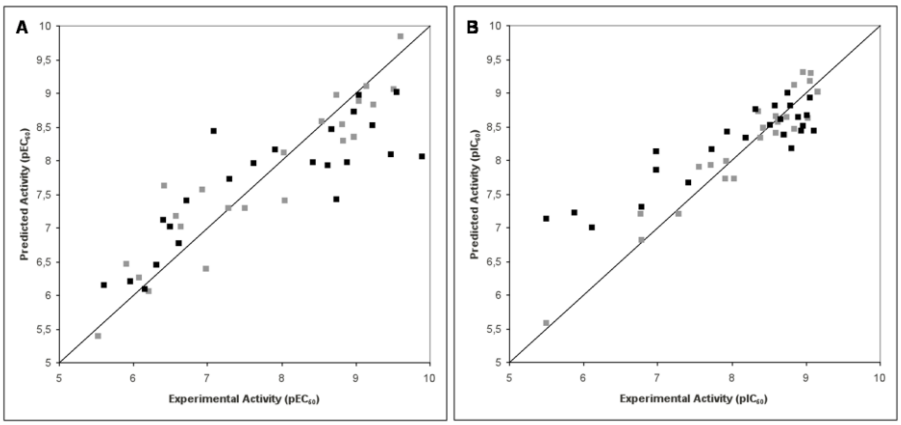



**Figure 5.** Scatter plots of the (A) $pEC_{50}$ and (B) $pIC_{50}$ models applied to the training set (colored in gray) and the test set (colored in black).

Table 2 shows the accuracy of the two models in predicting the transactivation activity and binding affinity of the training and test sets. The predictions of different activities have been classified according to the following residual scale (*i.e.*, residual is computed as the difference between the experimental activity and the estimated activity): residuals lower than 0.8 are considered good predictions; residuals between 0.8 and 1.6 are considered weak predictions and residuals higher than 1.6 are considered poor predictions. The predictions were good for the majority of the training molecules of both models. The predictions for the molecules from the test set are also good for most cases. Only 5 (*i.e.*, the compounds **5_2**, **9**, **20**, **22_2**, **24_2** and **6_2**) and 4 (*i.e.*, the compounds **3**, **4**, **28_2** and **ent-18**) out of 22 compounds have a weak prediction and only one compound for each model (*i.e.*, compound **11_2** for the

transactivation model and compound **5** for the binding model) displayed a poor prediction. All these values indicate a reasonably good correlation between the predicted and experimental activities and validate the use of both models.

**Table 2.** Experimental and predicted values of the transactivation activity ($pEC_{50}$) and binding affinity ($pIC_{50}$) of the 49 tyrosine-based molecules used for the construction of the 3D-QSAR models. Residuals lower than 0.8 are considered good predictions; residuals between 0.8 and 1.6 are considered weak predictions (marked with an asterisk) and residual higher than 1.6 are considered bad predictions (marked with two asterisks).

| Ligands | Sets | Transactivation activity | | | Binding affinity | | |
|---|---|---|---|---|---|---|---|
| | | Experimental Activity | Predicted Activity | Residual | Experimental Activity | Predicted Activity | Residual |
| 2 | training | 6.64 | 7.02 | -0.38 | 7.93 | 7.99 | -0.06 |
| 15 | training | 6.21 | 6.06 | 0.15 | 7.29 | 7.2 | 0.09 |
| 18 | training | 8.04 | 7.41 | 0.63 | 8.85 | 8.47 | 0.38 |
| 30 | training | 6.07 | 6.26 | -0.19 | 6.79 | 6.81 | -0.02 |
| 13_2 | training | 8.98 | 8.35 | 0.63 | 8.72 | 8.38 | 0.34 |
| 14_2 | training | 9.61 | 9.83 | -0.22 | 9.07 | 9.29 | -0.22 |
| 15_2 | training | 8.82 | 8.54 | 0.28 | 9.05 | 9.18 | -0.13 |
| 16_2 | training | 8.74 | 8.97 | -0.23 | 8.85 | 9.12 | -0.27 |
| 18_2 | training | 5.91 | 6.46 | -0.55 | 7.56 | 7.9 | -0.34 |
| 19_2 | training | 5.52 | 5.39 | 0.13 | 7.91 | 7.72 | 0.19 |
| 20_2 | training | 7.51 | 7.29 | 0.22 | 8.59 | 8.41 | 0.18 |
| 21_2 | training | 7.29 | 7.29 | 0 | 6.77 | 7.2 | -0.43 |
| 23_2 | training | 6.93 | 7.56 | -0.63 | 8.36 | 8.72 | -0.36 |
| 36_3 | training | 8.54 | 8.58 | -0.04 | 8.39 | 8.33 | 0.06 |
| 4_2 | training | 10 | 10.33 | -0.33 | 8.96 | 9.31 | -0.35 |
| 49_2 | training | 6.98 | 6.39 | 0.59 | 8.03 | 7.73 | 0.3 |
| 5_2 | training | 6.42 | 7.62 | -1.2 * | 8.59 | 8.65 | -0.06 |
| 58_3 | training | 8.83 | 8.29 | 0.54 | 9.03 | 8.62 | 0.41 |
| 59_3 | training | 9.04 | 8.89 | 0.15 | 8.74 | 8.64 | 0.1 |
| 63_3 | training | 9.15 | 9.11 | 0.04 | 8.43 | 8.48 | -0.05 |
| 65_3 | training | 9.52 | 9.06 | 0.46 | 8.62 | 8.57 | 0.05 |
| 66_3 | training | 9.24 | 8.82 | 0.42 | 9.01 | 8.66 | 0.35 |
| 7_2 | training | 8.03 | 8.11 | -0.08 | 9.16 | 9.02 | 0.14 |
| 70_2 | training | 6.57 | 7.17 | -0.6 | 7.71 | 7.92 | -0.21 |
| ent-2 | training | 4.66 | 4.44 | 0.22 | 5.5 | 5.58 | -0.08 |
| 3 | test | 6.31 | 6.45 | -0.14 | 5.88 | 7.22 | -1.34 * |
| 4 | test | 6.16 | 6.09 | 0.07 | 6.12 | 7 | -0.88 * |
| 6 | test | 5.6 | 6.15 | -0.55 | 5.5 | 7.14 | -1.64 ** |
| 9 | test | 4.78 | 5.94 | -1.16 * | 6.79 | 7.3 | -0.51 |
| 16 | test | 7.3 | 7.73 | -0.43 | 8.19 | 8.33 | -0.14 |
| 20 | test | 9.47 | 8.09 | 1.38 * | 8.94 | 8.43 | 0.51 |
| 10_2 | test | 7.91 | 8.17 | -0.26 | 8.32 | 8.76 | -0.44 |
| 11_2 | test | 9.9 | 8.06 | 1.84 ** | 8.8 | 8.17 | 0.63 |
| 12_2 | test | 9.22 | 8.53 | 0.69 | 8.96 | 8.51 | 0.45 |
| 17_2 | test | 8.68 | 8.46 | 0.22 | 9.06 | 8.92 | 0.14 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 22_2 | test | 8.74 | 7.43 | 1.31 * | 9.11 | 8.44 | 0.67 |
| 24_2 | test | 8.89 | 7.97 | 0.92 * | 8.66 | 8.6 | 0.06 |
| 25_2 | test | 8.62 | 7.93 | 0.69 | 9.01 | 8.66 | 0.35 |
| 26_2 | test | 7.63 | 7.95 | -0.32 | 8.58 | 8.81 | -0.23 |
| 28_2 | test | 6.62 | 6.76 | -0.14 | 6.98 | 7.86 | -0.88 * |
| 29_2 | test | 5.96 | 6.21 | -0.25 | 7.41 | 7.67 | -0.26 |
| 56_3 | test | 9.55 | 9.01 | 0.54 | 8.79 | 8.81 | -0.02 |
| 57_2 | test | 6.49 | 7.01 | -0.52 | 7.73 | 8.16 | -0.43 |
| 6_2 | test | 7.09 | 8.43 | -1.34 * | 8.7 | 8.38 | 0.32 |
| 64_3 | test | 9.04 | 8.97 | 0.07 | 8.52 | 8.52 | 0 |
| 76_2 | test | 6.72 | 7.4 | -0.68 | 7.94 | 8.42 | -0.48 |
| 8_2 | test | 8.97 | 8.73 | 0.24 | 8.75 | 9 | -0.25 |
| 9_2 | test | 8.43 | 7.96 | 0.47 | 8.9 | 8.63 | 0.27 |
| ent-18 | test | 6.41 | 7.11 | -0.7 | 6.98 | 8.13 | -1.15 * |

**Table 3.** Experimental and predicted values of the transactivation activity ($pEC_{50}$) and binding affinity ($pIC_{50}$) of the 6 thiazolidinediones used as an external set to validate the models. Residuals lower than 0.8 are considered good prediction; residuals between 0.8 and 1.6 are considered weak predictions (marked with an asterisk) and residual higher than 1.6 are considered bad predictions (marked with two asterisks).

| Ligands | Transactivation activity | | | Binding affinity | | |
|---|---|---|---|---|---|---|
| | Experimental Activity | Predicted Activity | Residual | Experimental Activity | Predicted Activity | Residual |
| 2_ciglitazone | 4.64 | 6.58 | 1.94 ** | 5.51 | 7.14 | 1.63 ** |
| 2_pioglitazone | 6.23 | 7.42 | 1.19 * | 5.91 | 7.60 | 1.69 ** |
| 2_troglitazone | 6.27 | 6.82 | 0.55 | 6.52 | 7.26 | 0.74 |
| 2_rosiglitazone | 7.05 | 7.48 | 0.43 | 7.33 | 7.86 | 0.53 |
| 2_BRL48482 | 7.95 | 7.92 | -0.03 | 7.57 | 7.87 | 0.30 |
| 2_AD7057 | 8.5 | 8.27 | -0.23 | 8.37 | 7.93 | -0.44 |

For a practical assessment of the study, the predictability of the 3D-QSAR models was evaluated using two external test sets of 6 thiazolidinediones [10] and 68 indanyacetic acid derivates [14]. The results of the predictions of the two external sets are shown in Tables 3 and 4. In general, the predictions are good, although the models over-predict the values of the molecules that have the lowest experimental values. This is most likely due to the fact that the test and training sets do not contain molecules with experimental values in these ranks of activity. Other than this limitation, our models can reasonably predict the order of the activity, *i.e.*, the predicted activity of the molecules with the lowest experimental activities tend to be lower than the predicted activity of molecules with higher experimental values. This point is relevant because when the 3D-QSAR model is applied to the results of a virtual screening, it is more important to know which compounds have the highest activity values rather than knowing the exact activity value for each compound. Therefore, these results are very

encouraging in view of future applications of the study aimed at a) the priorization of analogues of active ligands resulting from virtual screening and b) the optimization of known PPARγ full agonists.

**Table 4.** Experimental and predicted values of the transactivation activity ($pEC_{50}$) of the 68 indanyacetic acid derivates used as an external set to validate the transactivation model. Residuals lower than 0.8 are considered good predictions; residuals between 0.8 and 1.6 are considered weak predictions (marked with an asterisk) and residual higher than 1.6 are considered bad predictions (marked with two asterisks).

| Ligands | Transactivation activity | | | Ligands | Transactivation activity | | |
|---|---|---|---|---|---|---|---|
| | Exp. Activity | Pred. Activity | Residual | | Exp. Activity | Pred. Activity | Residual |
| 10_17a | 5.00 | 6.56 | 1.56 ** | 10_34u | 6.32 | 7.06 | 0.74 |
| 10_17j | 5.00 | 6.51 | 1.51 ** | 10_34b | 6.43 | 6.89 | 0.46 |
| 10_17w | 5.06 | 6.80 | 1.74 ** | 10_34an | 6.44 | 6.96 | 0.52 |
| 10_17i | 5.10 | 6.43 | 1.33 * | 10_34i | 6.49 | 6.80 | 0.31 |
| 10_17b | 5.21 | 6.65 | 1.44 ** | 10_17n | 6.52 | 6.71 | 0.19 |
| 10_17v | 5.25 | 6.98 | 1.73 ** | 10_29g | 6.52 | 6.98 | 0.46 |
| 10_17c | 5.25 | 6.71 | 1.45 ** | 10_34ae | 6.52 | 6.75 | 0.23 |
| 10_29b | 5.39 | 6.65 | 1.26 * | 10_34o | 6.53 | 7.18 | 0.65 |
| 10_17e | 5.52 | 6.50 | 0.98 * | 10_34j | 6.55 | 7.00 | 0.44 |
| 10_17l | 5.66 | 6.56 | 0.90 * | 10_34aa | 6.62 | 7.72 | 1.10 * |
| 10_29i | 5.74 | 6.56 | 0.82 * | 10_34t | 6.67 | 7.10 | 0.43 |
| 10_17q | 5.78 | 6.71 | 0.93 * | 10_34s | 6.74 | 7.09 | 0.35 |
| 10_17r | 5.78 | 6.79 | 1.01 * | 10_34am | 6.75 | 7.24 | 0.49 |
| 10_17m | 5.82 | 6.34 | 0.52 | 10_17x | 6.80 | 6.93 | 0.13 |
| 10_17t | 5.91 | 6.96 | 1.05 * | 10_34n | 6.94 | 7.56 | 0.62 |
| 10_34l | 5.92 | 7.00 | 1.08 * | 10_34m | 7.03 | 7.66 | 0.64 |
| 10_17d | 5.99 | 6.97 | 0.98 * | 10_34q | 7.08 | 7.42 | 0.34 |
| 10_34ab | 6.00 | 6.76 | 0.76 | 10_34af | 7.16 | 7.13 | -0.03 |
| 10_29f | 6.01 | 6.73 | 0.71 | 10_34f | 7.19 | 6.80 | -0.39 |
| 10_34c | 6.02 | 7.37 | 1.35 * | 10_34h | 7.26 | 7.25 | -0.01 |
| 10_17f | 6.03 | 7.11 | 1.08 * | 10_34w | 7.31 | 6.94 | -0.37 |
| 10_34ai | 6.04 | 7.19 | 1.15 * | 10_34a | 7.32 | 6.80 | -0.52 |
| 10_34ac | 6.05 | 7.37 | 1.32 * | 10_34aj | 7.35 | 7.16 | -0.19 |
| 10_34d | 6.06 | 7.37 | 1.31 * | 10_34e | 7.35 | 7.13 | -0.22 |
| 10_34ag | 6.12 | 6.94 | 0.81 * | 10_34ak | 7.36 | 7.74 | 0.39 |
| 10_29c | 6.13 | 7.21 | 1.08 * | 10_34r | 7.38 | 7.24 | -0.14 |
| 10_17s | 6.15 | 6.48 | 0.32 | 10_34ah | 7.39 | 7.13 | -0.26 |
| 10_29a | 6.19 | 6.79 | 0.61 | 10_34al | 7.47 | 7.41 | -0.05 |
| 10_17g | 6.20 | 6.43 | 0.24 | 10_34k | 7.48 | 7.52 | 0.04 |
| 10_17o | 6.24 | 6.54 | 0.30 | 10_34ad | 7.57 | 7.22 | -0.35 |
| 10_17u | 6.26 | 6.67 | 0.41 | 10_34x | 7.70 | 7.35 | -0.35 |
| 10_29d | 6.26 | 6.85 | 0.59 | 10_34g | 7.74 | 7.41 | -0.34 |
| 10_29h | 6.26 | 6.90 | 0.64 | 10_34v | 7.77 | 7.06 | -0.71 * |
| 10_29e | 6.31 | 6.90 | 0.58 | 10_34p | 7.92 | 7.26 | -0.66 |

Figures 6 shows the representation of the 3D-QSAR models. In these types of figures, the cubes that represent the model are displayed and colored according to the sign of their coefficient values. Blue and red cubes are usedfor positive and negative coefficients, respectively, and indicate regions that increase or decrease the analyzed parameter. One of the advantages of using these representations is that the position of the cubes in the 3D-QSAR model can be compared with the positions of the amino acid residues in the active site. This might provide insight into which functional groups are desirable or undesirable at certain positions of a molecule. Figure 6A shows the favorable and unfavorable regions for the transactivation activity. The favorable region for transactivation is mainly located in regions that interact with arm II of the LBD of PPARγ. The contribution of this region is mainly hydrophobic with some contribution of electron-withdraw (results not shown). This region corresponds to the effector part of PPARγ full agonists and is the most variable region in the molecules analyzed (see Figure 1). Although this region is far away from Tyr473 and H12, it explains the differences in affinity and potency between very similar glitazones; *i.e.*, the larger the effector module, more hydrophobic interactions will occur with arm II of the LBD of PPARγ, thus stabilizing better the PPARγ-ligand complex. Figures 6C and 6E display the cubes of the 3D-QSAR model grid that are occupied by two of the compounds analyzed. These representations show which parts of the ligand have a positive or negative contribution to the parameter analyzed, which is, in this case, the transactivation activity. Figure 6C shows the **4_2** compound, which has one of the highest transactivation activities in this series ($pEC_{50}$=10). Our 3D-QSAR model explains the high transactivation activity of this compound because it provides a heterocyclic and a phenyl ring at the effector module. In comparison, the compound from our ligand dataset with the lowest transactivation activity (*i.e.*, **ent-2** with $pEC_{50}$=4.66 in Figure 6E) only interact with arm II through one phenyl ring. Moreover, this compound has a (R)-configuration, and it is known that compounds with an (S)-configuration, derived from naturally occurring L-tyrosine, are more active as PPARγ full agonists [10].

Figures 6B, 6D and 6F show a representation of the binding model. In this model, the binding affinity of PPARγ is the analyzed variable. Figure 6B shows the favorable and unfavorable regions for binding. The favorable region is again located at arm II, and the main contribution to the binding affinity of PPARγ is also caused by hydrophobic interactions (results not shown). The unfavorable regions are located at arm I and at the beginning of arm II. Regarding the first unfavorable region, it occupies the place of one internal hydrogen bond in the 2-aminobenzophenone moiety. Therefore, the presence of some hydrophobic residues at this position will not allow the formation of this internal interaction. The second unfavorable region is due to steric clashes with residues Arg288, Leu330 and Ile341. Figures 6D and 6F show the 3D-QSAR models represented only by the cubic volume elements that are occupied by one of the compounds with the highest (*i.e.*, the **7_2** compound with a $pIC_{50}$ of 9.16)

and lowest (*i.e.*, the **ent-2** compound with a $pIC_{50}$ of 5.5) binding affinities, respectively. The **7_2** compound fits the blue areas of the 3D-QSAR model perfectly, as it can interact with the receptor through the favorable regions at arm II (see Figure 6D). However, the **ent-2** compound does not occupy all of arm II (see Figure 6F), and it is not able to make the favorable interactions shown as blue cubes in Figure 6.

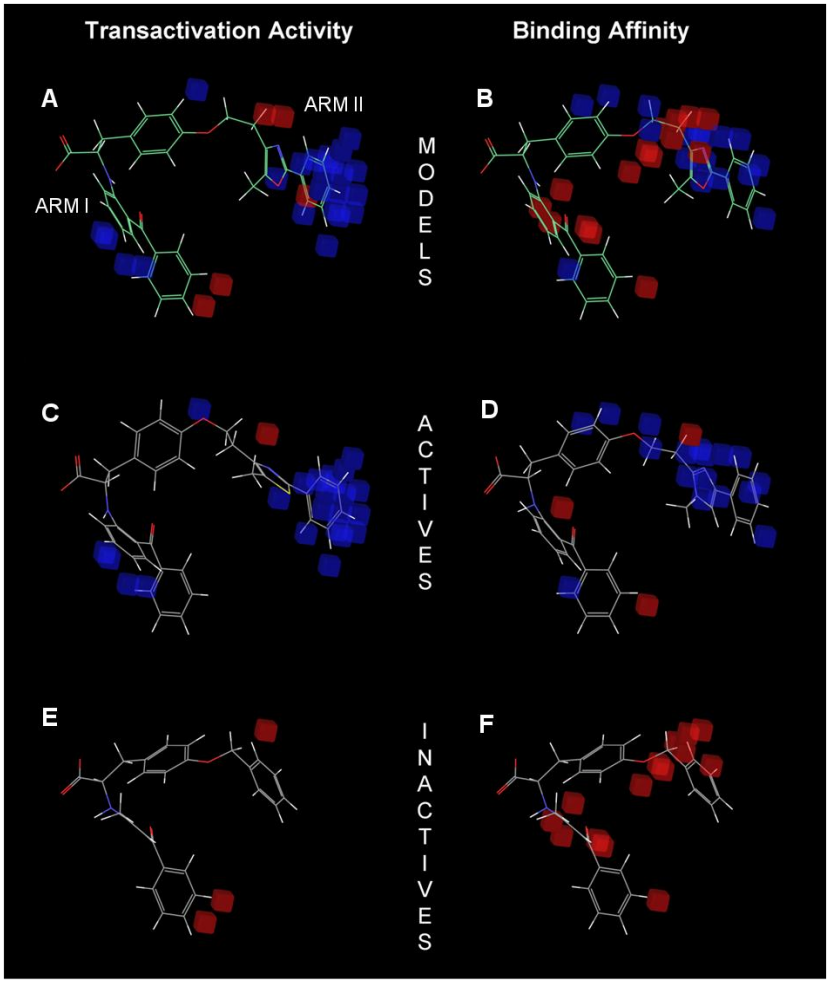



**Figure 6.** Representation of the $pEC_{50}$ (A) and the $pIC_{50}$ (B) models. Blue and red cubes indicate, respectively, regions that are favorable and unfavorable for the transactivation activity or binding affinity to the LBD of PPARγ. Panels (C) and (E) show the cubic volume elements that are occupied by one of the compounds of the series with the highest (*i.e.*, **4_2**) and the lowest (*i.e.*, **ent-2**) transactivation activity, respectively. Panels (D) and (F) show the cubic volume elements that are occupied by one of the compounds of the series with the highest (*i.e.*, **7_2**) and the lowest (*i.e.*, **ent-2**) binding affinity, respectively. All panels are presented in the same relative orientation in order to allow for an easier comparison.

## Conclusion

In this work, a structure-based docking strategy was used for aligning the molecules, and highly predictive 3D-QSAR transactivation and binding models were developed for PPARγ full agonists. These models match well with the known features of the different parts of the PPARγ binding site and show that the binding portion of the PPARγ agonists is essential for the transactivation activity of PPARγ. The hydrophobic interactions between the effector portion and the receptor are also important for increasing the transactivation activity of PPARγ.

## References

1. Yessoufou A, Wahli W. (2010) Multifaceted roles of peroxisome proliferator-activated receptors (PPARs) at the cellular and whole organism levels. Swiss Medical Weekly 140: w13071.

2. Garcia-Vallve S, Palau J. (1998) Nuclear receptors, nuclear-receptor factors, and nuclear-receptor-like orphans form a large paralog cluster in homo sapiens. Mol Biol Evol 15(6): 665-682.

3. Berger J, Moller DE. (2002) The mechanisms of action of PPARs. Annu Rev Med 53: 409-435.

4. Willson TM, Lambert MH, Kliewer SA. (2001) Peroxisome proliferator-activated receptor gamma and metabolic disease. Annu Rev Biochem 70: 341-367.

5. Lehmann JM, Moore LB, Smith-Oliver T A, Wilkison WO, Willson TM, et al. (1995) An antidiabetic thiazolidinedione is a high affinity ligand for peroxisome proliferator-activated receptor gamma (PPAR gamma). The Journal of Biological Chemistry 270(22): 12953-12956.

6. Michalik L, Wahli W. (2007) Peroxisome proliferator-activated receptors (PPARs) in skin health, repair and disease. Biochimica Et Biophysica Acta {(BBA)} - Molecular and Cell Biology of Lipids 1771(8): 991-998.

7. Sertznig P, Seifert M, Tilgen W, Reichrath J. (2008) Peroxisome proliferator-activated receptors (PPARs) and the human skin: Importance of PPARs in skin physiology and dermatologic diseases. American Journal of Clinical Dermatology 9(1): 15-31.

8. Behshad R, Cooper KD, Korman NJ. (2008) A retrospective case series review of the peroxisome proliferator-activated receptor ligand rosiglitazone in the treatment of atopic dermatitis. Arch Dermatol 144(1): 84-88.

9. Schadendorf D. (2009) Peroxisome proliferator-activating receptors: A new way to treat melanoma? J Invest Dermatol 129(5): 1061-1063.

10. Henke BR, Blanchard SG, Brackeen MF, Brown KK, Cobb JE, et al. (1998) N-(2-benzoylphenyl)-L-tyrosine PPARgamma agonists. 1. discovery of a novel series of potent antihyperglycemic and antihyperlipidemic agents. J Med Chem 41(25): 5020-5036.

11. Collins JL, Blanchard SG, Boswell GE, Charifson PS, Cobb JE, et al. (1998) N-(2-benzoylphenyl)-L-tyrosine PPARgamma agonists. 2. structure-activity relationship and optimization of the phenyl alkyl ether moiety. J Med Chem 41(25): 5037-5054.

12. Cobb JE, Blanchard SG, Boswell EG, Brown KK, Charifson PS, et al. (1998) N-(2-benzoylphenyl)-L-tyrosine PPARgamma agonists. 3. structure-activity relationship and optimization of the N-aryl substituent. J Med Chem 41(25): 5055-5069.

13. Azukizawa S, Kasai M, Takahashi K, Miike T, Kunishiro K, et al. (2008) Synthesis and biological evaluation of (S)-1,2,3,4-tetrahydroisoquinoline-3-carboxylic acids: A novel series of PPAR gamma agonists. Chem Pharm Bull (Tokyo) 56(3): 335-345.

14. Rudolph J, Chen L, Majumdar D, Bullock WH, Burns M, et al. (2007) Indanylacetic acid derivatives carrying 4-thiazolyl-phenoxy tail groups, a new class of potent PPAR alpha/gamma/delta pan agonists: Synthesis, structure-activity relationship, and in vivo efficacy. J Med Chem 50(5): 984-1000.

15. Gampe RT, Montana VG, Lambert MH, Miller AB, Bledsoe RK, et al. (2000) Asymmetry in the PPARgamma/RXRalpha crystal structure reveals the molecular basis of heterodimerization among nuclear receptors. Mol Cell 5(3): 545-555.

16. Zoete V, Grosdidier A, Michielin O. (2007) Peroxisome proliferator-activated receptor structures: Ligand specificity, molecular switch and interactions with regulators. Biochim Biophys Acta 1771(8): 915-925.

17. Guasch L, Sala E, Valls C, Blay M, Mulero M, et al. (2011) Structural insights for the design of new PPARgamma partial agonists with high binding affinity and low transactivation activity. J Comput Aided Mol Des .

18. Giaginis C, Theocharis S, Tsantili-Kakoulidou Anna. (2009) A QSAR study on indole-based PPAR gamma agonists in respect to receptor binding and gene transactivation data. QSAR & Combinatorial Science 28(8): 802-805.

19. Rathi L, Kashaw SK, Dixit A, Pandey G, Saxena AK. (2004) Pharmacophore identification and 3D-QSAR studies in N-(2-benzoyl phenyl)-L-tyrosines as PPAR gamma agonists. Bioorg Med Chem 12(1): 63-69.

20. Liao C, Xie A, Shi L, Zhou J, Lu X. (2004) Eigenvalue analysis of peroxisome proliferator-activated receptor gamma agonists. J Chem Inf Comput Sci 44(1): 230-238.

21. Hyun KH, Lee DY, Lee B, Kim CK. (2004) Receptor-based 3D QSAR studies on PPAR-gamma agonists using CoMFA and CoMSIA approaches. QSAR & Combinatorial Science 23(8): 637-649.

22. Kurogi Y. (1999) Three-dimensional quantitative structure-activity relationships (3D-QSAR) of antidiabetic thiazolidinediones. Drug Des Discov 16(2): 109-118.

23. Kapoor M, McCann M, Liu S, Huh K, Denton CP, et al. (2009) Loss of peroxisome proliferator-activated receptor gamma in mouse fibroblasts results in increased susceptibility to bleomycin-induced skin fibrosis. Arthritis Rheum 60(9): 2822-2829.

24. Dixon SL, Smondyrev AM, Knoll EH, Rao SN, Shaw DE, et al. (2006) PHASE: A new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. methodology and preliminary results. J Comput Aided Mol Des 20(10-11): 647-671.

## Supporting Information

**Table S1.** Smiles of the 49 tyrosine-based compounds used to generate the 3D-QSAR models

| Title | Smile |
|---|---|
| 2 | c1ccccc1C(=O)\C=C(\C)N[C@H](C([O-])=O)Cc2ccc(cc2)OCc3ccccc3 |
| 3 | c1ccccc1C(=O)\C=C(\C)N[C@H](C(=O)N)Cc2ccc(cc2)OCc3ccccc3 |
| 4 | c1ccccc1C(=O)\C=C(\C)N[C@H](C(=O)OC)Cc2ccc(cc2)OCc3ccccc3 |
| 4_2 | c1ccccc1C(=O)c2c(cccc2)N[C@H](C([O-])=O)Cc3ccc(cc3)OCCc4c(C)sc(n4)-c5ccccc5 |
| 5_2 | c1ccccc1C(=O)c2c(cccc2)N[C@H](C([O-])=O)Cc3ccc(cc3)OCCc4c(C)nc([nH]4)-c5ccccc5 |
| 6 | [O-]C(=O)[C@@H]([NH3+])Cc1ccc(cc1)OCc2ccccc2 |
| 6_2 | c1ccccc1C(=O)c2c(cccc2)N[C@H](C([O-])=O)Cc3ccc(cc3)OCCc4c(C)nc(n4COC)-c5ccccc5 |
| 7_2 | c1ccccc1C(=O)c2c(cccc2)N[C@H](C([O-])=O)Cc3ccc(cc3)OCCc(n4)n(C)cc4-c5ccccc5 |
| 8_2 | c1ccccc1C(=O)c2c(cccc2)N[C@H](C([O-])=O)Cc3ccc(cc3)OCCn4nc(cc4C)-c5ccccc5 |
| 9 | c1ccccc1C(=O)c2c(cccc2)N[C@H](C([O-])=O)Cc3ccc(cc3)OCc4ccccc4 |
| 9_2 | c1ccccc1C(=O)c2c(cccc2)N[C@H](C([O-])=O)Cc3ccc(cc3)OCCc4cc(C)n(n4)-c5ccccc5 |
| 10_2 | c1ccccc1C(=O)c2c(cccc2)N[C@H](C([O-])=O)Cc3ccc(cc3)OCCn4nc(nc4C)-c5ccccc5 |
| 11_2 | c1ccccc1C(=O)c2c(cccc2)N[C@H](C([O-])=O)Cc3ccc(cc3)OCCc4c(C)oc(n4)-c5ccc(F)cc5 |
| 12_2 | c1ccccc1C(=O)c2c(cccc2)N[C@H](C([O-])=O)Cc3ccc(cc3)OCCc4c(C)oc(n4)-c5ccc(cc5)OC |
| 13_2 | c1ccccc1C(=O)c2c(cccc2)N[C@H](C([O-])=O)Cc3ccc(cc3)OCCc4c(C)oc(n4)-c(c5C)scc5 |
| 14_2 | c1ccccc1C(=O)c2c(cccc2)N[C@H](C([O-])=O)Cc3ccc(cc3)OCCc4c(C)oc(n4)-c(s5)ccc5C |
| 15 | c1ccccc1C(=O)c2c(cccc2)N[C@H](C([O-])=O)Cc3ccc(cc3)OCC4(C)CCCCC4 |
| 15_2 | c1ccccc1C(=O)c2c(cccc2)N[C@H](C([O-])=O)Cc3ccc(cc3)OCCc4c(C)sc(n4)-c(no5)cc5C |
| 16 | c1ccccc1C(=O)c2c(cccc2)N[C@H](C([O-])=O)Cc3ccc(cc3)OCc(nc4)ccc4CC |
| 16_2 | c1ccccc1C(=O)c2c(cccc2)N[C@H](C([O-])=O)Cc3ccc(cc3)OCCc4c(C)oc(n4)-c5ccncc5 |
| 17_2 | c1ccccc1C(=O)c2c(cccc2)N[C@H](C([O-])=O)Cc3ccc(cc3)OCCc4c(C)sc(n4)-c5ccncc5 |
| 18 | c1ccccc1C(=O)c2c(cccc2)N[C@H](C([O-])=O)Cc3ccc(cc3)OCCN(C)c4cccc[nH+]4 |
| 18_2 | c1ccccc1C(=O)c2c(cccc2)N[C@H](C([O-])=O)Cc3ccc(cc3)OCCc4c(C)sc(n4)N(C)C |
| 19_2 | c1ccccc1C(=O)c2c(cccc2)N[C@H](C([O-])=O)Cc3ccc(cc3)OCCc4c(C)sc(n4)NCCC[NH+](C)C |
| 20 | c1ccccc1C(=O)c2c(cccc2)N[C@H](C([O-])=O)Cc3ccc(cc3)OCCc4c(C)oc(n4)-c5ccccc5 |
| 20_2 | c1ccccc1C(=O)c2c(cccc2)N[C@H](C([O-])=O)Cc3ccc(cc3)OCCc4c(C)sc(n4)NCCOC |
| 21_2 | c1ccccc1C(=O)c2c(cccc2)N[C@H](C([O-])=O)Cc3ccc(cc3)OCCc4c(C)oc(n4)N5CCCCC5 |
| 22_2 | c1ccccc1C(=O)c2c(cccc2)N[C@H](C([O-])=O)Cc3ccc(cc3)OCCc4c(C)sc(n4)N5CCOCC5 |
| 23_2 | c1ccccc1C(=O)c2c(cccc2)N[C@H](C([O-])=O)Cc3ccc(cc3)OCCc4c(C)sc(n4)N5CC[NH2+]CC5 |
| 24_2 | c1ccccc1C(=O)c2c(cccc2)N[C@H](C([O-])=O)Cc3ccc(cc3)OCCc4c(C)sc(n4)N5CC[NH+](C)CC5 |
| 25_2 | c1ccccc1C(=O)c2c(cccc2)N[C@H](C([O-])=O)Cc3ccc(cc3)OCCc4c(C)sc(n4)N(CC5)CCN5C(=O)OC(C)(C)C |
| 26_2 | c1ccccc1C(=O)c2c(cccc2)N[C@H](C([O-])=O)Cc3ccc(cc3)OCCc4c(C)sc(n4)N5CCN(CC5)[S@@](=O)OC |
| 28_2 | c1ccccc1C(=O)c2c(cccc2)N[C@H](C([O-])=O)Cc3ccc(cc3)OCc(cc4)ccc4C(C)C |
| 29_2 | c1ccccc1C(=O)c2c(cccc2)N[C@H](C([O-])=O)Cc3ccc(cc3)OCc4ccc(Cl)cc4 |
| 30 | c1ccccc1C(=N/N)\c2c(cccc2)N[C@H](C([O-])=O)Cc3ccc(cc3)OCCN(C)c4cccnc4 |
| 36_3 | C1CCCCC1C(=O)c2c(cccc2)N[C@H](C([O-])=O)Cc3ccc(cc3)OCCN(C)c(n4)oc(c45)cccc5 |

| | |
|---|---|
| 49_2 | c1ccccc1C(=O)c2c(cccc2)N[C@H](C([O-])=O)Cc3ccc(cc3)OCCc4ccc(Cl)cc4 |
| 56_3 | C1CCCCC1C(=O)c2c(cccc2)N[C@H](C([O-])=O)Cc3ccc(cc3)OCCc4c(C)oc(n4)-c5ccccc5 |
| 57_2 | c1ccccc1C(=O)c2c(cccc2)N[C@H](C([O-])=O)Cc3ccc(cc3)OCCc4c(C)ncs4 |
| 58_3 | c1ncccc1C(=O)c2c(cccc2)N[C@H](C([O-])=O)Cc3ccc(cc3)OCCc4c(C)oc(n4)-c5ccccc5 |
| 59_3 | c1cnccc1C(=O)c2c(cccc2)N[C@H](C([O-])=O)Cc3ccc(cc3)OCCc4c(C)oc(n4)-c5ccccc5 |
| 63_3 | COC(=O)c1c(cccc1)N[C@H](C([O-])=O)Cc2ccc(cc2)OCCc3c(C)oc(n3)-c4ccccc4 |
| 64_3 | CCOC(=O)c1c(cccc1)N[C@H](C([O-])=O)Cc2ccc(cc2)OCCc3c(C)oc(n3)-c4ccccc4 |
| 65_3 | CCCOC(=O)c1c(cccc1)N[C@H](C([O-])=O)Cc2ccc(cc2)OCCc3c(C)oc(n3)-c4ccccc4 |
| 66_3 | CC(C)OC(=O)c1c(cccc1)N[C@H](C([O-])=O)Cc2ccc(cc2)OCCc3c(C)oc(n3)-c4ccccc4 |
| 70_2 | c1ccccc1C(=O)c2c(cccc2)N[C@H](C([O-])=O)Cc3ccc(cc3)OCCOc4ccc(Br)cc4 |
| 76_2 | c1ccccc1C(=O)c2c(cccc2)N[C@H](C([O-])=O)Cc3ccc(cc3)OCCSc4ccc(Cl)cc4 |
| ent-2 | c1ccccc1C(=O)\C=C(\C)N[C@@H](C([O-])=O)Cc2ccc(cc2)OCc3ccccc3 |
| ent-18 | c1ccccc1C(=O)c2c(cccc2)N[C@@H](C([O-])=O)Cc3ccc(cc3)OCCN(C)c4cccc[nH+]4 |

**Table S2.** Smiles of the 6 thiazolidinediones used as an external validation set.

| Title | Smile |
|---|---|
| 2_AD7057 | c1ccccc1-c(n2)oc(C)c2CCOc(cc3)ccc3C[C@@H](C4=O)SC(=O)N4 |
| 2_BRL48482 | c1cccc(c12)oc(n2)C[NH2+]CCOc(cc3)ccc3C[C@@H](C4=O)SC(=O)N4 |
| 2_ciglitazone | N1C(=O)S[C@H](C1=O)Cc2ccc(cc2)OCC3(C)CCCCC3 |
| 2_pioglitazone | CCc1ccc(nc1)CCOc(cc2)ccc2C[C@@H](C3=O)SC(=O)N3 |
| 2_rosiglitazone | c1cc[nH+]c(c1C)NCCOc(cc2)ccc2C[C@@H](C3=O)SC(=O)N3 |
| 2_troglitazone | N1C(=O)S[C@H](C1=O)Cc2ccc(cc2)OC[C@@](C)(CC3)Oc(c34)c(C)c(C)c(O)c4C |

**Table S3.** Smiles of the 68 indanyacetic acid derivates used as an external validation set.

| Title | Smile |
|---|---|
| 10_17a | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc3ccccc3 |
| 10_17b | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(cc3)ccc3CC |
| 10_17c | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(cc3)ccc3C(F)(F)F |
| 10_17d | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(cc3)ccc3OC(F)(F)F |
| 10_17e | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc3ccc(cc3)OC |
| 10_17f | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc3ccc(cc3)OCC |
| 10_17g | N#Cc1ccc(cc1)OCCCOc(cc2)cc(c23)CC[C@H]3CC([O-])=O |
| 10_17i | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(cc3C)ccc3 |
| 10_17j | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(c3)ccc(C)c3C |
| 10_17l | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(c3C)ccc(C)c3 |
| 10_17m | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(c3CCC)cccc3 |
| 10_17n | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(cc3)c(CCC)cc3C(F)(F)F |

| | |
|---|---|
| 10_17o | N#Cc1cc(CCC)c(cc1)OCCCOc(cc2)cc(c23)CC[C@H]3CC([O-])=O |
| 10_17q | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(cc3)c(OC)cc3C |
| 10_17r | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(cc3)c(OC)cc3CC |
| 10_17s | N#Cc1cc(OC)c(cc1)OCCCOc(cc2)cc(c23)CC[C@H]3CC([O-])=O |
| 10_17t | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(cc3)c(cc3C)OCC |
| 10_17u | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(cc3)ccc3-n4cncn4 |
| 10_17v | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(c3NC(=O)C)ccc(c3)-n4ccnn4 |
| 10_17w | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc3c(Cl)cc(cc3)-n4cnnc4 |
| 10_17w | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc3c(Cl)cc(cc3)-n4cnnc4 |
| 10_17x | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(c3C)ccc(c3)-c(sn4)nc4C(F)(F)F |
| 10_29a | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(cc3)ccc3-c4ccsc4 |
| 10_29b | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(cc3)ccc3-c4ccoc4 |
| 10_29c | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(cc3)ccc3-c(cc4)cc(c45)[nH]cc5 |
| 10_29d | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(cc3)ccc3-c4cccnc4 |
| 10_29e | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(cc3)ccc3-c4c(OC)ccnc4 |
| 10_29f | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(cc3)ccc3-c4cncnc4 |
| 10_29g | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(cc3)ccc3-c4c(OC)nc(nc4)OC |
| 10_29h | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(cc3)ccc3-c(n4)cccc4C |
| 10_29i | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(cc3)ccc3-c(nc4)ccc4C(F)(F)F |
| 10_34a | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(c3CCC)ccc(c3)-c4nccs4 |
| 10_34aa | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(c3CCC)ccc(c3)-c(s4)nc(C)c4C([O-])=O |
| 10_34ab | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(cc3)c(OC)cc3-c(s4)nc(C)c4C([O-])=O |
| 10_34ac | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(c3CCC)ccc(c3)-c(s4)nc(CO)c4C([O-])=O |
| 10_34ad | [O-]C(=O)Cc1csc(n1)-c(c2)ccc(c2CCC)OCCCOc(cc3)cc(c34)CC[C@H]4CC([O-])=O |
| 10_34ae | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(cc3)ccc3-c(n4)scc4OC |
| 10_34af | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(cc3)c(OC)cc3-c(n4)scc4OC |
| 10_34ag | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(cc3)ccc3-c(n4)scc4OCC |
| 10_34ah | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(c3CCC)ccc(c3)-c(n4)scc4OCC |
| 10_34ai | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(cc3)c(OC)cc3-c(n4)scc4OCC |
| 10_34aj | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(c3CCC)ccc(c3)-c(n4)scc4OC(C)C |
| 10_34ak | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(cc3)c(OC)cc3-c(n4)scc4OC(C)C |
| 10_34al | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(c3CCC)ccc(c3)-c(s4)nc(c4C)OCC |
| 10_34am | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(cc3)c(OC)cc3-c(s4)nc(c4C)OCC |
| 10_34an | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(cc3)c(OC)cc3-c(s4)nc(c4CC)OCC |
| 10_34b | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(cc3)c(OC)cc3-c4nccs4 |
| 10_34c | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(cc3)c(OC)cc3-c(n4)scc4C |
| 10_34d | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(cc3)ccc3-c(n4)scc4CC |
| 10_34e | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(c3CCC)ccc(c3)-c(n4)scc4CC |
| 10_34f | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(cc3)c(OC)cc3-c(n4)scc4CC |
| 10_34g | CC(C)(C)c1csc(n1)-c(c2)ccc(c2CCC)OCCCOc(cc3)cc(c34)CC[C@@H]4C([O-])([O-])C |
| 10_34h | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(c3CCC)ccc(c3)-c(n4)scc4C(F)(F)F |
| 10_34i | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(cc3)c(OC)cc3-c(n4)scc4C(F)(F)F |
| 10_34j | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(cc3)ccc3-c(n4)sc(C)c4C |

| | |
|---|---|
| 10_34k | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(cc3)c(OC)cc3-c(n4)sc(C)c4C |
| 10_34l | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(cc3)ccc3-c(n4)sc(c45)CCC5 |
| 10_34m | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(c3CCC)ccc(c3)-c(n4)sc(c45)CCC5 |
| 10_34n | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(cc3)c(OC)cc3-c(n4)sc(c45)CCC5 |
| 10_34o | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(cc3)ccc3-c(n4)sc(c45)CCCC5 |
| 10_34p | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(c3CCC)ccc(c3)-c(n4)sc(c45)CCCC5 |
| 10_34q | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(cc3)c(OC)cc3-c(n4)sc(c45)CCCC5 |
| 10_34r | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(c3CCC)ccc(c3)-c(s4)nc(c45)OCCC5 |
| 10_34s | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(cc3)c(OC)cc3-c(s4)nc(c45)OCCC5 |
| 10_34t | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(cc3)c(OC)cc3-c(n4)sc(c45)cccc5 |
| 10_34u | CC(=O)c1c(C)nc(s1)-c2ccc(cc2)OCCCOc(cc3)cc(c34)CC[C@H]4CC([O-])=O |
| 10_34v | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(c3CCC)ccc(c3)-c(s4)nc(C)c4C(=O)C |
| 10_34w | CC(=O)c1c(C)nc(s1)-c2cc(OC)c(cc2)OCCCOc(cc3)cc(c34)CC[C@H]4CC([O-])=O |
| 10_34x | [O-]C(=O)C[C@@H]1CCc(c12)cc(cc2)OCCCOc(c3CCC)ccc(c3)-c(s4)nc(C)c4C(=O)N(C)C |

CHAPTER 2

# STRUCTURAL INSIGHTS FOR THE DESIGN OF NEW PPARγ PARTIAL AGONISTS WITH HIGH BINDING AFFINITY AND LOW TRANSACTIVATION ACTIVITY

***Guasch L**, Sala E, Valls C, Blay M, Mulero M, Arola L, Pujadas G, Garcia-Vallvé S. J Comput Aided Mol Des DOI 10.1007/s10822-011-9446-9*

## ABSTRACT

PPARγ full agonists are molecules with powerful insulin-sensitizing action that are used as antidiabetic drugs. Unfortunately, these compounds also present various side effects. Recent results suggest that effective PPARγ agonists should show a low transactivation activity but a high binding affinity to inhibit phosphorylation at Ser273. We use several structure activity relationship studies of synthetic PPARγ agonists to explore the different binding features of full and partial PPARγ agonists with the aim of differentiating the features needed for binding and those needed for the transactivation activity of PPARγ. Our results suggest that effective partial agonists should have a hydrophobic moiety and an acceptor site with an appropriate conformation to interact with arm II and establish a hydrogen bond with Ser342 or an equivalent residue at arm III. Despite the fact that interactions with arm I increase the binding affinity, this region should be avoided in order to not increase the transactivation activity of potential PPARγ partial agonists.

## Introduction

Peroxisome Proliferator-Activated Receptor γ (PPARγ) is a ligand-activated transcription factor and a member of the nuclear receptor superfamily that plays an important role in adipogenesis and glucose homeostasis [1]. PPARγ is activated by polyunsaturated fatty acids and their metabolites. This transcription factor regulates the expression of adipocyte-specific genes [2]; its function is, therefore, essential to fat cell formation, and PPARγ full agonists stimulate triglyceride storage and the differentiation of preadipocytes into adipocytes [1]. Some PPARγ full agonists, such as thiazolidinediones (TZDs), also have a powerful insulin-sensitizing action and are used as antidiabetic drugs [3]. Unfortunately, TZDs present various side effects, including weight gain, increased adipogenesis, renal fluid retention, bone fracture and increased incidence of cardiovascular events [4-6]. Other compounds with poor agonist activities for PPARγ, called PPARγ modulators or PPARγ partial agonists, retain very good antidiabetic effects without these undesired side effects [4]. Therefore, several partial agonists of PPARγ are being developed as new-antidiabetic drugs [6-8]. Analyses of a large number of crystallographic structures of the PPARγ ligand-binding domain (LBD) bound to an agonist have revealed that PPARγ has at least two binding modes in a single binding site. These two binding modes correspond to full and partial agonists [9]. The binding pocket of PPARγ has a Y-shaped form, consisting of an entrance (arm III) that branches off into two pockets. Arm I is extended toward H12, and arm II is situated between helix H3 and a β-sheet [10]. Arm I is the only substantially polar cavity of the PPARγ ligand-binding domain, whereas arms II and III are mainly hydrophobic [10]. Full agonists occupy arm I, making a net of hydrogen bonds with the side chains of Ser289, His323, His449 and Tyr473 [9, 11]. These interactions stabilize H12 and are responsible for the transactivation activity of PPARγ [9, 11]. In addition, full agonists also occupy arm II through a hydrophobic tail that is present in all ligands of this class [9, 11]. However, partial agonists interact mainly with arm III through a hydrogen bond with Ser342, but also with arm II through several hydrophobic interactions [12, 13]. This binding mode causes a lesser degree of H12 stabilization and an increase in the stabilization of H3 that affects the recruitment of coactivators and decreases the transactivation activity of PPARγ [7, 14].

However, the previous model does not explain why compounds with different PPARγ transactivation activities show the same insulin-sensitizing power. Recently, Choi and coworkers [15] revealed a new mechanism of action for the antidiabetic effect of some PPARγ agonists. This mechanism is completely independent of the classical PPARγ transactivation activity and relies instead on inhibition of the phosphorylation of PPARγ at Ser273, thereby preventing the unregulated expression of some genes, including adipsin (a fat-cell-selective gene, the expression of which is

altered in obesity) and adiponectin (an insulin-sensitizing adipokine) [15]. This alternative mechanism could clarify a long-standing paradox of why PPARγ activation by a wide range of ligands does not always correlate with the ligands' in vivo efficacy [16]. With this new knowledge, many research groups have had to accordingly shift their focus from their past drug discovery efforts on PPARγ, which were focused exclusively on potency and agonist activity. It is now necessary to develop effective and safe antidiabetic therapies that maximize the inhibition of PPARγ phosphorylation at Ser273 and reduce the side effects observed with current PPARγ drugs [8, 15]. It seems likely that at least some of the problematic side effects of PPARγ full agonists, such as weight gain or fluid retention, may occur through classical agonist action and that a substantial portion of the therapeutic benefits of full and partial PPARγ agonists occurs through the inhibition of the PPARγ phosphorylation at Ser273 [15]. Thus, an effective partial agonist of PPARγ would have a weak transactivation activity and high phosphorylation inhibitory activity on PPARγ at Ser273. This kind of compound would maintain its antidiabetic effects while reducing undesired side effects. Until researchers shift their focus to study the potency of the phosphorylation inhibitory activity at Ser273, binding affinity would be used instead to evaluate potential drug candidates. In this sense, the antidiabetic potency of PPARγ ligand drugs correlates very well with their binding affinities [17]. In the present study, after reviewing the binding features of full and partial agonists, we use several structure activity relationship (SAR) studies of synthetic PPARγ agonists to explore the different binding features of full and partial PPARγ agonists. Our goal was to differentiate the features needed for binding from those needed for the transactivation activity of PPARγ. Thus, our rationale consists of defining which interactions between the ligand-binding domain of PPARγ and its ligands increases the binding affinity without increasing the PPARγ transactivation activity. This information would allow us to predict the features that will produce optimal PPARγ agonists for use as antidiabetic drugs.

## Computational Methods

**Datasets.** A dataset of 205 PPARγ agonists with measured $IC_{50}$ values (i.e., binding affinity measured by the displacement of a radiolabeled full agonist) and transactivation activity was assembled from several SAR studies [18-29] (see Table 1). The $IC_{50}$ (nM) values were then transformed to -log $IC_{50}$ ($pIC_{50}$) (see Supporting Information Table S1). The transactivation activities were expressed as the percentage of maximal activation relative to the full agonist rosiglitazone (% max activation) (see Supporting Information Table S2). All compounds were drawn with ChemDraw Ultra v11.0 (CambridgeSoft Corporation, Cambridge, MA, USA; http://www.cambridgesoft.com), and their 3D structures were minimized with the LigPrep v2.4 program (Schrödinger LLC., Portland, USA) using an OPLS_2005 force field at pH 7.0 with the rest of the parameters at default.

**Table 1**. Structure activity relationship (SAR) studies of the PPARγ agonists used in the current study.

| series | cluster | transactivation activity (% max. activation)[a] | | binding affinity $pIC_{50}$ | | ref |
|---|---|---|---|---|---|---|
| | | Nº. ligands | activity range | Nº. ligands | activity range | |
| *sar1* | 1 | 18 | 4 - 33 | 30 | 6.50 – 8.70 | [18] |
| *sar2* | 2 | 11 | 20 - 51 | 13 | 7.77 – 9.00 | [19] |
| *sar3* | 2 | 13 | 21 – 97 | 17 | 5.20 – 9.00 | [20] |
| *sar4* | 2 | 8 | 19 - 33 | 11 | 6.96 – 9.00 | [21] |
| *sar5* | 2 | 18 | 14 - 47 | 19 | 6.14 – 9.00 | [22] |
| *sar6* | 3 | - | - | 16 | 5.44 – 7.00 | [23] |
| *sar7* | 1 | - | - | 30 | 5.03 – 9.00 | [24] |
| *sar8* | 4 | 12 | 26 - 65 | 11 | 4.91 – 8.10 | [25] |
| *sar9* | 4 | 9 | 24 - 71 | - | - | [26] |
| *sar10* | 4 | 10 | 30 - 92 | - | - | [27] |
| *sar11* | 3 | 20 | 25 - 89 | - | - | [28] |
| *sar12* | 5 | 17 | 19 - 93 | - | - | [29] |

[a] % of maximal activation relative to the full agonist rosiglitazone.

**Clustering.** A structural similarity analysis of all PPARγ agonists was performed using the Canvas v1.2 program (Schrödinger LLC., Portland, USA; http://www.schrodinger.com). For each compound, a set of MOLPRINT2D fingerprints was calculated using the default parameters. A similarity matrix, based on the Tanimoto similarities between each set of fingerprints, was calculated. In order to classify the compounds into several clusters, a Tanimoto cutoff of 0.8 was used. Seven groups of compounds were then obtained. Some of the groups were grouped together because the compounds they contained were chemically very similar, i.e. they contained the same core scaffold, obtaining at the end five different clusters of compounds. The compounds of each cluster represent therefore a group of very similar compounds, with an average Tanimoto coefficient of their MOLPRINT2D fingerprints greater than 0.8. The similarity matrix was also used as an input for the DendroUPGMA server (http://genomes.urv.es/UPGMA/) [30] to represent, as a dendrogram, the chemical similarities between molecules.

**Pharmacophore construction.** Energetically optimized, structure-based pharmacophores were constructed with the Glide v5.6 program (Schrödinger LLC., Portland, USA; http://www.schrodinger.com). This program accurately characterizes protein-ligand interactions based on energetic contributions such that energetically favorable features are incorporated into the pharmacophore [31]. The Glide XP scoring function was used to obtain an energetic description of each complex. The pharmacophore sites are ranked based on the Glide XP energies, and the most favorable sites are selected for the pharmacophore hypothesis. Aromatic rings were considered as hydrophobic groups. The PPARγ residues that interact with the sites of the above pharmacophores were visualized with the LigandScout v2.03 program (Inte:ligand, Vienna, Austria, http://www.inteligand.com/ligandscout/) [32].

**Molecular alignments.** The most crucial step for a 3D-QSAR construction model is the alignment of the molecules. We chose a structure-based docking strategy that was carried out using the poses predicted by docking using the Glide v5.6 program (Schrödinger LLC., Portland, USA; http://www.schrodinger.com). We only analyzed compounds with a similar chemical structure that we predicted that have very similar binding features to the receptor. These compounds were docked within the binding site of the 2Q5P PDB structure. The binding site was defined using the Receptor Grid Generation panel with the default options. Standard-precision (SP) docking was selected for screening the ligands. We selected the flexible docking mode, meaning that Glide internally generated conformations during the docking process. We did not request any constraints for docking. Each docking run recorded at most twenty poses per ligand that survived the post-docking minimization. GlideScore was used as the fitness function. The best scoring pose was selected for each ligand and used as an input structure for subsequent 3D-QSAR analyses. Moreover, to confirm that the docked poses that we obtained were realistic, we inspected manually the group of best scoring poses for each compound of the selected clusters to confirm that they contain the important intermolecular interactions with the receptor that we detect at the binding features analysis (and that, obviously, are also present in the 2Q5P complex). In addition, a cross docking analysis of the molecules used to build the 3D-QSAR with other PPARγ conformations derived from PDB complexes (i.e., 2Q5S and 2P4Y) showed similar results to the ones obtained with 2Q5P. Thus, this knowledge-based selection of docked poses ensures their realism.

**Generation of the 3D-QSAR models.** The selected conformations of the ligands, obtained with the previously described alignment protocol, were used for the generation of a pair of 3D-QSAR models (one for $pIC_{50}$ and another for the percentage of maximal activation). The Phase v3.2 program (Schrödinger LLC., Portland, USA; http://www.schrodinger.com) was employed to carry out the calculations using the Atom-Based 3D-QSAR panel. In the atom-based 3D-QSAR, a molecule is treated as a set of overlapping van der Waals spheres. To encode the basic characteristics of the

local chemical structure, each atom (and hence each sphere) is placed into one of six categories according to a simple set of rules: hydrogen atoms attached to polar atoms are classified as hydrogen bond donors (D); carbons, halogens, and C–H hydrogens are classified as hydrophobic/non-polar (H); atoms with an explicit negative ionic charge are classified as negative ionic (N); atoms with an explicit positive ionic charge are classified as positive ionic (P); non-ionic nitrogen and oxygen atoms are classified as electron-withdrawing (W); and all other types of atoms are classified as miscellaneous (X) [33]. The docking-predicted conformations of each ligand were first imported into the program together with their activity data. Then, training and test sets were chosen randomly using the Phase program. We chose a high training set percentage (80%) because the main aim of our models was to explain the relation between the selected ligands and their activities, not to predict activity values. Furthermore, in order to discard a possible influence of the splitting of the ligands into the training and test subsets on the resulting pair of 3D-QSAR models, we (1) randomly selected other 10 different training/test sets, (2) obtained their corresponding pairs of 3D-QSAR models, and (3) check their similarity relative to the initial pair of models by visual inspection. The 3D-QSAR model partitions the space occupied by the ligands into a cubic grid. Any structural component can occupy part of one or more cubes. The size of the cubes selected was 1 Å. The independent variables in the regression were given by the binary-valued occupancies ("bits") of the cubes (by structural components), while the dependent variables were the transactivation activity or the binding affinity. The regression was done by constructing a series of models with an increasing number of partial least square (PLS) factors. The accuracy of the models increases when the number of PLS factors increases until over-fitting starts to occur.

**Statistical validations of the 3D-QSAR models.** The performance of the initial pair of 3D-QSAR models was evaluated by measuring the accuracy of the predictions. The statistical parameters that were used to evaluate the predictions for the training set were: a) the coefficient of determination ($R^2$); b) the standard deviation of regression (SD); c) the F statistic, which measures the overall significance of the model; d) the statistical significance (P), which measures the probability that the correlation could occur by chance; and e) a stability value, which has a maximum value of 1 and measures the stability of the model predictions with changes in the training set composition. The parameters used to evaluate the predictions for the test set were: a) $Q^2$, the equivalent of $R^2$ for the test set; b) the root-mean-square error (RMSE); and c) the Pearson correlation coefficient (r).

## Results and Discussion

**Binding features of partial agonists.** A total of 205 structures of PPARγ agonists were retrieved from 12 SAR studies (sar1-12) of synthetic PPARγ agonists (Table 1) [18-29]. Based on their chemical similarities, these 205 compounds can be grouped into five clusters (Figure 1). Cluster 1 is composed of ligands from sar1 (aryl indole-2-carboxylic acids) and sar7 (N-sulfonyl-2-indole carboxamides), which consist of an indole system that contains a carboxylic group or a sulfonyl group at the second position [18, 24]. In essence, all of the ligands from this cluster have two lipophilic parts on either side of an acidic center. Cluster 2 is the largest family and contains compounds from sar2 (3-acylindole-1-benzylcarboxylic acids) [19], sar3 (benzoyl 2-methyl indoles) [20], sar4 (N-benzyl-indoles) [21] and sar5 (7-azaindoles) [22]. Compounds from this cluster have an indole group, like the compounds from cluster 1, but otherwise follow a different pattern. These compounds are made up of an acidic head and a lipophilic tail.

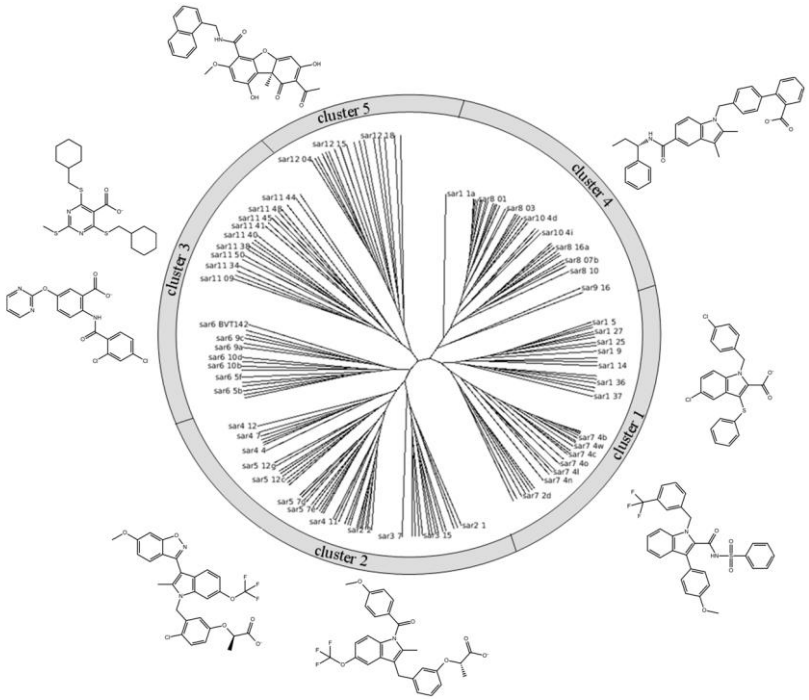



**Figure 1.** Representation of PPARγ partial agonist clusters. A total of 205 synthetic compounds from 12 SAR series were clustered by comparing their MOLPRINT2D fingerprints. The 2D structure of a representative member of each cluster is also shown.

With the aim of analyzing the binding differences between different PPARγ agonists, we constructed at least one energetically optimized pharmacophore [31] for each of the above clusters. This methodology quantifies the importance of each pharmacophore feature and allowed us to analyze the differences of receptor-ligand contacts between clusters. Table 2 shows the eight energy-based pharmacophores constructed from eight PDB structures that contain the PPARγ LBD crystallized with a partial agonist. Most of the pharmacophore sites are aromatic rings, highlighting the importance of hydrophobic interactions for the binding of PPARγ agonists with the receptor. Another significant feature is the presence of an acceptor site together with a negative site in the majority of the pharmacophores. This site corresponds to a carboxylic group present in the majority of the PPARγ partial agonists that forms a hydrogen bond with the Ser342 from the LBD of PPARγ. A comparison of the energy-based pharmacophores between clusters shows that the pharmacophores from clusters 1, 2 and 3 are similar, although the positions of the hydrophobic sites vary. The pharmacophores from clusters 4 and 5 are, however, slightly different. They contain an additional donor site, and the hydrophobic sites occupy a different region when compared with the pharmacophores from clusters 1, 2 and 3. Table 2 also shows the energy-based pharmacophore of a PPARγ full agonist. This pharmacophore has sites similar to those of the previous pharmacophores, but their locations are very different, highlighting the binding differences between full and partial PPARγ agonists.

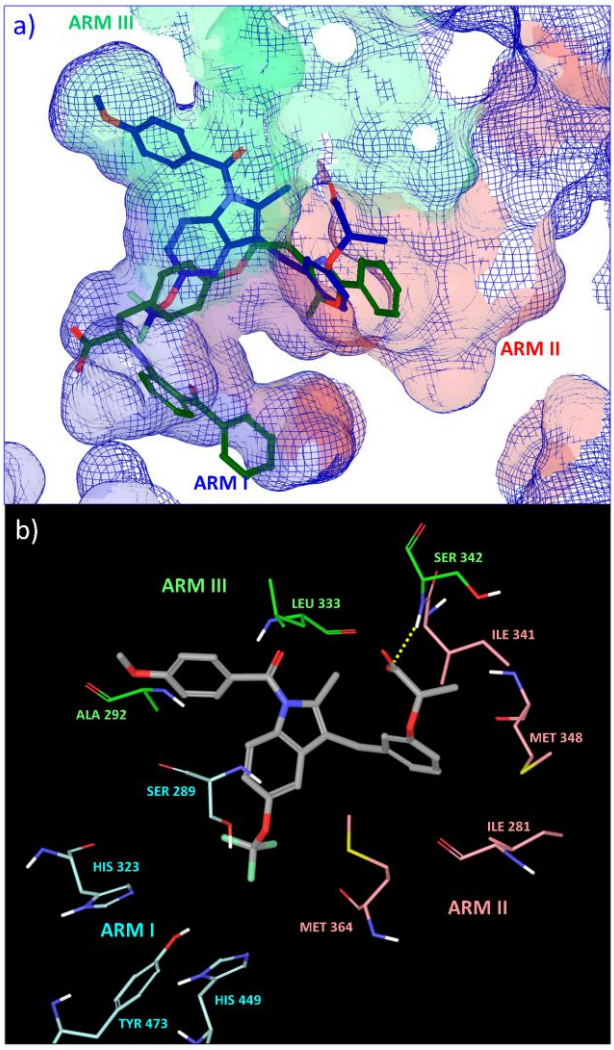
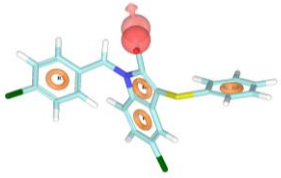
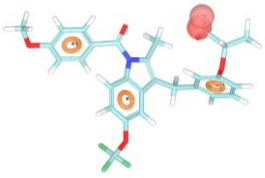



**Figure 2. A)** The ligand-binding domain (LBD) of PPARγ complexed with a partial agonist, a benzoyl 2-methyl indole derivate (MRL-24 from PDB 2Q5P), colored in blue, superimposed with the structure of a full agonist, farglitazar, colored in green. The partial agonist occupies mainly arm II and arm III of the LDB of PPARγ, but the full agonist occupies mainly arm I and arm II. **B)** The main interactions between the PPARγ partial agonist MRL-24 and the LBD of PPARγ. The conserved hydrophobic interactions between Ile281, Ala292, Ile326, Ile341, Leu333, Met348 and Met364 and the hydrophobic sites of MRL-24 are shown. These interactions are common to nearly all PPARγ agonists, including full agonists. Hydrogen bonds between Ser342 and the carboxylic acid from MRL-24 are also shown by a dashed yellow line. This hydrogen bond is conserved between some PPARγ partial agonists but not for full agonists.

**Table 2**. Energy-based pharmacophores for eight structures that contain the PPARγ ligand-binding domain crystallized with a partial agonist plus one structure complexed with a full agonist. Pink spheres represent hydrogen bond acceptors, green spheres represent hydrophobic groups, orange rings represent aromatic rings, light-blue spheres represent hydrogen bond donors, and red spheres represent negative ionizable groups. The PPARγ residues involved in the interaction with the ligand are also shown. All interactions correspond to hydrophobic interactions, with the exception of the residues marked with an asterisk, which form hydrogen bonds with the polar groups of the ligands. The interactions conserved in the majority of the structures that contain a PPARγ agonist are shown in bold. All energetic pharmacophores are presented in the same relative orientation in order to allow for an easier comparison. Contact residues were defined using LigandScout.

| Cluster | PDB code | Energetic Pharmacophore | Contact Residues | | |
|---|---|---|---|---|---|
| | | | ARM I | ARM II | ARM III |
| 1 | 2Q5S |  | **Ile326**<br>Phe363 | Ile249<br>**Ile281**<br>**Val339**<br>**Ile341**<br>**Met348**<br>**Leu353**<br>**Met364** | **Ala292**<br>**Leu330**<br>**Ser342*** |
| | 2HFP |  | **Ile326** | **Ile281**<br>Met329<br>**Val339**<br>**Ile341**<br>**Met348**<br>**Leu353**<br>**Met364** | Ile262<br>Lys265*<br>Arg288<br>**Ala292**<br>**Leu330**<br>**Leu333**<br>**Ser342*** |
| 2 | 2Q5P |  | **Ile326**<br>Tyr327 | Phe264<br>**Ile281**<br>**Val339**<br>**Ile341**<br>**Met348** | Arg288<br>**Ala292**<br>**Leu330**<br>**Leu333**<br>**Ser342*** |
| | 2P4Y |  | **Ile326**<br>Tyr327<br>Phe363 | **Ile281**<br>Met329<br>**Val339**<br>**Ile341**<br>**Met348**<br>**Leu353**<br>**Met364** | Phe287<br>**Ala292**<br>**Leu330**<br>**Leu333**<br>**Ser342*** |

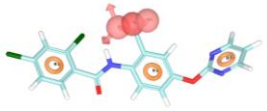
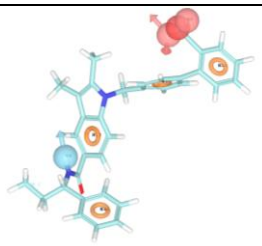
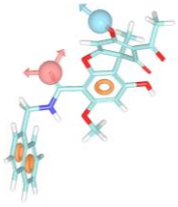
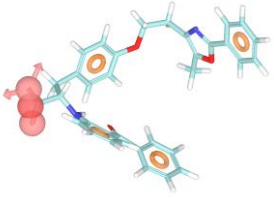
| | | | | | |
|---|---|---|---|---|---|
| 3 | 2Q6S |  | - | Ile249<br>Leu255<br>Phe264<br>**Val339**<br>**Met364** | **Ala292**<br>**Leu330**<br>**Leu333**<br>**Ser342*** |
| | Cmpd 50 |  | **Ile326** | Leu255<br>**Ile281**<br>**Val339**<br>**Ile341**<br>**Leu353**<br>**Met364** | Ile262<br>Thr268<br>**Ala292**<br>**Leu330**<br>**Leu333**<br>**Ser342*** |
| 4 | 3KMG |  | Phe282<br>Ser289 *<br>**Ile326**<br>Tyr327*<br>Phe363<br>Leu453<br>Leu465<br>Leu469 | **Ile281**<br>**Val339**<br>**Ile341**<br>**Met348**<br>**Met364** | Arg288*<br>**Ala292**<br>**Leu330**<br>**Leu333**<br>**Ser342*** |
| 5 | 3LMP |  | **Ile326** | Met334<br>**Leu353**<br>**Met364** | **Leu330**<br>Cys285* |
| Full agonist | 1FM9 |  | Phe282<br>Ser289*<br>**Ile326**<br>Phe360<br>Phe363<br>Leu453<br>Ile456<br>Leu465<br>Leu469<br>Tyr473* | **Ile281**<br>**Val339**<br>**Ile341**<br>**Met348**<br>**Leu353**<br>**Met364** | **Leu330** |

Table 2 also shows the binding differences between full and partial agonists (and between some partial agonists) from the receptor point of view. In this table, the PPARγ residues that interact with each site of the pharmacophores are shown. The hydrophobic interactions between Ile281, Ala292, Ile326, Ile341, Leu330, Leu333, Val339, Met348, Leu353 and Met364 and the hydrophobic sites of the ligands are conserved in the majority of the structures (see Table 2), even for the full agonist. The residue that interacts through a hydrogen bond with the acceptor/negative site of the ligand differs depending on whether the ligand is a full or partial agonist. Partial agonists (with the exception of compounds from cluster5) form a hydrogen bond with

Ser342. However, the residues that interact through a hydrogen bond in full agonists are usually Ser289 and Tyr473. There are other binding differences between partial and full agonists. If we split the LBD of PPARγ into three parts, arm I, arm II and arm III (see Figure 2), we observe that the partial agonists (with the exception of compounds from cluster4) basically interact with arms II and III, but the full agonists basically interact with arms I and II (see Table 2). Thus, in agreement with previous results [10], our analyses show that full and partial agonists show different binding patterns for the LBD of PPARγ. The binding patterns of different partial agonists are also slightly different. Ligands from cluster4 occupy arm I, like full agonists, and also make two additional hydrogen bonds with Ser289 and Tyr327 (see Table 2). Ligands from cluster5 make few contacts with the LBD of PPARγ because they are surrounded by several water molecules. The binding profiles of compounds from clusters 1, 2 and 3 are similar (see Table 2).

**Generation of 3D-QSAR models.** We selected the sar1, sar2, sar3, sar4 and sar5 series of PPARγ agonists for the construction of two 3D-QSAR models. 3D-QSAR techniques have efficiently provided models in reasonable agreement with those deduced by the crystal structure of PPARγ complexes [34]. We used this methodology not for predicting the activity or binding affinity of putative PPARγ agonists, but rather to analyze which interactions between the LBD of PPARγ and its ligands increase the binding affinity without increasing the PPARγ transactivation activity. The sar1, sar2, sar3, sar4 and sar5 series were selected because for these compounds we have a wide range of measured $IC_{50}$ (i.e., binding affinity measured by the displacement of a radiolabeled full agonist) values, tested under the same assay conditions, and values for the transactivation activity (see Table 1). Ligands from sar8 were not used because their binding mode is quite different from that used by the agonists from clusters 1 and 2 (see Table 2). The sar1, sar2, sar3, sar4 and sar5 series form clusters 1 and 2 in Figure 1 and contain a set of 82 indole-based PPARγ agonist derivates with a similar binding profile. With these compounds, we constructed two atom-based 3D-QSAR models, one analyzing the binding affinity between the ligands and PPARγ (called the $pIC_{50}$ model) and one analyzing the transactivation activity of PPARγ (called the transactivation model). For the first model, we used values of $pIC_{50}$, and for the second, we used the percentage of maximal activation relative to the full agonist rosiglitazone. Activation levels that reach the maximal activation of rosiglitazone are considered full agonists, while those reaching 20-60% of rosiglitazone maximal activation are considered partial agonists. Table 3 and Figure 3 show the statistic fits of the constructed 3D-QSAR models. For both models, as it is shown in Figure 3, the activity values of the ligands from the training and test sets are homogeneously distributed along all the activity range. In both models, to avoid an over-fitting effect, two PLS factors were chosen. The Pearson correlation coefficient of the pIC50 model was 0.77 with an $R^2$ for the training set and a $Q^2$ for the test set of 0.67 and 0.55, respectively. For the transactivation model, the Pearson correlation

coefficient was 0.72 with an $R^2$ for the training set and a $Q^2$ for the test set of 0.71 and 0.40, respectively. The low $Q^2$ values imply that there is a greater difference between the experimental values of binding affinity and transactivation activity and the values predicted by each model. This difference is more important when predicting the transactivation activity of full agonists (see Figure 3b). This may be due to the fact that the great majority of compounds in the dataset used are partial agonists. The $R^2$ values for the training set are better. As our main purpose was to use the 3D-QSAR models for analyzing the interactions between the LBD of PPARγ and a group of similar PPARγ agonists, the $R^2$ values are more relevant. These $R^2$ values and the scatter plots for the training set in Figure 3 indicate a reasonably good correlation between the predicted and experimental activities and validate the use of both models.

**Table 3.** Statistics of the best 3D-QSAR models for analyzing the binding affinity ($pIC_{50}$ model) and the transactivation activity of PPARγ (% max activation model) derived from an 80% randomly selected training set. See the Computational Methods section for the meaning of the statistical parameters used. To avoid an over-fitting effect, two factor models were chosen.

| Model | # | SD | $R^2$ | F | P | Stability | RMSE | $Q^2$ | Pearsonr |
|---|---|---|---|---|---|---|---|---|---|
| | 1 | 0.62 | 0.31 | 26.1 | 3.8e-06 | 0.94 | 0.68 | 0.14 | 0.43 |
| | **2** | **0.43** | **0.67** | **58.5** | **1.53e-14** | **0.49** | **0.49** | **0.55** | **0.77** |
| **$pIC_{50}$** | 3 | 0.36 | 0.78 | 64.5 | 3.66e-18 | 0.38 | 0.57 | 0.40 | 0.63 |
| | 4 | 0.24 | 0.90 | 121.8 | 1.2e-26 | 0.15 | 0.63 | 0.26 | 0.51 |
| | 5 | 0.19 | 0.94 | 166.7 | 1.55e-31 | -0.02 | 0.63 | 0.25 | 0.50 |
| | 1 | 13.64 | 0.42 | 38.2 | 9.36e-08 | 0.85 | 17.83 | 0.28 | 0.55 |
| | **2** | **9.69** | **0.71** | **64.4** | **8.57e-15** | **0.55** | **16.27** | **0.40** | **0.72** |
| **% max activation** | 3 | 6.4 | 0.88 | 121.1 | 3.42e-23 | 0.29 | 17.61 | 0.30 | 0.63 |
| | 4 | 4.75 | 0.93 | 175.7 | 8.78e-29 | 0.22 | 17.31 | 0.32 | 0.64 |
| | 5 | 2.98 | 0.97 | 372.3 | 9.95e-38 | 0.15 | 16.33 | 0.40 | 0.72 |

Figures 4 and 5 show the representation of the 3D-QSAR models. In these figures, the cubes that represent the model are displayed and colored according to the sign of their coefficient values. Blue and red cubes are used, respectively, for positive and negative coefficients and indicate regions that increase or decrease the analyzed parameter. One of the advantages of using these representations is that the position of the cubes of the 3D-QSAR model can be compared with the positions of the amino acid residues in the active site. This might give an insight as to which functional groups are desirable or undesirable at certain positions in a molecule.

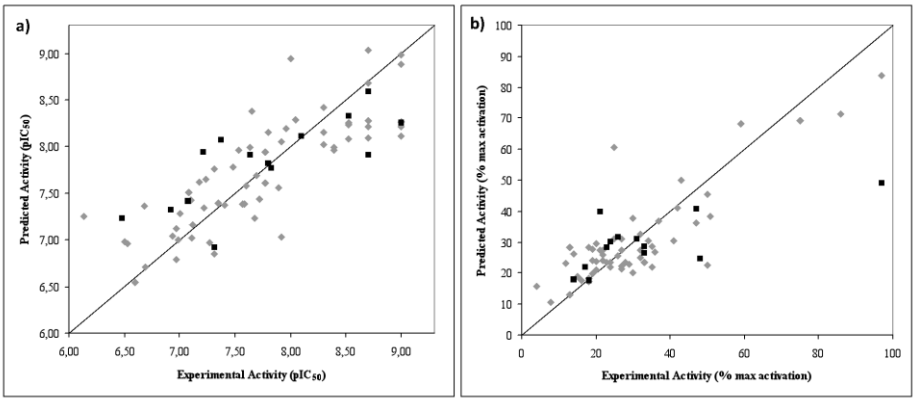



**Figure 3.** Scatter plots for the two factors (a) $pIC_{50}$ and (b) percentage of maximal activation models applied to the training set (colored in gray) and the test set (colored in black).

Figure 4b shows the favorable and unfavorable regions for the binding affinity. Similar favorable and unfavorable regions were obtained when ten additional 3D-QSAR models were generated using different training set selections obtained at random (results not shown). The favorable regions for binding are located at regions that interact with arms I and II and the right part of arm III (which includes Ser342) of the LBD of PPARγ. When viewing the 3D-QSAR model by atom type, we see that the hydrophobic (Figure 4c) and the electron-withdrawing contributions (Figure 4d) are the most important for the binding affinity of the compounds analyzed (whereas the rest of the contributions have a very limited role in binding affinity; results not shown). The electron-withdrawing contributions are favorable at arm I, where hydrogen bonds with Ser289, His323, His449 and Tyr473 can be established, and arm II, where a hydrogen bond with Ser342 is common to most PPARγ partial agonists. Hydrophobic interactions are the most important binding forces between PPARγ agonists and the LBD of PPARγ. The representation of the 3D-QSAR model in Figure 4c suggests that, when more hydrophobic interactions occur with arm I and arm II of the LBD of PPARγ, a greater binding affinity is seen in the compound. Figures 4e and 4f display the cubes of the 3D-QSAR model grid that are occupied by two compounds from the SAR series analyzed. In these representations, we can see which parts of the ligand have a positive or a negative contribution to the parameter analyzed, which is, in this case, the binding affinity. Figure 4e shows the sar1_24 compound [18], which has one of the lowest binding affinities in this series. Our 3D-QSAR model explains the low binding affinity of this compound because, although it can partially interact with arm II through a carboxylic group (see the upper blue cubes in Figure 4e), it lacks a hydrophobic moiety at indole position 6 to interact with arm I. This compound also contains a trifluoromethyl group, a group with high electronegativity, located in the hydrophobic environment of arm III (see the red cubes in Figure 4e). When the

compound from our ligand dataset with the highest binding affinity (i.e., sar2_1 in Figure 4f) is considered in the context of the model, we see an excellent fit with the blue areas of the model (see Figure 4f).

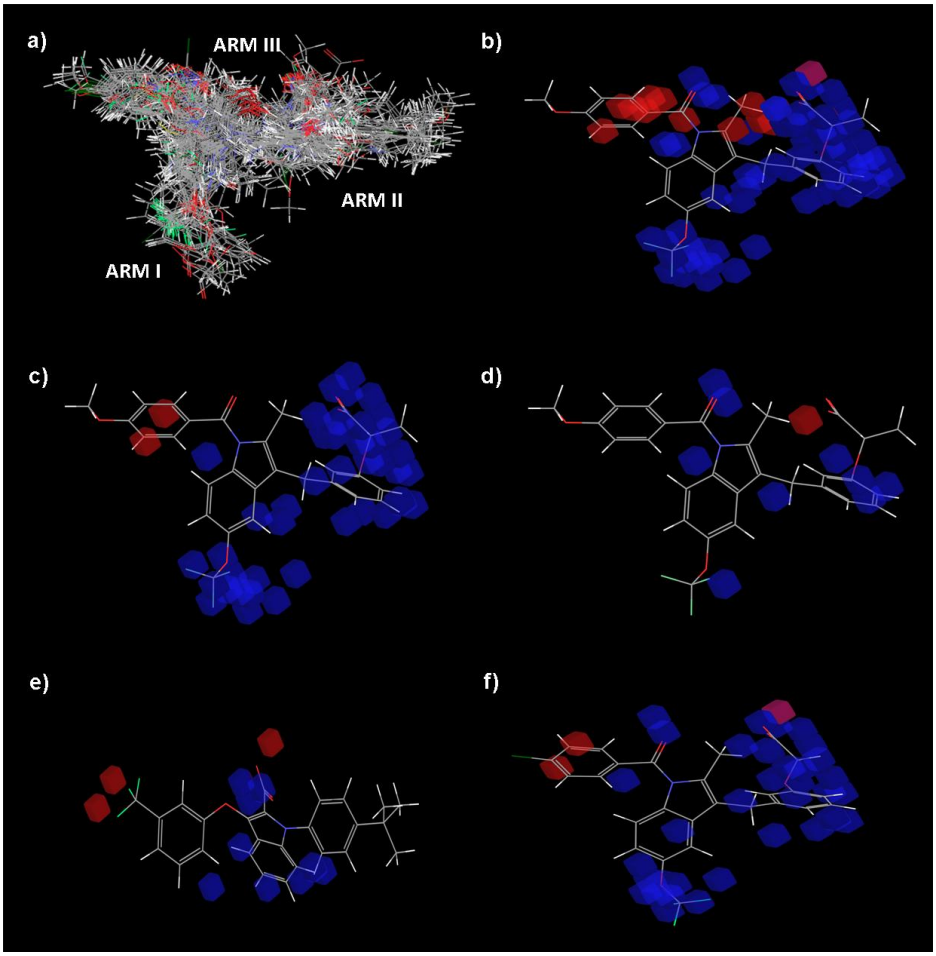



**Figure 4.** Representation of the $pIC_{50}$ model. The structural alignment of the selected docking poses of all compounds (a) was used to construct an atom-based 3D-QSAR model. Blue and red cubes indicate, respectively, regions that are favorable and unfavorable for binding to the LBD of PPARγ. The thresholds used for considering a region with a positive or negative contribution were 1.0 e-02 and -1.4 e-02, respectively. The complete 3D-QSAR model is displayed in panel (b), whereas panels (c) and (d) show the hydrophobic and electron-withdrawing contributions, respectively. Panels (e) and (f) show the cubic volume elements that are occupied by one of the compounds of the series with the lowest (i.e., sar1_24) and the highest (i.e., sar2_1) binding affinity, respectively. All panels are presented in the same relative orientation in order to allow for an easier comparison.

Figure 5 shows a representation of the transactivation model. In this model, the transactivation activity of PPARγ is the variable analyzed. Figure 5a shows the favorable and unfavorable regions for transactivation activity. Similar favorable and unfavorable regions were obtained when ten additional 3D-QSAR models were generated using different training set selections obtained at random (results not shown). The favorable regions are located at arm I and at part of arm III. Interestingly, unfavorable regions are located at arm II and the right part (which corresponds to Ser342) of arm III. Figures 5b-d show that the main contribution to the transactivation activity of PPARγ is caused by hydrophobic interactions, specifically the hydrophobic interactions that can be established with the hydrophobic residues of the LBD of PPARγ that form arm II and part of arm I (see Figure 5b). Hydrophobic interactions with arm II and part of arm III are marked as unfavorable in the model (Figure 5b). This effect is due to the fact that partial agonists do not occupy arm I but do occupy arm II and the right part of arm III. In addition, an unfavorable hydrophobic interaction is also localized at arm I (see the red cubes at the bottom and left side of Figure 5b). A carboxylic group that makes a hydrogen bond with Ser289 may occupy this part of the ligand, especially for full PPARγ agonists. This interaction is crucial for the stabilization of H12 and for the transactivation activity of PPARγ. Thus, when this region is occupied by a hydrophobic group, a hydrogen bond cannot be established, and the transactivation activity of PPARγ decreases.

The importance of this interaction for the transactivation activity of PPARγ is also visualized in Figures 5c and 5d, when the electron-withdrawing and negatively charged contributions are represented in the transactivation model. In both figures, a blue cube at arm I (at the bottom of the figures) represents the importance of polar interactions in this region. Figure 5d also shows that the negatively charged contributions at the right part of arm III are unfavorable for the transactivation activity. This negative contribution reflects the fact that most partial agonists have a carboxylic group at this region that forms a hydrogen bond with Ser342. This hydrogen bond neither stabilizes H12 nor activates the transactivation activity of PPARγ. As the majority of partial agonists form this hydrogen bond and their transactivation activity is low, the model marks this interaction as unfavorable for the transactivation activity. Figures 5e and 5f show, respectively, the 3D-QSAR model represented only by the cubic volume elements that are occupied by one of the most inactive compounds (i.e., the sar1_6 compound) and the most active compound (i.e., the sar3_12 compound) in terms of transactivation activity. The sar1_6 compound [18] has only a maximal transactivation activity of 8% relative to rosiglitazone. Figure 5e shows that this ligand basically occupies arm III and arm II of the receptor and makes hardly any of the favorable interactions shown as blue cubes in Figure 5. The sar3_12 compound [20] has a maximal transactivation activity of 97% relative to rosiglitazone. This compound fits the blue areas of the 3D-QSAR model perfectly, as it can interact with the receptor through the favorable regions at arms I, II and III (see Figure 5f).

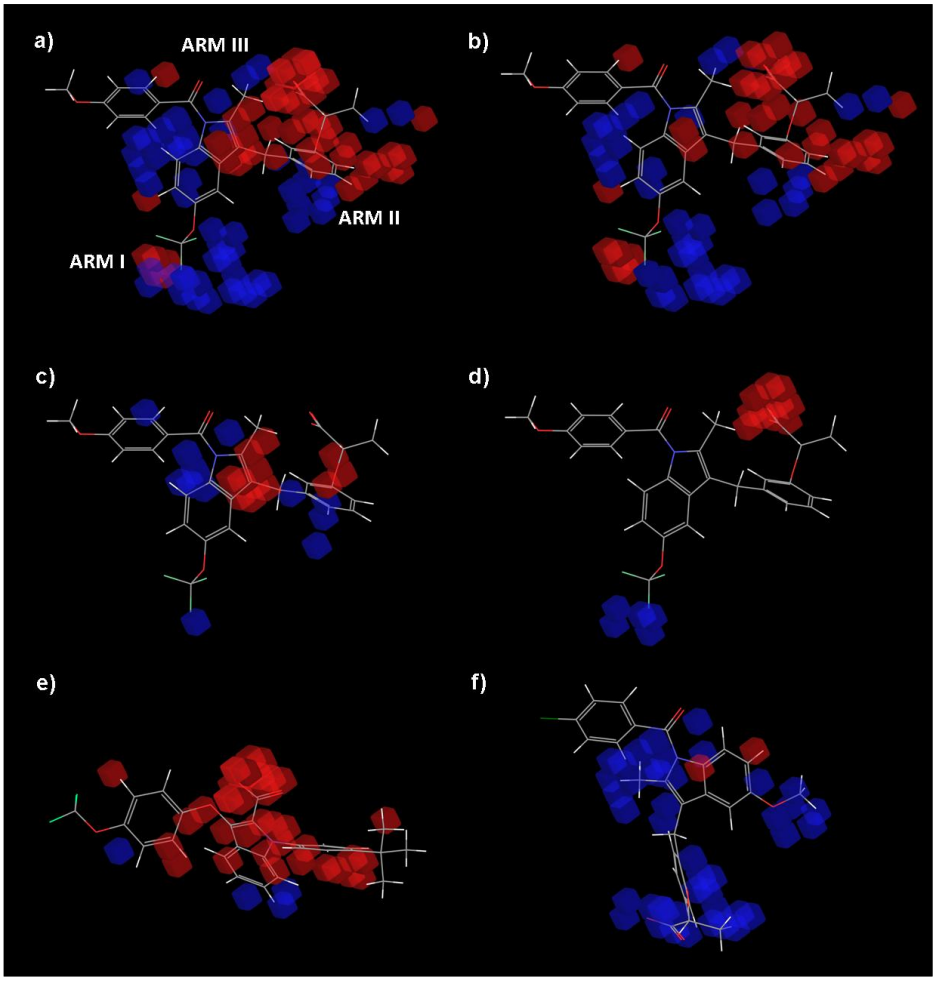



**Figure 5.** Representation of the % of maximal transactivation model. Blue and red cubes indicate, respectively, regions that are favorable and unfavorable for the transactivation activity of PPARγ. The thresholds used for considering a region with a positive or negative contribution were 2.0 e-01 and -2.0 e-01, respectively. The complete 3D-QSAR model is displayed in panel (a), whereas panels (b), (c) and (d) show the hydrophobic, electron-withdrawing and negatively charged contributions, respectively. Panels (e) and (f) show, respectively, the cubic volume elements that are occupied by one of the least actives (i.e., sar1_6) and one of the most actives (i.e., sar3_12) compounds. All panels are presented in the same relative orientation in order to allow for an easier comparison.

Arm I of the LBD of PPARγ is an important part for the binding and the transactivation activity of PPARγ ligands. The ligands that occupy this arm interact with PPARγ through a series of hydrophobic interactions and a net of hydrogen bonds with the side chains of Ser289, His323, His449 and Tyr473. These interactions

stabilize H12 and are responsible for the transactivation activity of PPARγ [9, 11]. When a hydrophobic group occupies the region of the carboxyl group responsible for the net of hydrogen bonds with the side chains of Ser289, His323, His449 and Tyr473, the transactivation activity of PPARγ decreases (see Figure 5). Other regions of the LBD of PPARγ that also contribute to the transactivation activity of PPARγ include the regions of arms III and II that are closer to arm I. Hydrophobic interactions in these regions are favorable for the transactivation activity of PPARγ (see Figure 5). However, the region of arm III furthest from arm II does not contribute to this activity. This region, which includes Ser342, is the region most occupied by partial agonists. Hydrophobic interactions between the PPARγ residues from arms I and II and the hydrophobic groups of PPARγ ligands are very important for their binding (see Figure 4). In addition, partial agonists can establish a hydrogen bond with Ser342. Because arm I contributes significantly to the transactivation activity of PPARγ, this region must not be occupied by potential PPARγ partial agonists.

## Conclusion

The ideal PPARγ partial agonists to be used as antidiabetic compounds should show a low transactivation activity but a high binding affinity to inhibit phosphorylation at Ser273. Our models suggest that effective partial agonists should have a hydrophobic moiety and an acceptor site with an appropriate conformation to interact with arm II and to establish a hydrogen bond with Ser342 or an equivalent residue. Despite the fact that interactions with arm I increase the binding affinity, this region should be avoided in order to decrease the transactivation activity of potential PPARγ partial agonists.

## Acknowledgements

This manuscript was edited for English language fluency by American Journal Experts. This study was supported by Grant Number AGL2008-00387/ALI from the Ministerio de Educación y Ciencia of the Spanish Government and the ACC1Ó (TECCT10-1-0008) program (Generalitat de Catalunya). The authors wish to thank the Servei de Disseny de Fàrmacs (Drug Design Service) of the Catalonia Supercomputer Center (CESCA) for providing access to Schrödinger software.

## References

1. Berger J, Moller DE (2002) Annu. Rev. Med. 53: 409

2. Willson TM, Lambert MH, Kliewer SA (2001) Annu. Rev. Biochem. 70: 341

3. Lehmann JM, Moore LB, Smith-Oliver TA, Wilkison WO, Willson TM, Kliewer SA (1995) J. Biol. Chem. 270: 12953

4. Feldman PL, Lambert MH, Henke BR (2008) Curr. Top. Med. Chem. 8: 728

5. Pourcet B, Fruchart J, Staels B, Glineur C (2006) Expert. Opin. Emerg. Drugs. 11: 379

6. Jones D (2010) Nat. Rev. Drug Discov. 9: 668

7. Gelman L, Feige JN, Desvergne B (2007) Biochim. Biophys. Acta 1771: 1094

8. Grether U, Klaus W, Kuhn B, Maerki HP, Mohr P, Wright MB (2010) ChemMedChem. 5: 1973

9. Farce A, Renault N, Chavatte P (2009) Curr. Med. Chem. 16: 1768

10. Zoete V, Grosdidier A, Michielin O (2007) Biochim. Biophys. Acta. 1771: 915

11. Pochetti G, Godio C, Mitro N, Caruso D, Galmozzi A, Scurati S, Loiodice F, Fracchiolla G, Tortorella P, Laghezza A, Lavecchia A, Novellino E, Mazza F, Crestani M (2007) J. Biol. Chem. 282: 17314

12. Bruning JB, Chalmers MJ, Prasad S, Busby SA, Kamenecka TM, He Y, Nettles KW, Griffin PR (2007) Structure 15: 1258

13. Montanari R, Saccoccia F, Scotti E, Crestani M, Godio C, Gilardi F, Loiodice F, Fracchiolla G, Laghezza A, Tortorella P, Lavecchia A, Novellino E, Mazza F, Aschi M, Pochetti G (2008) J. Med. Chem. 51: 7768

14. Lu I, Huang C, Peng Y, Lin Y, Hsieh H, Chen C, Lien T, Lee H, Mahindroo N, Prakash E, Yueh A, Chen H, Goparaju CMV, Chen X, Liao C, Chao Y, Hsu JT, Wu S (2006) J. Med. Chem. 49: 2703

15. Choi JH, Banks AS, Estall JL, Kajimura S, Boström P, Laznik D, Ruas JL, Chalmers MJ, Kamenecka TM, Blüher M, Griffin PR, Spiegelman BM (2010) Nature 466: 451

16. Houtkooper RH, Auwerx J (2010) Nature 466: 443

17. Willson TM, Cobb JE, Cowan DJ, Wiethe RW, Correa ID, Prakash SR, Beck KD, Moore LB, Kliewer SA, Lehmann JM (1996) J. Med. Chem. 39: 665

18. Dropinski JF, Akiyama T, Einstein M, Habulihaz B, Doebber T, Berger JP, Meinke PT, Shi GQ (2005) Bioorg. Med. Chem. Lett. 15: 5035

19. Liu W, Liu K, Wood HB, McCann ME, Doebber TW, Chang CH, Akiyama TE, Einstein M, Berger JP, Meinke PT (2009) J. Med. Chem. 52: 4443

20. Acton JJ, Black RM, Jones AB, Moller DE, Colwell L, Doebber TW, Macnaul KL, Berger J, Wood HB (2005) Bioorg. Med. Chem. Lett. 15: 357

21. Liu K, Black RM, Acton JJ, Mosley R, Debenham S, Abola R, Yang M, Tschirret-Guth R, Colwell L, Liu C, Wu M, Wang CF, MacNaul KL, McCann ME, Moller DE, Berger JP, Meinke PT, Jones AB, Wood HB (2005) Bioorg. Med. Chem. Lett. 15: 2437

22. Debenham SD, Chan A, Lau FW, Liu W, Wood HB, Lemme K, Colwell L, Habulihaz B, Akiyama TE, Einstein M, Doebber TW, Sharma N, Wang CF, Wu M, Berger JP, Meinke PT (2008) Bioorg. Med. Chem. Lett. 18: 4798

23. Thor M, Beierlein K, Dykes G, Gustavsson AL, Heidrich J, Jendeberg L, Lindqvist B, Pegurier C, Roussel P, Slater M, Svensson S, Sydow-Bäckman M, Thornström U, Uppenberg J (2002) Bioorg. Med. Chem. Lett. 12: 3565

24. Hopkins CR, O'neil SV, Laufersweiler MC, Wang Y, Pokross M, Mekel M, Evdokimov A, Walter R, Kontoyianni M, Petrey ME, Sabatakos G, Roesgen JT, Richardson E, Demuth TP (2006) Bioorg. Med. Chem. Lett. 16: 5659

25. Lamotte Y, Martres P, Faucher N, Laroze A, Grillot D, Ancellin N, Saintillan Y, Beneton V, Gampe RT (2010) Bioorg. Med. Chem. Lett. 20: 1399

26. Goebel M, Clemenz M, Staels B, Unger T, Kintscher U, Gust R (2009) ChemMedChem. 4: 445

27. Goebel M, Staels B, Unger T, Kintscher U, Gust R (2009) ChemMedChem. 4: 1136

28. Seto S, Okada K, Kiyota K, Isogai S, Iwago M, Shinozaki T, Kitamura Y, Kohno Y, Murakami K (2010) J. Med. Chem. 53: 5012

29. Furukawa A, Arita T, Satoh S, Wakabayashi K, Hayashi S, Matsui Y, Araki K, Kuroha M, Ohsumi J (2010) Bioorg. Med. Chem. Lett. 20: 2095

30. Garcia-Vallvé S, Palau J, Romeu A (1999) Mol. Biol. Evol. 16: 1125

31. Salam NK, Nuti R, Sherman W (2009) J Chem Inf Model. 49: 2356

32. Wolber G, Langer T (2005) J. Chem. Inf. Model. 45: 160

33. Dixon SL, Smondyrev AM, Knoll EH, Rao SN, Shaw DE, Friesner RA (2006) J. Comput. Aided Mol. Des. 20: 647

34. Giaginis C, Theocharis S, Tsantili-Kakoulidou A (2009) Mini Rev Med Chem. 9: 1075

## Supporting information

**Table S1.** Experimental and predicted values of the $pIC_{50}$ (i.e. the binding affinity measured by the displacement of a radio-labelled full agonist).

| Ligand Name | QSAR Set | Experimental Activity | Predicted Activity | Residual Activity |
|---|---|---|---|---|
| sar5_7a | training | 6.14 | 7.25 | -1.11 |
| sar1_24 | training | 6.50 | 6.98 | -0.48 |
| sar1_14 | training | 6.53 | 6.96 | -0.43 |
| sar1_34 | training | 6.60 | 6.54 | 0.06 |
| sar3_7 | training | 6.68 | 7.36 | -0.68 |
| sar1_38 | training | 6.69 | 6.71 | -0.02 |
| sar1_9 | training | 6.94 | 7.04 | -0.10 |
| sar1_10 | training | 6.97 | 7.12 | -0.15 |
| sar5_12c | training | 6.98 | 6.79 | 0.19 |
| sar1_13 | training | 6.99 | 7.01 | -0.02 |
| sar3_10 | training | 7.00 | 7.29 | -0.28 |
| sar1_18 | training | 7.08 | 7.51 | -0.43 |
| sar3_15 | training | 7.10 | 7.43 | -0.33 |
| sar1_8 | training | 7.11 | 7.02 | 0.09 |
| sar1_33 | training | 7.12 | 7.17 | -0.05 |
| sar5_7b | training | 7.18 | 7.62 | -0.44 |
| sar1_11 | training | 7.22 | 7.35 | -0.13 |
| sar1_39 | training | 7.24 | 7.65 | -0.41 |
| sar5_12b | training | 7.27 | 6.97 | 0.29 |
| sar1_06 | training | 7.31 | 6.85 | 0.46 |
| sar3_13 | training | 7.31 | 7.77 | -0.46 |
| sar3_16 | training | 7.35 | 7.40 | -0.05 |
| sar1_23 | training | 7.41 | 7.38 | 0.03 |
| sar4_5 | training | 7.48 | 7.78 | -0.30 |
| sar4_4 | training | 7.54 | 7.96 | -0.42 |
| sar1_15 | training | 7.57 | 7.39 | 0.18 |
| sar1_1a | training | 7.59 | 7.38 | 0.21 |
| sar5_12a | training | 7.60 | 7.58 | 0.02 |
| sar5_12f | training | 7.64 | 8.00 | -0.36 |
| sar5_12h | training | 7.66 | 8.38 | -0.72 |
| sar1_35 | training | 7.68 | 7.23 | 0.45 |
| sar4_9 | training | 7.70 | 7.69 | 0.01 |
| sar1_16 | training | 7.72 | 7.44 | 0.28 |
| sar1_17 | training | 7.77 | 7.61 | 0.16 |
| sar2_12i | training | 7.77 | 7.95 | -0.18 |
| sar2_12k | training | 7.80 | 8.16 | -0.36 |

| | | | | |
|---|---|---|---|---|
| sar1_19 | training | 7.89 | 7.56 | 0.33 |
| sar1_5 | training | 7.92 | 7.03 | 0.89 |
| sar2_12f | training | 7.92 | 8.06 | -0.14 |
| sar2_12b | training | 7.96 | 8.20 | -0.24 |
| sar5_7d | training | 8.00 | 8.94 | -0.94 |
| sar2_12d | training | 8.05 | 8.29 | -0.24 |
| sar2_12j | training | 8.30 | 8.02 | 0.28 |
| sar4_7 | training | 8.30 | 8.16 | 0.15 |
| sar5_7c | training | 8.30 | 8.42 | -0.12 |
| sar2_12c | training | 8.40 | 7.97 | 0.43 |
| sar5_7f | training | 8.40 | 7.99 | 0.41 |
| sar3_25 | training | 8.52 | 8.25 | 0.27 |
| sar3_28 | training | 8.52 | 8.09 | 0.43 |
| sar5_12e | training | 8.52 | 8.24 | 0.28 |
| sar2_2 | training | 8.70 | 8.10 | 0.60 |
| sar3_20 | training | 8.70 | 8.69 | 0.01 |
| sar4_12 | training | 8.70 | 8.28 | 0.42 |
| sar5_7i | training | 8.70 | 8.22 | 0.48 |
| sar1_29 | training | 8.70 | 9.04 | -0.34 |
| sar2_1 | training | 9.00 | 8.21 | 0.79 |
| sar2_12a | training | 9.00 | 8.11 | 0.89 |
| sar3_22 | training | 9.00 | 8.88 | 0.12 |
| sar5_7h | training | 9.00 | 8.27 | 0.73 |
| sar5_7j | training | 9.00 | 8.99 | 0.01 |
| sar3_9 | test | 6.48 | 7.24 | -0.76 |
| sar1_12 | test | 6.92 | 7.33 | -0.41 |
| sar3_12 | test | 7.08 | 7.42 | -0.34 |
| sar4_14 | test | 7.21 | 7.95 | -0.74 |
| sar1_7 | test | 7.31 | 6.92 | 0.39 |
| sar4_13 | test | 7.38 | 8.07 | -0.69 |
| sar3_19 | test | 7.64 | 7.91 | -0.27 |
| sar5_7e | test | 7.80 | 7.82 | -0.02 |
| sar5_12d | test | 7.82 | 7.78 | 0.05 |
| sar2_12e | test | 8.10 | 8.12 | -0.02 |
| sar5_12g | test | 8.52 | 8.33 | 0.19 |
| sar4_11 | test | 8.70 | 7.92 | 0.78 |
| sar5_7g | test | 8.70 | 8.59 | 0.11 |
| sar3_24 | test | 9.00 | 8.25 | 0.75 |

**Table S2.** Experimental and predicted values of the % of maximal transactivation activities relative to the full agonist rosiglitazone.

| Ligand Name | QSAR Set | Experimental Activity | Predicted Activity | Residual Activity |
|---|---|---|---|---|
| sar1_39 | training | 4 | 16 | -12 |
| sar1_06 | training | 8 | 11 | -3 |
| sar1_29 | training | 12 | 23 | -11 |
| sar1_14 | training | 13 | 28 | -15 |
| sar1_8 | training | 13 | 13 | 0 |
| sar5_7j | training | 14 | 26 | -12 |
| sar1_9 | training | 15 | 19 | -4 |
| sar1_15 | training | 16 | 18 | -2 |
| sar1_33 | training | 18 | 17 | 1 |
| sar1_1a | training | 18 | 28 | -10 |
| sar1_7 | training | 18 | 18 | 0 |
| sar4_14 | training | 19 | 24 | -5 |
| sar5_7f | training | 19 | 20 | -1 |
| sar5_7g | training | 19 | 28 | -9 |
| sar5_12g | training | 20 | 30 | -10 |
| sar2_2 | training | 20 | 24 | -4 |
| sar2_12c | training | 20 | 21 | -1 |
| sar5_12e | training | 21 | 28 | -7 |
| sar1_17 | training | 22 | 26 | -4 |
| sar5_12f | training | 22 | 27 | -5 |
| sar4_13 | training | 22 | 24 | -2 |
| sar5_7a | training | 23 | 23 | 0 |
| sar1_23 | training | 24 | 22 | 2 |
| sar4_12 | training | 24 | 23 | 1 |
| sar5_12c | training | 25 | 31 | -6 |
| sar3_10 | training | 25 | 61 | -36 |
| sar4_5 | training | 26 | 26 | 0 |
| sar1_19 | training | 27 | 22 | 5 |
| sar2_12j | training | 27 | 27 | 0 |
| sar5_7i | training | 27 | 21 | 6 |
| sar2_12e | training | 27 | 31 | -4 |
| sar4_11 | training | 28 | 24 | 4 |
| sar5_12a | training | 29 | 23 | 6 |
| sar2_12a | training | 30 | 20 | 10 |
| sar2_1 | training | 30 | 38 | -8 |
| sar1_18 | training | 32 | 27 | 5 |
| sar5_7b | training | 32 | 25 | 7 |

| | | | | |
|---|---|---|---|---|
| sar4_4 | training | 32 | 33 | -1 |
| sar1_34 | training | 33 | 23 | 10 |
| sar3_25 | training | 34 | 31 | 3 |
| sar5_7h | training | 35 | 29 | 6 |
| sar2_12k | training | 35 | 22 | 13 |
| sar5_7e | training | 36 | 27 | 9 |
| sar5_12d | training | 37 | 37 | 0 |
| sar2_12i | training | 41 | 30 | 11 |
| sar3_13 | training | 42 | 41 | 1 |
| sar5_7c | training | 43 | 50 | -7 |
| sar5_7d | training | 47 | 36 | 11 |
| sar3_9 | training | 50 | 46 | 4 |
| sar2_12b | training | 50 | 23 | 27 |
| sar2_12f | training | 51 | 38 | 13 |
| sar3_16 | training | 59 | 68 | -9 |
| sar3_22 | training | 75 | 69 | 6 |
| sar3_15 | training | 86 | 71 | 15 |
| sar3_12 | training | 97 | 84 | 13 |
| sar1_35 | test | 14 | 18 | -4 |
| sar1_38 | test | 17 | 22 | -5 |
| sar1_5 | test | 18 | 18 | 0 |
| sar3_24 | test | 21 | 40 | -19 |
| sar5_12b | test | 23 | 28 | -5 |
| sar3_28 | test | 24 | 30 | -6 |
| sar5_12h | test | 26 | 32 | -6 |
| sar4_7 | test | 31 | 31 | 0 |
| sar3_7 | test | 33 | 29 | 4 |
| sar4_9 | test | 33 | 26 | 7 |
| sar3_19 | test | 47 | 41 | 6 |
| sar2_12d | test | 48 | 25 | 23 |
| sar3_20 | test | 97 | 49 | 48 |

CHAPTER 3

# IDENTIFICATION OF NOVEL PPARγ PARTIAL AGONISTS BY A VIRTUAL SCREENING OF NATURAL PRODUCTS

## ABSTRACT

We have developed a virtual screening procedure based on structure-based pharmacophore construction, protein-ligand docking and electrostatic/shape similarity to discover novel scaffolds of PPARγ partial agonists. From an initial set of 89,165 natural products and natural product derivatives, 135 compounds were identified as potential PPARγ partial agonists with good ADME properties. Ten compounds that represent ten new chemical scaffolds for PPARγ partial agonists were selected for *in vitro* biological testing. Five out of eight of these compounds were confirmed as PPARγ partial agonists: they bind to PPARγ, do not or only moderately stimulate the transactivation activity of PPARγ, do not induce adipogenesis of preadipocyte cells and stimulate the insulin-induced glucose uptake of adipocytes. These results demonstrate that our virtual screening procedure is able to find novel scaffolds for PPARγ partial agonists.

## Introduction

Peroxisome proliferator-activated receptors (PPARs) are members of the nuclear receptor superfamily that regulate the gene expression of proteins involved in energy, glucose and lipid metabolism, adipocyte proliferation and differentiation and insulin sensitivity [1]. PPARs act as cellular sensors that activate transcription in response to the binding of natural or synthetic ligands. Three subtypes, PPARα, PPARβ/δ and PPARγ, have been identified. Although the subtypes share a high level of sequence and structural homology [2], they exhibit differences in tissue expression and physiological function [3]. Agonists of PPARα and PPARγ are currently approved for treating dyslipidemia and type 2 diabetes, respectively [4,5]. Thiazolidinediones (TZDs) are one important class of synthetic agonists of PPARγ. TZDs are antidiabetic agents that target adipose tissue and improve insulin sensitivity, and they are currently being used in the treatment of type 2 diabetes. Despite the clinical benefit of TZDs, they have been associated with adverse side effects including weight gain, increased adipogenesis, renal fluid retention and a possible increased incidence of cardiovascular events [6-8]. Therefore, new PPARγ ligands with enhanced therapeutic efficacy and reduced adverse effects are needed. A promising new class of such ligands is that of the selective PPARγ modulators (i.e., SPPARγMs) [6-8]. These compounds act as partial agonists of PPARγ and display different binding properties than do full agonists [9].

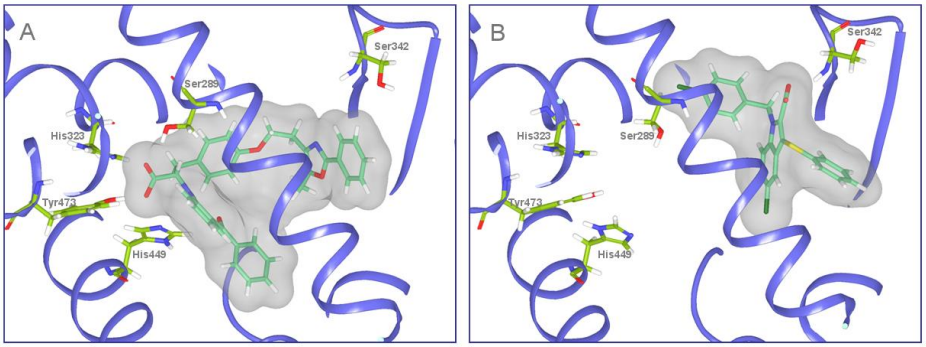



**Figure 1.** Binding models of (A) the PPARγ full agonist Farglitazar (crystal structure 1FM9) and (B) the PPARγ partial agonist nTZDpa (crystal structure 2Q5S). Important binding residues are depicted as wireframes with green carbon atoms. Oxygen, nitrogen, and hydrogen atoms are colored red, blue and white, respectively.

The mechanism of PPARγ activation by full agonists is mediated by a molecular switch of the H12 α-helix [10]. H12 forms part of the ligand-dependent activation domain, AF-2, that closes on the ligand-binding site in response to ligand binding. The resulting active form can bind to several co-activator proteins that activate the cellular

transcriptional machinery [10]. Full agonists occupy the large binding site of PPARγ in a U conformation and generally consist of a polar head and a hydrophobic tail [11]. The polar head makes a net of hydrogen bonds with the Ser289, His323, His449 and Tyr473 PPARγ side chains (Figure 1A), and this net is responsible for the conformational change of H12 and the activation of PPARγ. In contrast, partial agonists activate PPARγ by an H12-independent mechanism [12,13], and consequently, the key interactions between partial agonists and the ligand-binding domain (LBD) of PPARγ are different than those of the full agonists [9] (i.e., partial agonists do not bind to PPARγ by the net of hydrogen bonds used by full agonists). This causes a lower degree of H12 stabilization, which affects the recruitment of coactivators and, in turn, decreases the transcriptional activity of PPARγ [14,15]. With minor exceptions, most of the currently known partial agonists interact with the LBD of PPARγ through a hydrogen bond with Ser342 [11] and several hydrophobic interactions that are similar to those that occur with full agonist binding (Figure 1B). Recently, a new mechanism by which partial and full PPARγ agonists act to improve insulin sensitivity independent of receptor agonism has been suggested. This mechanism consists of blocking the phosphorylation of PPARγ at Ser 273 [16] and may explain how partial agonists can exhibit similar or higher antidiabetic effects than those of full agonists. This mechanism might also be the reason for the differing side-effect profiles of the two types of agonists [8]. It is possible that partial and full agonists achieve comparable efficacy in insulin sensitization through a similar inhibitory effect on PPARγ phosphorylation, whereas the differences in their agonistic potency could be linked to the differences in side effects [8].

Although there are successful examples of the discovery of new PPARγ agonists [14,17-20], including from natural origins [21-24], it has recently been of great interest to identify new PPARγ partial agonists from natural products [25,26]. Consequently, the goal of this work was to design and apply a virtual screening (VS) workflow to identify novel PPARγ partial agonists among natural products. To achieve this goal, we (a) designed a VS workflow that includes a filter to remove PPARγ full agonist candidates from the sample; (b) validated the performance of the VS with samples of known PPARγ agonists (either full or partial) and decoys; (c) applied the VS to a database of natural or derivatives of natural compounds; (d) clustered the VS hits with known PPARγ partial agonists; and (e) selected 10 different VS hits (from 10 clusters where no known PPARγ partial agonists were present) for testing their bioactivity as PPARγ partial agonists. Our results show that our VS workflow performs well and is able to discover new chemical scaffolds for the design of effective antidiabetics with fewer side effects than PPARγ full agonists.

## Results and Discussion

**Virtual Screening: Description, Validation and Application.** The VS workflow applied in this study is summarized in Figure 2. It consists of several steps that were applied one after another (i.e., the output molecules of one step were the input molecules for the next step). The discriminatory power of the VS workflow to identify PPARγ partial agonists was evaluated by applying it to a group of 135 known PPARγ full agonists (Supporting Information Table 1), 19 known PPARγ partial agonists (Supporting Information Table 2) and 3,122 decoys obtained from the DUD database [27].

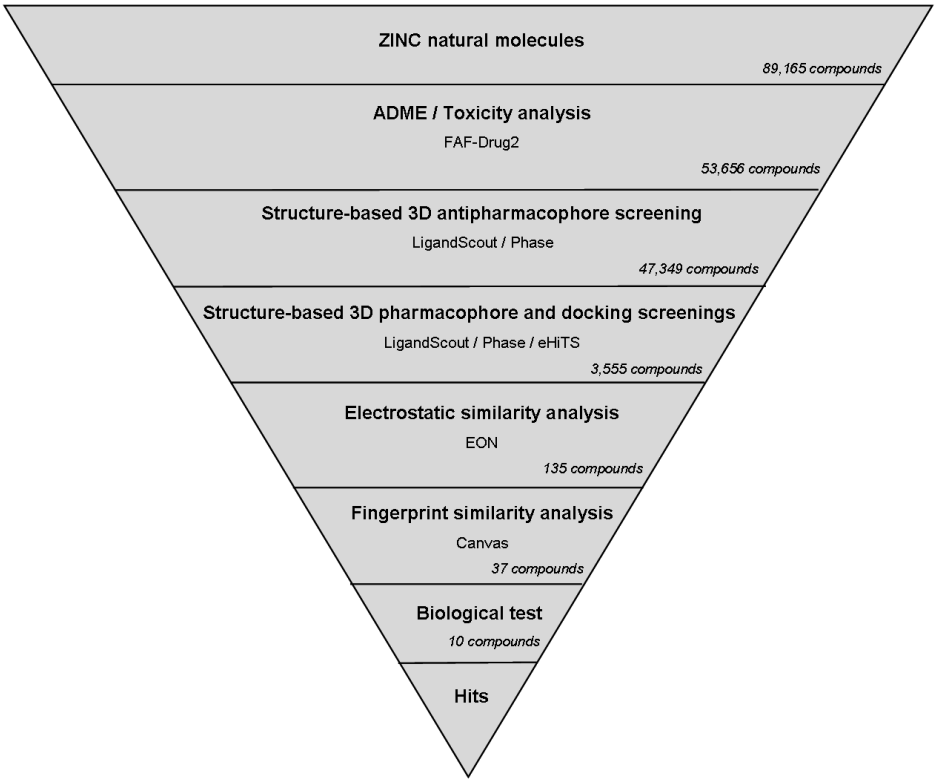



**Figure 2.** Schematic overview of the VS workflow and the procedure used for selecting the VS hits whose bioactivity was experimentally tested. The number of compounds that passed each step and the programs used are shown. From an initial set of 89,165 compounds, 135 compounds were identified as putative PPARγ partial agonists by the VS workflow. Ten of these 135 compounds were selected for *in vitro* testing.

**Table 1**. **Validation of each method used in the virtual screening workflow.** A dataset of 19 known PPARγ partial agonists, 135 known PPARγ full agonists and 3122 decoys extracted from the DUD database were used. The values represent the number of compounds from each set that survived each step when applied sequentially.

| Set of Compounds | Nº of Compounds | Structure-based pharmacophore screening | | | Electrostatic/shape similarity analysis | Global virtual screening |
|---|---|---|---|---|---|---|
| | | anti pharmacophore | partial agonist pharmacophore | | | |
| | | *in vacuo conformations* | *in vacuo conformations* | *docking poses* | | |
| Partial Agonists | 19 | 12 | 10 | 8 | 5 | 5 |
| Full Agonists | 135 | 31 | 11 | 7 | 1 | 1 |
| Decoys | 3122 | 2204 | 964 | 382 | 16 | 16 |
| Enrichment Factor | | 2.45 | 1.90 | 1.98 | 11.28 | 39.19 |
| $EF_{max}$ | | 24.27 | 187.25 | 98.50 | 49.63 | 172.42 |
| Sensitivity (Se) | | 77.04% | 83.33% | 80.00% | 62.50% | 26.32% |
| Specificity (Sp) | | 29.45% | 56.38% | 60.10% | 95.63% | 99.49% |

Table 1 shows how many of these molecules survived each VS step and several quantitative measures for model quality. Because we were interested in discovering novel PPARγ partial agonists but not full agonists, we first developed a structure-based pharmacophore, called the antipharmacophore, to exclude possible full agonists. We used this strategy because full agonists have more clearly defined structural features than partial agonists. Using 19 validated crystal structures from known full agonists complexed with PPARγ, we created the antipharmacophore model that represented the common features of full agonists. This antipharmacophore consisted of 5 sites (Figure 3A): 2 of them are involved in a hydrogen bond network between the ligand and the receptor, and 3 are hydrophobic sites. The PPARγ residues that interact with the two sites involved in the hydrogen bond network are Ser289, Tyr473, His323 and His449. Because the hydroxyl group from serine and tyrosine and the nitrogen from the histidine side chain can act as donors and acceptors simultaneously, the two sites involved in the hydrogen bond network were defined as having possible dual behavior as a hydrogen bond donor and acceptor. Both sites were considered to be essential. Two out of the three hydrophobic sites were also mandatory, whereas the site located at the effector end (side HF3 in Figure 3A) was defined as optional because it regulates the affinity and potency of ligands11. Table 1 shows that 104 out of the 135 PPARγ full agonists used in the validation process were identified as full agonists by our antipharmacophore model, as were 918 out of 3122 decoys (29%) and

7 out of the 19 partial agonists (37%). This represents an enrichment factor (EF) of 2.45, or 10% of the EF maximum value of 24.3 that would be obtained if all 135 full agonists were identified as positive hits in this analysis. Importantly, this is the only VS step for which the active set was composed of PPARγ full agonists and the inactive set was decoys and PPARγ partial agonists; therefore, the statistics for this VS step were calculated using these considerations. The sensitivity (Se) and the specificity (Sp) of this step were 77.04% and 29.45%, respectively. The high percentage of partial agonists misidentified as full agonists shows that it is sometimes difficult to distinguish between both sets. This was confirmed when some partial agonists were similar to full agonists and clustered together using a fingerprint similarity analysis (results not shown). However, because the aim of the antipharmacophore step was to minimize the presence of full agonists, loss of some possible partial agonists at this step was tolerated.

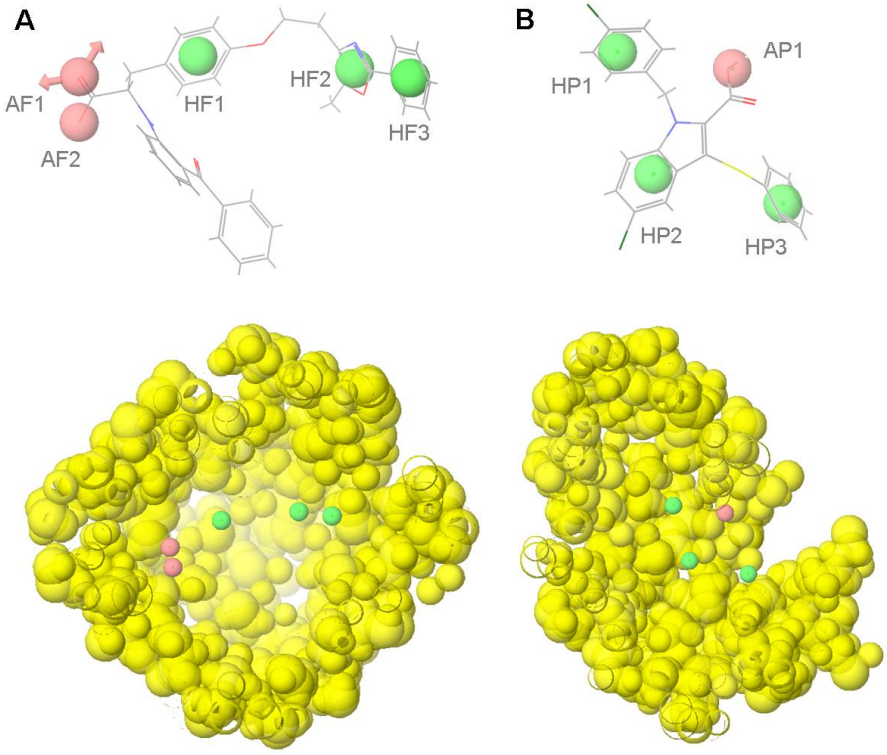



**Figure 3.** Pharmacophores used for the identification of (a) PPARγ full agonists and (B) PPARγ partial agonists. Hydrophobic and acceptor/donor sites are colored in green and pink, respectively. Excluded volumes are showed as yellow spheres. The ligands farglitazar (from the PDB entry 1FM9) and nTZDpa (from the PDB entry 2Q5S) are also represented.

Next, a common PPARγ partial agonist pharmacophore, derived from 12 structures of PPARγ crystallized with a partial agonist, was applied. This pharmacophore consisted of one hydrogen bond acceptor, located on a carboxylic group of the ligand that forms a hydrogen bond with Ser342, and three hydrophobic sites located on aromatic rings (Figure 3B). From the 12 partial agonists that survived the antipharmacophore step, 10 were identified as partial agonists by our partial agonist pharmacophore, as were 11 of the 31 full agonists and 964 of the 2204 decoys. Therefore, the EF for this VS step was 1.90, which represents 1.02% of the maximum EF of 187.25 obtained if all 12 partial agonists would have been identified as true positives in this step. The Se and Sp for this step were 83.33% and 56.38%, respectively (Table 1).

To find docking poses that were compatible with the partial agonist pharmacophore, the compounds that had at least one conformer, generated in vacuo, that matched the partial agonist pharmacophore were also docked to the PPARγ structure from 2Q5S. The best docking poses were then matched again to the partial agonist pharmacophore. We found that 8 out of 10 partial agonists, 7 out of 11 full agonists and 382 out of 964 decoys that survived the previous step have at least one docked pose that both was compatible with the PPARγ ligand-binding site and had functional groups that match the 3D location of the sites of the partial agonist pharmacophore. The EF, Se and Sp for this step were 1.98, 80.00% and 60.10%, respectively (Table 1).

To reduce the number of PPARγ partial agonist candidates, an electrostatic and shape similarity analysis was applied. Using the experimental poses of 5 known PPARγ partial agonists as queries, 5 out of 8 partial agonists, 1 out of 7 full agonists and 16 out 382 decoys were identified as partial agonist candidates by this VS step. The EF, Se and Sp for this step were 11.28 (out of an EF maximum of 49.63), 62.50% and 95.63%, respectively (see Table 1).

Table 1 shows that the full VS workflow identified 5 out of 19 partial agonists, 1 out of 135 full agonists and 16 out of 3122 decoys as partial agonists. Therefore, the global EF was 39.19 (22.73% of an EF maximum of 172.42) and the Se and Sp were 26.32% and 99.49%, respectively. The high Sp and the moderate Se of our procedure reflect, respectively, the correct assignment of inactive compounds and the loss of potential partial agonists. However, because of the high number of initial compounds and the difficulties in differentiating partial from full agonists, we preferred a very specific, but less sensible, VS workflow. Table 1 also shows that in terms of sensitivity, using the partial agonist pharmacophores was the best step, whereas in terms of specificity and EF, the best step was the electrostatic/shape similarity analysis. Therefore, the combination of the three steps seems adequate to obtain a VS workflow that combines the best of each method. Importantly, the Se and Sp of the

antipharmacophore step should not be compared with those for the other VS steps because the objective of the antipharmacophore step was to remove full agonists from the sample. In that sense, despite the low Sp for this step, its high Se (77.04%) suggests that it is adequate for this purpose.

Upon validation of the VS workflow, it was applied to the Natural Products subset of the ZINC database28. From an initial set of 89,165 molecules, compounds with poor ADME properties or potentially toxic compounds were discarded, resulting in an initial set of 53,656 molecules. After applying the VS workflow described above, a group of 135 PPARγ partial agonist candidates were finally identified. Figure 2 shows the number of molecules that survived each step of the VS workflow.

**Fingerprint similarity analysis.** To reduce the number of hits for biological testing while simultaneously increasing the significance of the results (i.e., by obtaining new chemical scaffolds for PPARγ partial agonists), a fingerprint cluster analysis was done. The 135 partial agonist candidates from the VS were combined with a group of 19 known partial agonists (Supporting Information Table 2), and their 2D fingerprints were calculated. A hierarchical cluster analysis classified the compounds into 51 clusters, and 37 of them did not contain any already known partial agonists and therefore represented new chemical scaffolds for PPARγ partial agonists.

**Biological testing of selected VS hits.** Figure 4 shows the chemical structures of the ten compounds (C1-C10) selected for bioactivity testing. They were selected from 10 of the 37 clusters that corresponded to new chemical scaffolds of PPARγ partial agonists, taking into account as a selection criteria their purity, price and availability. It is likely that at least some of the problematic side effects of PPARγ full agonists, such as weight gain or fluid retention, may be caused by classical agonist interactions. A substantial portion of the therapeutic benefits of full and partial PPARγ agonists occurs through the inhibition of the PPARγ phosphorylation at Ser273 [16]. Thus, an effective partial agonist of PPARγ would have weak or low transactivation activity while maintaining the stimulation of glucose uptake [16]. In this sense, the ten selected compounds were analyzed in vitro to check whether they bind to PPARγ, activate the transactivation activity of PPARγ, and stimulate glucose uptake and differentiation in adipocytes. Prior to these analyses, cytotoxicity and viability tests were also performed. None of the tested compounds (C1-10) at the concentrations analyzed significantly decreased the viability or increased the cytotoxicity in HepG2 and 3T3-L1 cells (results not shown).

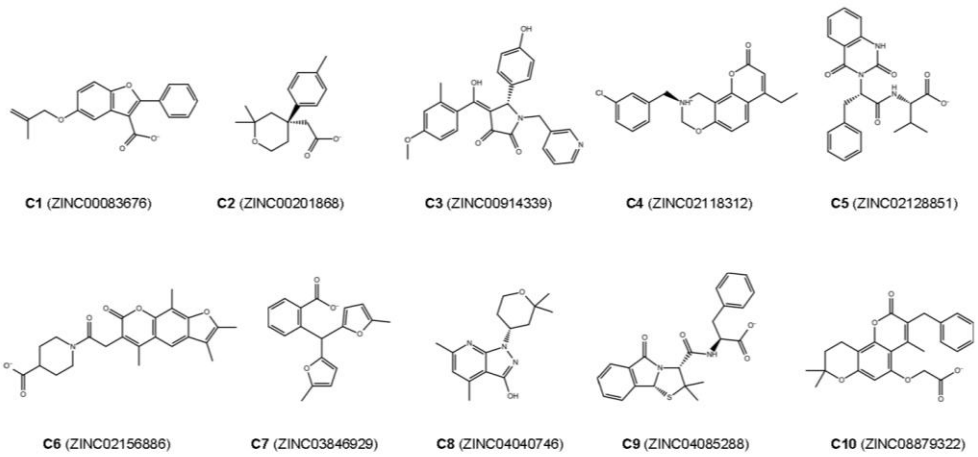



**Figure 4.** Chemical structures and ZINC codes of the 10 compounds suggested to be PPARγ partial agonists and selected for bioactivity testing.

**Table 2. Experimental $IC_{50}$ values and transactivation activity of the selected compounds.** A competitive binding assay was used to assess the ability of experimental compounds, FMOC or rosiglitazone to displace a fluorescent PPARγ ligand from a human-derived recombinant PPARγ ligand-binding domain. The concentration of the test compound that results in a half-maximal shift in the polarization value is defined as $IC_{50}$. This value is a measure of the relative affinity of the test compound for the PPAR ligand-binding domain. The transactivation capacity of selected compounds was also determined in HepG2 cells as described in Materials and Methods. Results represent the mean ± SEM of at least three separate experiments performed in triplicate. Results are expressed as arbitrary firefly luciferase units relative to arbitrary renilla luciferase units.

| Compound | Binding affinity $IC_{50}$ (μM) | Gene Reporter activity at 10 μM mean ± SEM | Gene Reporter activity at 100 μM mean ± SEM |
|---|---|---|---|
| Rosiglitazone | 0.39 | 6.827 ± 2.367** (i) | - |
| FMOC | 18.5 | 3.387 ± 0.422** | - |
| C1 | 12.4 | 1.767 ± 0.075 | 2.607 ± 0.323* |
| C5 | 1000 | 1.713 ± 0.231 | 1.990 ± 0.276 |
| C7 | 252 | 1.710 ± 0.078 | 3.437 ± 0.219** |
| C8 | 460 | 1.700 ± 0.361 | 2.127 ± 0.167 |
| C9 | 585 | 0.650 ± 0.052 | 1.043 ± 0.052 |

(i) The transactivation activity of Rosiglitazone was assayed at 1 μM. For compounds C2, C3 and C6, no binding was observed when assayed at concentrations up to 8 mM. Compounds C4 and C10 were not assayed because of solubility problems.

** $p<0.001$ * $p<0.05$ when compared to the control (DMSO) in an Anova test.

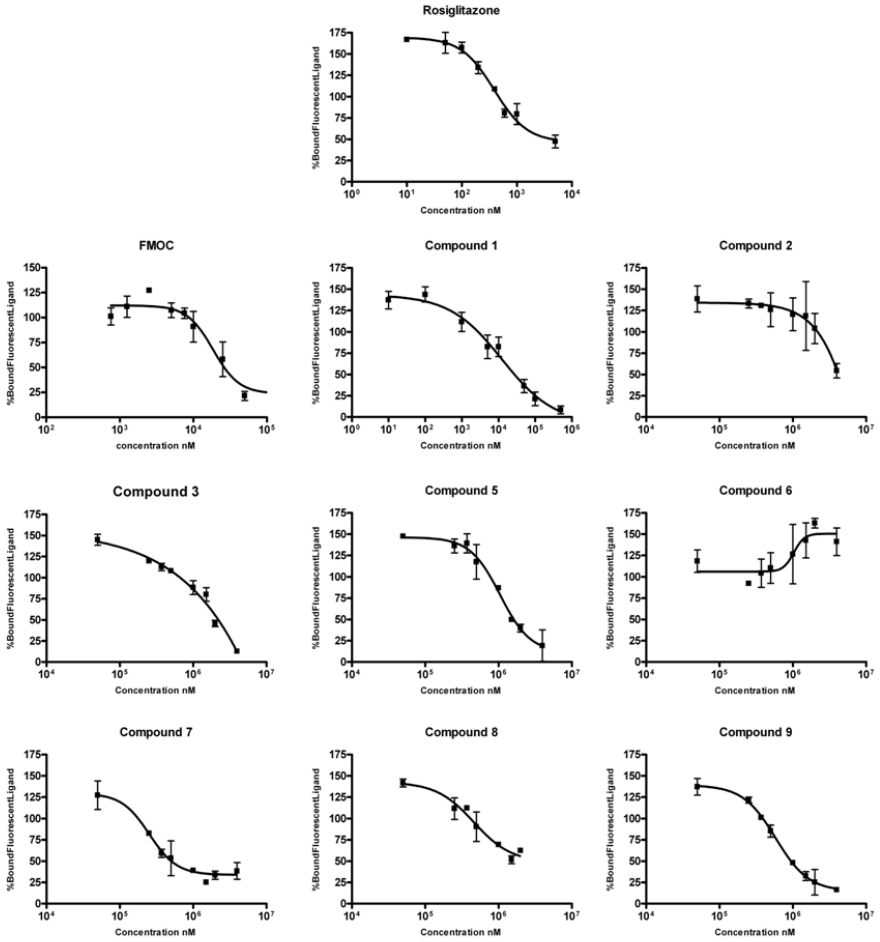



**Figure 5.** Results of the PolarScreen PPAR Competitor Assay for all selected compounds, except C4 and C10, which had solubility problems. One PPARγ full agonist (Rosiglitazone) and one partial agonist (FMOC) were also assayed. Error bars represent one standard deviation from the mean of triplicates.

Figure 5 shows the results of the PolarScreen PPARγ Competitor Assay to determine the binding affinity of the selected compounds. Compounds C4 and C10 were not assayed due to solubility problems. Compounds C1, C5, C7, C8 and C9 bound to PPARγ with different affinities (Table 2). C1 had a moderate binding affinity for PPARγ, similar to that of the known PPARγ partial agonist FMOC. Compounds C5, C7, C8 and C9 had lower binding affinities. The results for compounds C2, C3 and C6 were not conclusive when assayed at concentrations up to 8 mM. These results validate the predictions of the VS procedure, as five out of eight of the assayed

compounds were able to bind PPARγ. Table 2 also shows the transcription activity of the five compounds that were able to bind PPARγ and were potential PPARγ partial agonists (C1, C5, C7, C8 and C9). Only compounds C1 and C7 had moderate and significant transactivation activity at 100 μM, similar to that shown by FMOC (see Table 2). These results validate the antipharmacophore step, as none of the assayed compounds acted as a PPARγ full agonist.

Table 3 shows the in vitro effects of these five partial agonists (C1, C5, C7, C8 and C9) on the adipogenic activity and the stimulation of glucose uptake in adipocytes. As expected from the results of the PPARγ transactivation activity assay, none of the assayed compounds induced triglyceride accumulation in 3T3-L1 preadipocytes when compared with Rosiglitazone, a well-known PPARγ full agonist that stimulates adipogenesis. Table 3 also shows the results of the effects on insulin-induced glucose uptake. All of the selected compounds stimulated insulin-induced glucose uptake to the same extent or even more than Rosiglitazone and FMOC, with compounds C7, C8, and especially C5 being the most effective. Together, these results show that these five compounds can be considered to be partial agonists of PPARγ and validate the virtual screening protocol developed.

**Table 3. Adipogenic activity and stimulation of glucose uptake by some of the selected compounds measured *in vitro*.** The compounds that bind PPARγ were added to 3T3-L1 pre-adipocytes to test their adipogenic capacity, measured as triglyceride accumulation normalized to the effects of rosiglitazone. The same compounds were added to fully differentiated adipocytes to test their effects in insulin-stimulated 2-deoxy-[$H^3$]-glucose uptake (insulin stimulation is considered 100%). Data are mean ± SEM of at least three biological replicates

| Compound | Adipogenic activity | Glucose uptake stimulation |
|---|---|---|
| Rosiglitazone | 1.000 ± 0.051 | 118.99 ± 8.543 |
| FMOC | 0.530 ± 0.062** | 124.64 ± 7.295 |
| C1 | 0.494 ± 0.052** | 120.9 ± 13.561# |
| C5 | 0.492 ± 0.036** | 140.48 ± 17.385## |
| C7 | 0.519 ± 0.021** | 133.54 ± 13.508## |
| C8 | 0.562 ± 0.071** | 120.24 ± 6.680## |
| C9 | 0.641 ± 0.025** | 120.90 ± 20.410 |

**P< 0.05 VS. Rosiglitazone

## P< 0.05 # P< 0.1 VS. Vehicle

**Docking of novel PPARγ ligands.** To determine the putative binding mode and the potential ligand-target interactions of the five novel PPARγ partial agonists (C1, C5, C7, C8 and C9), these compounds were docked to the PPARγ LBD of PDB entry 2Q5S. Similar docking poses were determined for all sets of compounds (see Figure 6). The predicted binding modes of all compounds (with the exception of C8) included

one hydrogen bond with Ser342 or adjacent residues and several hydrophobic contacts with Ile281, Ala292, Ile326, Ile341, Leu330, Leu333, Val339, Met348, Leu353 or Met364 from arms II and III of the LBD of PPARγ. These interactions are typical of PPARγ partial agonists9. In addition, no hydrogen bond interaction between the five compounds and residues His323, Tyr327, His449 and Tyr473 from arm I of the LBD of PPARγ (typical of PPARγ full agonists) was predicted. This could explain the lack of (or moderate) transactivation activity determined for the five compounds.

The best docking pose for compound C5 shows that this compound could establish three hydrogen bonds with arms II and III of the LBD of PPARγ (Figure 6B). Two of these hydrogen bonds include those formed between the carboxylic moiety of the compound and the backbones of Ser342 and Glu343. Another hydrogen bond could be established between the nitrogen of the quinazoline-2,4-dione moiety and the side chain of Ile281 (Figure 6B). In addition, hydrophobic interactions were predicted between the rest of the ligand and residues Ile249, Leu255, Arg288, Met348, Val339, Ile341, and Met364 from arms II and III. The initial docking of compound C7 to the LBD of PPARγ showed that this compound could establish two hydrogen bonds with Ser342 and Glu343 (Figure 6C). The small size of this compound may allow the binding of a second ligand molecule. This 2:1 binding stoichiometry has been described or predicted for other PPARγ partial agonists [25,29]. Based on this, we performed a docking study to investigate the possibility that two copies of compound C7 could bind simultaneously to PPARγ. Figure 6F shows that an additional copy of the C7 compound could interact with arms I and III through several hydrophobic interactions. The experimental IC50 and transactivation activity of this compound also suggest this possibility. The plot of the PolarScreen PPAR competitor assay for compound C7 in Figure 5 shows that the full binding is delayed until high concentrations are reached. In the same way, we did not observe significant transactivation activity with 10 μM of this compound, but this activity increases significantly with 100 μM (see Table 2). These observations agree with a model in which one molecule of compound C7 binds to arms II and III of the LBD of PPARγ, and when the concentration increases, a second molecule occupies arm I and causes transactivation activity.

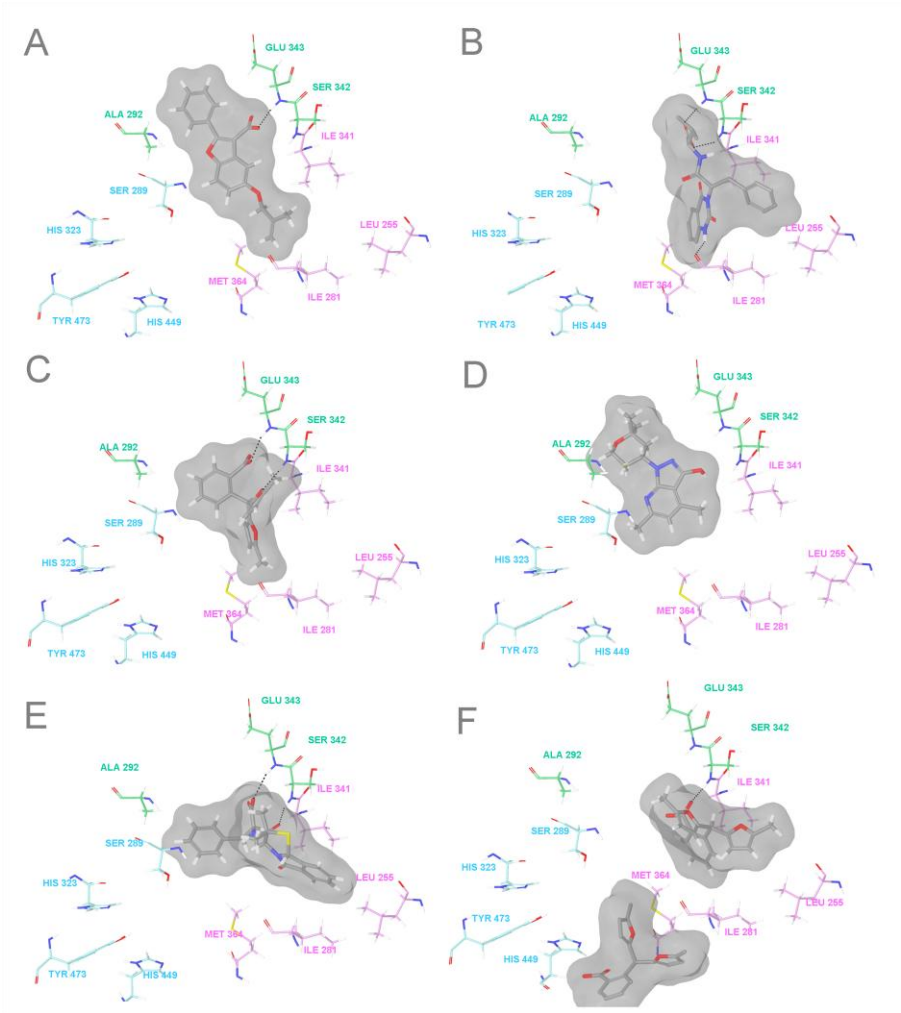



**Figure 6.** Putative ligand-PPARγ interactions of the best docking poses of compounds (A) C1, (B) C5, (C) and (F) C7, (D) C8 and (E) C9. The following key residues of the LBD of PPARγ are shown: Ser289, His 323, Tyr473 and His449 from arm I are colored in blue; Leu255, Ile281, Ile341 and Met364 from arm II are colored in pink; and Ala292, Ser342 and Glu343 from arm III are colored in green.

## Conclusions

We have shown that a VS workflow based on two structure-based 3D pharmacophores (one to exclude potential PPARγ full agonists), protein-ligand docking and electrostatic/shape similarity analysis is able to discover novel scaffolds for PPARγ partial agonists. Thus, from an initial set of 89,165 natural products and natural product derivatives, 135 compounds were defined as potential PPARγ partial

agonists. Using a fingerprint similarity analysis, 37 clusters that represent new chemical scaffolds for PPARγ partial agonists were defined. Ten compounds from ten of these clusters were chosen for bioactivity testing, but two of them were not assayed because of solubility problems. Five out of the remaining eight compounds can be considered as PPARγ partial agonists because they were able to bind PPARγ with a moderate affinity, did not stimulate adipogenesis and enhanced insulin-stimulated glucose uptake in vitro. Therefore, our results suggest that our VS workflow is able to identify compounds with a high chance of being effective PPARγ partial agonists in a molecule database and that this bioactivity is not trivial because their chemical structure does not resemble known PPARγ partial agonists. In addition, our data show that C5 is an appropriate compound for lead-optimization and the subsequent design of more potent and safe antidiabetic drugs.

## Experimental Section

**Dataset of PPARγ Structures Used.** Forty-nine structures of the PPARγ LBD co-crystallized with an agonist were downloaded from the RCBS Protein Data Bank (http://www.pdb.org) [30]. For each structure, we determined whether electron density maps were available at the Uppsala Electron Density Server (http://eds.bmc.uu.se/eds/) [31] and, if available, the goodness-of-fit between these maps and the structures of both the ligand and the PPARγ active site. After this preliminary analysis, 18 out of the 49 PDB complexes were not further considered in our study due to one of following reasons: (a) the electron density maps were not available; (b) either the ligand or the PPARγ active site did not fit well on the electron density maps; (c) the ligand was a fatty acid; or (d) the ligand could not be identified as either a full or partial agonist. The remaining 31 PDB complexes (Table 4) were superposed with the DeepView v3.7 program [32] to arrange them in the same relative orientation. Only the resulting re-oriented coordinates of the PDB complexes were used in the subsequent steps of the workflow.

**Table 4.** PDB codes of the ligand-protein complexes used for the generation of the structure-based pharmacophore models for PPARγ full agonists and PPARγ partial agonists.

| Full agonists | | | | Partial agonists | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | cluster 1 | cluster 2 | cluster 3 | cluster 4 | cluster 5 |
| 1FM9 | 1I7I | 1FM6 | 2GTK | 2G0G | 4PRG | 2Q6R | 2FVJ | 2Q6S |
| 1RDT | 1KNU | 1ZGY | 3B3K | 2G0H | | 2Q61 | | 2WM0 |
| 1K74 | 2F4B | 2PRG | 2ATH | | | 2Q5S | | |
| 3BC5 | 2HWQ | 2FVJ | 1NYX | | | 2Q5P | | |
| 2Q8S | 2HWR | 1ZEO | | | | 2HFP | | |
| | | | | | | 2P4Y | | |

**Generation of Structure-Based Pharmacophores.** LigandScout v2.03 (Inte:ligand, Vienna, Austria, http://www.inteligand.com/ligandscout/) [33,34] was used for the analysis of the 31 PPARγ structures from Table 4 and the analysis of the possible interactions between the crystallized ligands and the ligand-binding pocket of PPARγ. Individual pharmacophores for the 19 structures of PPARγ crystallized with a full agonist (Table 4) were visually inspected to construct a common structure-based pharmacophore of full agonists. This pharmacophore (Figure 3A) is formed by 5 sites (two hydrogen-bond acceptors and three hydrophobic sites) that are present in most of the complexes of full agonists analyzed and are therefore assumed to be responsible for the intermolecular interactions that are essential for the activity of PPARγ full agonists. We named this pharmacophore the antipharmacophore because we used it to exclude putative full agonists when searching for partial agonists. Taking into account ligand similarity, we classified the remaining 12 structures of PPARγ crystallized with a partial agonist into 5 clusters or families (Table 4). For each cluster, a common structure-based pharmacophore for PPARγ partial agonists was defined. The resulting pharmacophores contained 5 to 8 sites, mainly hydrophobic sites and some hydrogen-bond acceptors. A common pharmacophore of 4 sites was then constructed. This pharmacophore consisted of one hydrogen-bond acceptor site and three hydrophobic sites (Figure 3B). This common pharmacophore contained the sites in common with the pharmacophores of each cluster and, in our opinion, the sites that are important for the intermolecular interaction between PPARγ and its partial agonists. This common pharmacophore, which we called the partial agonist pharmacophore, was used in the VS workflow to identify putative PPARγ partial agonists.

Both pharmacophores were also completed with receptor-based excluded volumes, obtained either from 1FM9 for the antipharmacophore or from 2Q5S for the partial agonist pharmacophore, that schematically represent the location of the PPARγ residues that form the LBD (Figures 3C and 3D). Excluded volumes were added by applying the Receptor-Based Excluded Volumes graphic front-end from Phase v3.1 (Schrödinger LLC., Portland, USA; http://www.schrodinger.com) [35] and by setting the Sphere filter parameter values to (a) ignoring receptor atoms whose surfaces were within 0.25 Å of ligand surface and (b) limiting excluded volume shell thickness to 10 Å. The rest of the parameters used were the default values.

**Initial Dataset of Natural Compounds.** The initial dataset of the natural compounds that we used contained 89,165 compounds from the Natural Products subset of the ZINC database [28]. This dataset contains commercially available natural products and natural-product derivatives. The 3D structures of this initial dataset were processed with the LigPrep v2.3 program (Schrödinger LLC., Portland, USA; http://www.schrodinger.com) to clean them and obtain their corresponding low-energy structures. Only one low energy conformation was generated for each molecule. This process was carried out with the following parameter values: (a) the force field used

was OPLS 2005; (b) all possible ionization states at pH 7.0 ± 2.0 were generated with Ionizer; (c) the desalt option was activated; (d) tautomers were generated for all ionization states at pH 7.0 ± 2.0; (e) chiralities, when present, were determined from the 3D structure; and (f) one low-energy ring conformation per ligand was generated. When chirality was not defined, a maximum number of 32 stereoisomers were generated. Conformations were built with the Phase program, generating in vacuo a maximum number of 200 conformers per structure and using the default Phase options.

**Virtual Screening Workflow.** Briefly, the VS workflow consisted of several steps that must be applied one after another (i.e., the output molecules of one step were the input molecules for the next step). Thus, the filters applied and sorted according their usage were: (1) an ADME/Toxicity prediction; (2) a structure-based antipharmacophore screening for removing PPARγ full agonist candidates; (3) a structure-based pharmacophore screening; and (4) an electrostatic/shape similarity analysis.

The initial set of compounds was submitted to an ADME/Tox filter with the FAF-Drugs2 program [36]. The aim of this step was to discard those molecules that could have poor ADME properties or were potentially toxic. Thus, the drug-like properties of a compound were evaluated by means of the Lipinski rule [37], and only one violation of the rule was allowed. This rule is based on a set of property values (i.e., the number of hydrogen bond donors and acceptors, the molecular weight and the logP) that were derived from a large number of drugs with good ADME characteristics [37]. Hence, molecules that pass the Lipinski rule are expected to be orally active in humans. Moreover, molecules containing toxic groups were filtered by using the 204 substructures for "warhead" chelators, frequent hitters, promiscuous inhibitors and other undesirable functional groups available in the FAF-Drugs2 [36] tool.

Molecules with appropriate ADME/Tox properties were then filtered by a structure-based antipharmacophore with the aim of discarding potential PPAR□ full agonists. This filter removed from the sample those molecules that had at least one in vacuo-generated conformer that matched at least 4 out of 5 sites of the antipharmacophore. The fitting between the molecules and the pharmacophore was analyzed with the Phase program [33], using a site-matching tolerance of 2 Å for acceptor and donor sites and 2.5 Å for hydrophobic, aromatic and negative sites and applying the excluded volumes previously generated. The subset of molecules that did not match the antipharmacophore was then used to identify possible partial agonists. To do this, a second pharmacophore obtained from the common sites of known PPARγ partial agonists was used. Equivalent conditions were used for the pharmacophore-based searches. Molecules that had at least one in vacuo-generated

conformer and matched the 4 sites of the partial agonist pharmacophore were initially identified as putative PPARγ partial agonists.

To find docking poses that were compatible with the partial agonist pharmacophore, the molecules identified as putative PPARγ partial agonists were docked to the ligand-binding site of 2Q5S. The 32 best docked poses predicted by the eHiTS v2009 program (SimBioSys Inc., Toronto, Canada; http://www.simbiosys.ca/ehits) [38] were filtered again with Phase through the partial agonist pharmacophore, using the same filtering options of the first pharmacophore matching, with the exception that now re-orientation of the poses was not allowed during the search (*i.e.*, the score in place option was used).

The poses that passed the pharmacophore and docking screenings were submitted to an electrostatic/shape similarity analysis, using the experimental poses of the PPARγ partial agonists crystallized at the structures 2G0H, 4PRG, 2Q5S, 2FVJ and 2Q6S as queries. These five partial agonists are representative of each of the five clusters of PPARγ partial agonists defined in Table 4. This analysis was done with EON v2.0.1 (OpenEye Scientific Software, Inc., Santa Fe, New Mexico, USA; http://www.eyesopen.com) using the Electrostatic Tanimoto combo (ET_combo) score as similarity criteria. The ET_combo score is the sum of two calculations: the Shape Tanimoto (ST) score, which is a quantitative measure of three-dimensional overlap, where 1 corresponds to a perfect overlap, i.e., same shape, and the Possion-Boltzman Electrostatic Tanimoto (ET_pb) score that compares the electrostatic potential of two small molecules and ranges from 1 (identical potential) to negative values that result from the overlap of positive and negative charges. Molecules with an ET_pb score greater than 0.3 and an ST score greater than 0.5 were predicted to be potential PPARγ partial agonists by the VS workflow.

**Virtual Screening Workflow Validation**. The ability of the VS workflow to identify PPARγ partial agonists was tested by applying it to a group of 135 known PPARγ full agonists (Supporting Information Table 1), 19 known PPARγ partial agonists (Supporting Information Table 2) and 3,122 decoys obtained from the DUD database [27]. The structures of the 135 full agonists and the 19 partial agonists were built with ChemDraw Ultra v11.0 (CambridgeSoft Corporation, Cambridge, MA, USA; http://www.cambridgesoft.com/) [39] and cleaned using LigPrep v2.3 (Schrödinger LLC., Portland, USA; http://www.schrodinger.com). For each step of the VS workflow, an enrichment factor (EF) and a value for sensitivity (Se) and specificity (Sp) were calculated [40]. The EF was defined as the quotient of the fraction of active compounds in the sample that survived a particular VS step and the fraction of active compounds that were in the sample before applying this step. Therefore, the EF represents the ratio of the number of active compounds actually retrieved by a method compared with the number expected purely by chance. The

maximum EF value (EFmax) at each step was also estimated assuming only the known active compounds would survive at each step. Sensitivity (Se) describes how well the model correctly identifies active compounds, and it is calculated as the ratio between the number of active compounds that survived a particular VS step and the number of all active compounds that were in the sample before applying the VS step. Specificity (Sp) measures the correct assignment of inactive compounds, and it is calculated as the ratio between the number of inactive compounds that were discarded at a particular VS step and the number of all inactive molecules that were in the sample before applying the VS step. For the estimation of EF, EFmax, Se and Sp at the antipharmacophore step, the full agonist set was considered to be the set of active compounds. For the rest of the steps, the set of partial agonists was considered to be the active compounds. Global EF, EFmax, Se and Sp values for the entire VS process were also calculated using the number of active or inactive compounds that survived the entire VS workflow and the initial number of compounds before applying the VS procedure.

**Structural Similarity Analysis.** To select a representative dataset of VS hits for testing their bioactivity, the molecules that survived the electrostatic/shape similarity filter were merged with a group of 19 known partial agonists (Supporting Information Table 2) and clustered with Canvas v1.2 (Schrödinger LLC., Portland, USA; http://www.schrodinger.com). MOLPRINT2D fingerprints [41] with a fingerprint precision of 32 bits were calculated for each molecule, and then a hierarchical clustering based on Tanimoto similarities was obtained. The number of clusters was defined using the Kelley criterion [42]. Clusters that did not contain any known partial agonists were defined as clusters with new scaffolds for PPARγ partial agonists, and ten molecules from ten different clusters were selected for further bioactivity tests.

**Reagents and Materials for the Biological Tests.** The 10 selected compounds [ZINC00083676 (C1), ZINC00201868 (C2), ZINC00914339 (C3), ZINC02118312 (C4), ZINC02128851 (C5), ZINC02156886 (C6), ZINC03846929 (C7), ZINC04040746 (C8), ZINC04085288 (C9), and ZINC08879322 (C10)] were purchased from InterBioScreen Ltd. (Moscow, Russia). Their purities were higher than 92% or 95%. FMOC-L-Leucine (FMOC) was purchased from Calbiochem (Merck, Darmstadt, Germany). Rosiglitazone (BRL) was kindly provided by GlaxoSmithKline (Middlesex, UK). The test compounds were dissolved in DMSO, aliquoted and kept frozen until use. Cell culture reagents were obtained from BioWhittaker (Verviers, Belgium). Bradford protein reagent was obtained from Bio-Rad Laboratories (Life Sciences Group, Hercules, CA, USA). Insulin (Actrapid) was from Novo Nordisk (Bagsvaerd, Denmark). 2-deoxy-[H3]-glucose and ECL detection reagent were from Amersham Biosciences (Buckinghamshire, England).

**Polarscreen PPARγ Competitive Assay.** The PPAR ligand-binding competitive assay was performed with the PolarScreenTM PPARγ Competitor Assay Green according to the manufacturer's protocol. Briefly, the PPARγ LBD and the fluorescent PPARγ ligand form a complex with a high polarization value. Displacement of the fluorescent ligand by PPARγ ligands frees the fluormone in solution to tumble rapidly during its fluorescence lifetime, causing a low polarization value. The change in polarization value was used to determine the relative affinity of test compounds for the PPARγ LBD. Fluorescence polarization was measured using a POLARstar omega plate reader (BMG Labtech, Germany) at an excitation wavelength of 485 nm and an emission wavelength of 535 nm. Rosiglitazone, a compound with high affinity for PPARγ, was used as a positive control. Polarization values were plotted against the concentration of the test compound. To discard non-specific effects, DMSO was also tested at equivalent concentrations. The concentration of the test compound that resulted in a half-maximal shift in polarization value was defined as IC50. This value is a measure of the relative affinity of the test compound for the PPAR LBD. Curve fitting was performed using GraphPad Prism v4.0 (GraphPad Software, San Diego CA, USA; http://www.graphpad.com) following the program instructions.

**Dual-Luciferase Reporter Assay.** The activity of overexpressed PPARγ in response to its agonists was assessed in HepG2 cells using a PPARγ reporter (SABiosciences CCS-3026L). The PPAR reporter is a mixture of a PPAR-responsive luciferase construct and a constitutively expressed Renilla construct (40:1). The PPAR-responsive luciferase construct encodes the firefly luciferase reporter gene under the control of a minimal (m)CMV promoter and tandem repeats of the PPAR transcriptional response element. This construct monitors both increases and decreases in the transcriptional activity of PPAR. The constitutively expressed Renilla construct encodes the Renilla luciferase reporter gene under the control of a CMV immediately early enhancer/promoter and acts as an internal control for normalizing transfection efficiency and monitoring cell viability. Cells were co-transfected with the PPAR reporter and negative control along with the PPARγ expression vector in a 96-well plate. After 24 hours of transfection, cells were treated with the total agonist rosiglitazone (1 μM), partial agonist FMOC (10 μM) or the selected experimental compounds (10 and 100 μM). The dual-luciferase assay was performed with the Biotek FLx800 Multi-Detection Microplate Reader using the Promega dual luciferase reporter kit (E1910). Promoter activity values are expressed as arbitrary units using a Renilla reporter for internal normalization. Experiments were done in at least triplicate, and results represent the relative luciferase activity normalized to the untreated control. Statistical analysis was carried out by one-way analysis of variance (ANOVA) with Dunnett's post-hoc test using GraphPad Prism v4.0. Differences were considered significant when $P < 0.05$* or $P < 0.001$**.

**Cytotoxicity and Viability Assays of the Experimental Compounds in HepG2 Cells.** HepG2 cytotoxicity induced by the tested compounds was assessed by lactate dehydrogenase (LDH) leakage into the culture medium. Following a 24-h exposure to compounds C1-C10 (10 and 100 μM), the culture medium was aspirated and centrifuged at 3000 rpm for 5 min to obtain a cell-free supernatant. The activity of LDH in the medium was determined using a commercially available kit from QCA (Amposta, Spain). Aliquots of media and warm reagent were mixed in a 96-well plate (Falcon, 353075), and the decrease in absorbance was recorded using a microplate spectrophotometer system (Biochrom, UK). Results were analyzed with GraphPad Prism v4.0 and presented as LDH activity (mU/ml).

An MTT test was used to assess viability. HepG2 cells, cultured at a density of 5.0 × 104 in a 96-well plate in Dulbecco's modified Eagle's medium (DMEM), were treated with compounds C1-C10 (10 and 100 μM) for 24 hours. After the medium was changed, HepG2 cells were treated with 5 mg/ml MTT (Thiazolyl Blue Tetrazolium Bromide) solution (Sigma, M5655) for 4 hours. After cells were dissolved in DMSO, the level of formazane was analyzed by measuring the optical density at 570 nm against the optical density at 630 nm. Results were analyzed with GraphPad Prism v4.0 and are presented as the percent viability of control values.

**3T3-L1 Preadipocyte Cell Culture and Treatment.** The 3T3-L1 preadipocyte cell line was used to evaluate the adipogenic activity and the stimulation of the insulin-induced glucose uptake of selected compounds. 3T3-L1 pre-adipocytes were propagated and induced to differentiate in DMEM. Proliferating preadipocytes were maintained at low density in a culture medium (growth medium) that consisted of DMEM supplemented with 10% calf serum, 2 mM glutamine, 100 U/ml penicillin and 100 μg/ml streptomycin. For the differentiation assay, 2-days post-confluent preadipocytes were treated with 200 nM insulin and 1 μM test compound for 6 days in DMEM supplemented with 10% fetal bovine serum (FBS). The treatment medium was changed every 2 days. Then, toxicity and triglyceride content were measured. For the glucose uptake assay, cells were differentiated with a differentiation cocktail as previously described43. Briefly, cells were treated with 0.25 μM dexamethasone, 0.5 mM 3-isobutyl-methylxanthine, and 200 nM insulin for 2 days in DMEM containing 10% FBS , then switched to the same media containing insulin for 2 more days, and then switched to the same media without insulin. Ten days after differentiation was induced, cells were treated with the test compounds (1 μM) for 3 more days and used for the glucose uptake assay. Cells were also grown in 48-well plates and exposed to 1 μM of the selected compounds. Cellular viability was assessed by the neutral red assay [44].

**Evaluation of the Adipogenic Activity of the Selected Compounds.** Treated cells were rinsed twice with PBS, scraped into a 250-μl solution of 50 mM Tris-HCl, 1

mM EDTA and 1 mM b-mercaptoethanol and sonicated. The resulting cell lysates were used to determine the total triacylglyceride content, measured using the enzymatic glycerol-phosphate oxidase test, following the manufacturer's instructions (QCA, Amposta, Spain). Results were expressed as the mean ± SEM. The effects were assessed using a one-way ANOVA or Student's T-test. We used Tukey's Test of honestly significant differences to make pairwise comparisons. All calculations were performed using SPSS (IBM Corp., New York, USA).

**Glucose Uptake Assay.** After the treatment of 3T3-L1 adipocytes cultured on 12-well plates with the different compounds, the cells were serum-depleted for 3 hours, and 200 nM insulin or water (vehicle control) was added for 30 min. Glucose transport was determined by measuring the 2-deoxy-d-[3H]glucose uptake as previously described45. Protein content assessed by the Bradford method46 was used to normalize the glucose transport values. Each condition was run in triplicate. Results were expressed as the mean ± SEM. The effects were assessed using a one-way ANOVA or Student's T-test. We used Tukey's Test of honestly significant differences to make pairwise comparisons. All calculations were performed using SPSS.

**Docking of Novel PPARγ Partial Agonists**. Docking studies of the PPARγ partial agonists C1, C5, C7, C8 and C9 were performed with the software Glide v5.6 (Schrödinger LLC., Portland, USA; http://www.schrodinger.com) on the PPARγ crystal structure 2Q5S. For compound C7, an additional docking study was performed with the 2HFP structure. The binding site was defined using the Receptor Grid Generation panel with the default options. Extra-precision (XP) docking was selected for screening the ligands. We selected the flexible docking mode, meaning that Glide internally generated the conformations during the docking process. We did not request any constraints for docking. Each docking run recorded a maximum of ten poses per ligand that survived the post-docking minimization. GlideScore XP was used as the fitness function. The best docking poses for the novel PPARγ ligands were selected by taking into account not only the docking scores but also the results of the visual investigation of all docking poses. Maestro v9.2 and Glide XP Visualizer (Schrödinger LLC., Portland, USA; http://www.schrodinger.com) were used for analyzing and visually investigating the ligand-protein interactions of the docking poses.

## References

1. Francis, G. A. Fayard, E. Picard, F.; Auwerx, J. Nuclear receptors and the control of metabolism. Annu. Rev. Physiol. 2003, 65, 261-311.

2. Garcia-Vallvé, S.; Palau, J. Nuclear receptors, nuclear-receptor factors, and nuclear-receptor-like orphans form a large paralog cluster in Homo sapiens. Mol. Biol. Evol 1998, 15, 665-682.

3. Berger, J.; Moller, D. E. The mechanisms of action of PPARs. Annu. Rev. Med. 2002, 53, 409-435.

4. Shearer, B. G.; Billin, A. N. The next generation of PPAR drugs: do we have the tools to find them? Biochim. Biophys. Acta 2007, 1771, 1082-1093.

5. Willson, T. M. Brown, P. J. Sternbach, D. D.; Henke, B. R. The PPARs: from orphan receptors to drug discovery. J. Med. Chem. 2000, 43, 527-550.

6. Feldman, P. L. Lambert, M. H.; Henke, B. R. PPAR modulators and PPAR pan agonists for metabolic diseases: the next generation of drugs targeting peroxisome proliferator-activated receptors? Curr. Top. Med. Chem. 2008, 8, 728-749.

7. Pourcet, B. Fruchart, J.-C. Staels, B.; Glineur, C. Selective PPAR modulators, dual and pan PPAR agonists: multimodal drugs for the treatment of type 2 diabetes and atherosclerosis. Expert. Opin. Emerg. Drugs 2006, 11, 379-401.

8. Jones, D. Potential remains for PPAR-targeted drugs. Nat. Rev. Drug Discov. 2010, 9, 668-669.

9. Guasch, L. Sala, E. Valls, C. Blay, M. Mulero, M. Arola, L. Pujadas, G.; Garcia-Vallvé, S. Structural insights for the design of new PPARgamma partial agonists with high binding affinity and low transactivation activity. J. Comput. Aided Mol. Des. 2011.

10. Zoete, V. Grosdidier, A.; Michielin, O. Peroxisome proliferator-activated receptor structures: ligand specificity, molecular switch and interactions with regulators. Biochim. Biophys. Acta 2007, 1771, 915-925.

11. Farce, A. Renault, N.; Chavatte, P. Structural insight into PPARgamma ligands binding. Curr. Med. Chem. 2009, 16, 1768-1789.

12. Bruning, J. B. Chalmers, M. J. Prasad, S. Busby, S. A. Kamenecka, T. M. He, Y. Nettles, K. W.; Griffin, P. R. Partial agonists activate PPARgamma using a helix 12 independent mechanism. Structure 2007, 15, 1258-1271.

13. Pochetti, G. Godio, C. Mitro, N. Caruso, D. Galmozzi, A. Scurati, S. Loiodice, F. Fracchiolla, G. Tortorella, P. Laghezza, A. Lavecchia, A. Novellino, E. Mazza, F.; Crestani, M. Insights into the mechanism of partial agonism: crystal structures of the peroxisome proliferator-activated receptor gamma ligand-binding domain in the complex with two enantiomeric ligands. J. Biol. Chem. 2007, 282, 17314-17324.

14. Lu, I.-L. Huang, C.-F. Peng, Y.-H. Lin, Y.-T. Hsieh, H.-P. Chen, C.-T. Lien, T.-W. Lee, H.-J. Mahindroo, N. Prakash, E. Yueh, A. Chen, H.-Y. Goparaju, C. M. V. Chen, X. Liao, C.-C. Chao, Y.-S. Hsu, J. T.-A.; Wu, S.-Y. Structure-based drug design of a novel family of PPARgamma partial agonists: virtual screening, X-ray

crystallography, and in vitro/in vivo biological activities. J. Med. Chem. 2006, 49, 2703-2712.

15. Gelman, L. Feige, J. N.; Desvergne, B. Molecular basis of selective PPARgamma modulation for the treatment of Type 2 diabetes. Biochim. Biophys. Acta 2007, 1771, 1094-1107.

16. Choi, J. H. Banks, A. S. Estall, J. L. Kajimura, S. Boström, P. Laznik, D. Ruas, J. L. Chalmers, M. J. Kamenecka, T. M. Blüher, M. Griffin, P. R.; Spiegelman, B. M. Anti-diabetic drugs inhibit obesity-linked phosphorylation of PPARgamma by Cdk5. Nature 2010, 466, 451-456.

17. Lewis, S. N. Bassaganya-Riera, J.; Bevan, D. R. Virtual Screening as a Technique for PPAR Modulator Discovery. PPAR Res. 2010, 2010, 861238.

18. Markt, P. Petersen, R. K. Flindt, E. N. Kristiansen, K. Kirchmair, J. Spitzer, G. Distinto, S. Schuster, D. Wolber, G. Laggner, C.; Langer, T. Discovery of novel PPAR ligands by a virtual screening approach based on pharmacophore modeling, 3D shape, and electrostatic similarity screening. J. Med. Chem. 2008, 51, 6303-6317.

19. Choi, J. Park, Y. Lee, H. S. Yang, Y.; Yoon, S. 1,3-Diphenyl-1H-pyrazole derivatives as a new series of potent PPARγ partial agonists. Bioorg. Med. Chem. 2010, 18, 8315-8323.

20. Vidović, D. Busby, S. A. Griffin, P. R.; Schürer, S. C. A Combined Ligand- and Structure-Based Virtual Screening Protocol Identifies Submicromolar PPARγPartial Agonists. ChemMedChem. 2011, 6, 94-103.

21. Rau, O. Wurglics, M. Dingermann, T. Abdel-Tawab, M.; Schubert-Zsilavecz, M. Screening of herbal extracts for activation of the human peroxisome proliferator-activated receptor. Pharmazie 2006, 61, 952-956.

22. Salam, N. K. Huang, T. H.-W. Kota, B. P. Kim, M. S. Li, Y.; Hibbs, D. E. Novel PPAR-gamma agonists identified from a natural product library: a virtual screening, induced-fit docking and biological assay study. Chem. Biol. Drug Des. 2008, 71, 57-70.

23. Huang, T. H.-W. Kota, B. P. Razmovski, V.; Roufogalis, B. D. Herbal or natural medicines as modulators of peroxisome proliferator-activated receptors and related nuclear receptors for therapy of metabolic syndrome. Basic Clin. Pharmacol. Toxicol. 2005, 96, 3-14.

24. Tanrikulu, Y. Rau, O. Schwarz, O. Proschak, E. Siems, K. Müller-Kuhrt, L. Schubert-Zsilavecz, M.; Schneider, G. Structure-based pharmacophore screening for natural-product-derived PPARgamma agonists. Chembiochem. 2009, 10, 75-78.

25. Fakhrudin, N. Ladurner, A. Atanasov, A. G. Heiss, E. H. Baumgartner, L. Markt, P. Schuster, D. Ellmerer, E. P. Wolber, G. Rollinger, J. M. Stuppner, H.; Dirsch, V. M. Computer-aided discovery, validation, and mechanistic characterization of novel neolignan activators of peroxisome proliferator-activated receptor gamma. Mol. Pharmacol. 2010, 77, 559-566.

26. Petersen, R. K. Christensen, K. B. Assimopoulou, A. N. Fretté, X. Papageorgiou, V. P. Kristiansen, K.; Kouskoumvekaki, I. Pharmacophore-driven identification of PPARγ agonists from natural sources. J. Comput. Aided Mol. Des. 2010.

27. Huang, N. Shoichet, B. K.; Irwin, J. J. Benchmarking sets for molecular docking. J. Med. Chem. 2006, 49, 6789-6801.

28. Irwin, J. J.; Shoichet, B. K. ZINC--a free database of commercially available compounds for virtual screening. J. Chem. Inf. Model. 2005, 45, 177-182.

29. Hopkins, C. R. O'neil, S. V. Laufersweiler, M. C. Wang, Y. Pokross, M. Mekel, M. Evdokimov, A. Walter, R. Kontoyianni, M. Petrey, M. E. Sabatakos, G. Roesgen, J. T. Richardson, E.; Demuth, T. P., Jr Design and synthesis of novel N-sulfonyl-2-indole carboxamides as potent PPAR-gamma binding agents with potential application to the treatment of osteoporosis. Bioorg. Med. Chem. Lett. 2006, 16, 5659-5663.

30. Berman, H. M. Westbrook, J. Feng, Z. Gilliland, G. Bhat, T. N. Weissig, H. Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. Nucleic Acids. Res. 2000, 28, 235-242.

31. Kleywegt, G. J. Harris, M. R. Zou, J. Y. Taylor, T. C. Wählby, A.; Jones, T. A. The Uppsala Electron-Density Server. Acta Crystallogr. D Biol. Crystallogr. 2004, 60, 2240-2249.

32. Guex, N.; Peitsch, M. C. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 1997, 18, 2714-2723.

33. Wolber, G.; Langer, T. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. J. Chem. Inf. Model. 2005, 45, 160-169.

34. Wolber, G. Dornhofer, A. A.; Langer, T. Efficient overlay of small organic molecules using 3D pharmacophores. J. Comput. Aided Mol. Des 2006, 20, 773-788.

35. Dixon, S. L. Smondyrev, A. M. Knoll, E. H. Rao, S. N. Shaw, D. E.; Friesner, R. A. PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. J. Comput. Aided Mol. Des. 2006, 20, 647-671.

36. Lagorce, D. Sperandio, O. Galons, H. Miteva, M. A.; Villoutreix, B. O. FAF-Drugs2: free ADME/tox filtering tool to assist drug discovery and chemical biology projects. BMC Bioinformatics 2008, 9, 396.

37. Lipinski, C. A. Lombardo, F. Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv. Drug Deliv. Rev. 2001, 46, 3-26.

38. Zsoldos, Z. Reid, D. Simon, A. Sadjad, S. B.; Johnson, A. P. eHiTS: a new fast, exhaustive flexible ligand docking system. J. Mol. Graph. Model. 2007, 26, 198-212.

39. Mills, N. ChemDraw Ultra 10.0. J. Am. Chem. Soc. 2006, 128, 13649-13650.

40. Schuster, D.; Wolber, G. Identification of bioactive natural products by pharmacophore-based virtual screening. Curr. Pharm. Des. 2010, 16, 1666-1681.

41. Duan, J. Dixon, S. L. Lowrie, J. F.; Sherman, W. Analysis and comparison of 2D fingerprints: insights into database screening performance using eight fingerprint methods. J. Mol. Graph. Model. 2010, 29, 157-170.

42. Kelley, L. A. Gardner, S. P.; Sutcliffe, M. J. An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies. Protein Eng. 1996, 9, 1063-1065.

43. Ardévol, A. Bladé, C. Salvadó, M. J.; Arola, L. Changes in lipolysis and hormone-sensitive lipase expression caused by procyanidins in 3T3-L1 adipocytes. Int. J. Obes. Relat. Metab. Disord. 2000, 24, 319-324.

44. Borenfreund, E.; Puerner, J. A. Toxicity determined in vitro by morphological alterations and neutral red absorption. Toxicol. Lett. 1985, 24, 119-124.

45. Pinent, M. Blay, M. Bladé, M. C. Salvadó, M. J. Arola, L.; Ardévol, A. Grape seed-derived procyanidins have an antihyperglycemic effect in streptozotocin-induced diabetic rats and insulinomimetic activity in insulin-sensitive cell lines. Endocrinology 2004, 145, 4985-4990.

46. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 1976, 72, 248-254.

## Supporting Information

**Table S1.** Smiles of the 135 PPARγ full agonists used in the VS validation.

| Title | Smile |
|---|---|
| ppar001 | O=C(O)CCCCCCC/C=C/C=C/[C@@H](O)CCCCC |
| ppar002 | O=C(O)CCCCCCC[C@@H](O)\C=C\C=C\CCCCC |
| ppar003 | c1occc1C2=C(C(=O)OCC)C(=O)c(c23)cc(cc3)OC |
| ppar004 | COc(cc1)cc(c12)C(=O)C(C(=O)OCC)=C2c3ccccc3 |
| ppar005 | O=C(O)CSc(n1)nc(Cl)cc1Nc2cccc(C)c2C |
| ppar006 | N#Cc1c(SC)cc(oc1=O)-c2ccc(cc2)OCCSC |
| ppar007 | N1C(=O)S[C@H](C1=O)Cc2ccc(cc2)OCC3(C)CCCCC3 |
| ppar008 | O=C(O)CC[C@@H](O)\C=C\C=C\C/C=C/C/C=C/C/C=C/C/C=C/CC |
| ppar009 | CC(=O)/C=C(\C)N[C@H](C(=O)O)Cc1ccc(cc1)OCc2ccccc2 |
| ppar010 | N1C(=O)S[C@H](C1=O)Cc(c2)ccc(c23)O[C@H](CC3)Cc4ccccc4 |
| ppar011 | O=C(S1)NC(=O)/C1=C\c2ccc(cc2)OCCN(C)c3ccccn3 |
| ppar012 | c1cccc(c1CC(=O)O)OCC[C@H](C(=C)N2CCC)Oc(c23)cccc3 |
| ppar013 | CN[C@H](C(=O)O)Cc1ccc(cc1)OCCc2c(C)oc(n2)-c3ccccc3 |
| ppar014 | COc(cc1)cc(c12)C(=O)C(CC(=O)OCC)=C2C34C[C@@H]5C[C@H](C3)C[C@H](C4)C5 |
| ppar015 | c1cccc(c12)oc(n2)N(C)CCOc(cc3)ccc3[C@@H](C4=O)SC(=O)N4 |
| ppar016 | c1cccc(c1CC(=O)O)OCC[C@H](C(=O)N2CCCC)Oc(c23)cccc3 |
| ppar017 | c1cccc(c1CC(=O)O)OCC[C@@H](C(=O)N2CC(C)C)Oc(c23)cccc3 |
| ppar018 | c1ccccc1C(\C)=N/OCCOc(cc2)ccc2C[C@@H](C3=O)SC(=O)N3 |
| ppar019 | n1ccccc1C(\C)=N/OCCOc(cc2)ccc2C[C@@H](C3=O)SC(=O)N3 |
| ppar020 | c1ncccc1C(\C)=N/OCCOc(cc2)ccc2C[C@@H](C3=O)SC(=O)N3 |
| ppar021 | c1cnccc1C(\C)=N/OCCOc(cc2)ccc2C[C@@H](C3=O)SC(=O)N3 |
| ppar022 | CN(C)[C@H](C(=O)O)Cc1ccc(cc1)OCCc2c(C)oc(n2)-c3ccccc3 |
| ppar023 | O=C(S1)NC(=O)/C1=C\c2ccc(cc2)OCCN(C)c(n3)oc(c34)cccc4 |
| ppar024 | c1cccc(c12)oc(n2)N(C)CCOc(cc3)ccc3C[C@@H](C4=O)SC(=O)N4 |
| ppar025 | c1cccc(c1CC(=O)O)OCC[C@@H](Oc(c23)cccc2)C(=O)N3CCCCC |
| ppar026 | N1C(=O)O[C@@H](C1=O)CCCc2ccc(cc2)OCc3c(C)oc(n3)-c4ccccc4 |
| ppar027 | CC(=O)N[C@@H](C(=O)O)Cc1ccc(cc1)OCCc2c(C)oc(n2)-c3ccccc3 |
| ppar028 | c1ccccc1-c(n2)oc(C)c2CCOc(cc3)ccc3C[C@@H](C4=O)SC(=O)N4 |
| ppar029 | CC[C@@](C(=O)O)(CC1)Oc(c12)cc(cc2)OCCCOc3c(Cl)cc(F)cc3 |
| ppar030 | c1cccc(c1CC(=O)O)OCC[C@H](Oc(c23)cccc2)C(=O)N3CCCCC=C |
| ppar031 | CO[C@@H](C(=O)O)Cc1ccc(cc1)OCCOc(cc2)ccc2-c3ccc(F)cc3 |
| ppar032 | c1cccc(c1CC(=O)O)OCC[C@@H](Oc(c23)cccc2)C(=O)N3CCCCCC |
| ppar033 | c1cccc(c1CC(=O)O)OCC[C@H](Oc(c23)cccc2)C(=O)N3CCCCCC |
| ppar034 | c1cccc(c1CC(=O)O)OCC[C@H](Oc(c23)cccc2)C(=O)N3CCCCCO |
| ppar035 | c1cccc(c1CC(=O)O)OCC[C@H](Oc(c23)cccc2)C(=O)N3CCCCOC |
| ppar036 | c1cccc(c1CC(=O)O)OCC[C@@H](Oc(c23)cccc2)C(=O)N3CCCCOC |
| ppar037 | c1ccccc1C(=O)\C=C(\C)N[C@@H](C(=O)N)Cc2ccc(cc2)OCc3ccccc3 |

| | |
|---|---|
| ppar038 | c1ccccc1C(=O)\C=C(\C)N[C@@H](C(=O)O)Cc2ccc(cc2)OCc3ccccc3 |
| ppar039 | c1cccn1[C@@H](C(=O)O)Cc2ccc(cc2)OCCc3c(C)oc(n3)-c4ccccc4 |
| ppar040 | CO[C@@H](C(=O)O)Cc1ccc(cc1)OCCOc(cc2)ccc2-c(cc3)ccc3CC |
| ppar041 | C1CCCN1[C@H](C(=O)O)Cc2ccc(cc2)OCCc3c(C)oc(n3)-c4ccccc4 |
| ppar042 | N1C(=O)O[C@@H](C1=O)CCCc2ccc(cc2)OCc3c(C)sc(n3)-c4ccccc4 |
| ppar043 | c1cccc(c1CC(=O)O)OCC[C@@H](Oc(c23)cccc2)C(=O)N3CCC4CCCC4 |
| ppar044 | CC(=O)CCCCN1C(=O)[C@H](Oc(c12)cccc2)CCOc(c3CC(=O)O)cccc3 |
| ppar045 | c1cccc(c1CC(=O)O)OCC[C@@H](Oc(c23)cccc2)C(=O)N3CCCCCCC |
| ppar046 | c1cccc(c1CC(=O)O)OCC[C@@H](Oc(c23)cccc2)C(=O)N3CCCCC(C)C |
| ppar047 | N1C(=O)O[C@@H](C1=O)CCCc2cc(OC)c(cc2)OCc3c(C)oc(n3)-c4ccco4 |
| ppar048 | c1cccc(c1CC(=O)O)OCC[C@H](Oc(c23)cccc2)C(=O)N3CCCCCCO |
| ppar049 | c1cccc(c1CC(=O)O)OCC[C@H](Oc(c23)cccc2)C(=O)N3CCCC[C@@H](C)O |
| ppar050 | c1cccc(c1CC(=O)O)OCC[C@H](Oc(c23)cccc2)C(=O)N3CCCCCCF |
| ppar051 | c1cccc(c1CC(=O)O)OCC[C@@H](Oc(c23)cccc2)C(=O)N3CCCCCCF |
| ppar052 | c1ccccc1C(=O)\C=C(\C)N[C@@H](C(=O)OC)Cc2ccc(cc2)OCc3ccccc3 |
| ppar053 | c1cccc(c1CC(=O)O)OCC[C@H](Oc(c23)cccc2)C(=O)N3CCSCCC |
| ppar054 | CC[C@](C1)(C(=O)O)Oc(c12)ccc(c2)OCCCOc(cc3)c(Cl)cc3C(C)(C)C |
| ppar055 | C1CCCCN1[C@@H](C(=O)O)Cc2ccc(cc2)OCCc3c(C)oc(n3)-c4ccccc4 |
| ppar056 | c1cccc(c1CC(=O)O)OCC[C@H](Oc(c23)cccc2)C(=O)N3CCCC4CCCC4 |
| ppar057 | c1cccc(c1CC(=O)O)OCC[C@H](Oc(c23)cccc2)C(=O)N3CCCC4CCCC4 |
| ppar058 | CC(=O)CCCCCN1C(=O)[C@@H](Oc(c12)cccc2)CCOc(c3CC(=O)O)cccc3 |
| ppar059 | c1cccc(c1CC(=O)O)OCC[C@H](Oc(c23)cccc2)C(=O)N3CCCCCCCC |
| ppar060 | c1cccc(c1CC(=O)O)OCC[C@@H](Oc(c23)cccc2)C(=O)N3CCCCC(C)(C)C |
| ppar061 | NC(=O)CCCCCN1C(=O)[C@H](Oc(c12)cccc2)CCOc(c3CC(=O)O)cccc3 |
| ppar062 | O\N=C(C)\CCCCN1C(=O)[C@H](Oc(c12)cccc2)CCOc(c3CC(=O)O)cccc3 |
| ppar063 | c1cccc(c1CC(=O)O)OCC[C@H](Oc(c23)cccc2)C(=O)N3CCCCCCCO |
| ppar064 | c1cccc(c1CC(=O)O)OCC[C@H](Oc(c23)cccc2)C(=O)N3CCCCC(C)(C)O |
| ppar065 | c1cccc(c1CC(=O)O)OCC[C@H](Oc(c23)cccc2)C(=O)N3CCCCCCCF |
| ppar066 | n1cccnc1N[C@@H](C(=O)O)Cc2ccc(cc2)OCCc3c(C)oc(n3)-c4ccccc4 |
| ppar067 | CS(=O)(=O)N[C@@H](C(=O)O)Cc1ccc(cc1)OCCc2c(C)oc(n2)-c3ccccc3 |
| ppar068 | Cc1ccc(C)n1[C@@H](C(=O)O)Cc2ccc(cc2)OCCc3c(C)oc(n3)-c4ccccc4 |
| ppar069 | S1SCC[C@H]1CCCCC(=O)N(C)CCCOc(cc2)c(Cl)cc2CC(=O)O |
| ppar070 | CC(C)(C)Cc1cc(Cl)c(cc1)OCCCOc(c2)ccc(c23)O[C@@](C3)(C(=O)O)CC |
| ppar071 | O=C(O)Cn(cc1)c(c12)ccc(c2)OCCCOc(cc3)ccc3C(=O)c4ccc(F)cc4 |
| ppar072 | c1cccc(c1CC(=O)O)OCC[C@H](Oc(c23)cccc2)C(=O)N3CCCCC(F)(F)C |
| ppar073 | c1cccc(c1CC(=O)O)OCC[C@H](Oc(c23)cccc2)C(=O)N3CCCCC(F)(F)C |
| ppar074 | N#CC(C)(C)CCCCN1C(=O)[C@H](Oc(c12)cccc2)CCOc(c3CC(=O)O)cccc3 |
| ppar075 | c1ccccc1C(=O)c2c(cccc2)N[C@@H](C(=O)O)Cc3ccc(cc3)OCc4ccccc4 |
| ppar076 | c1ccccc1C(=O)c2c(cccc2)N[C@H](C(=O)O)Cc3ccc(cc3)OCc4ccccc4 |
| ppar077 | c1cccc(c1CC(=O)O)OCC[C@H](Oc(c23)cccc2)C(=O)N3CCCCCCCCC |
| ppar078 | CO\N=C(C)\CCCCN1C(=O)[C@H](Oc(c12)cccc2)CCOc(c3CC(=O)O)cccc3 |
| ppar079 | S1SCC[C@@H]1CCCCC(=O)NCCOc(cc2)ccc2C[C@@H](C3=O)SC(=O)N3 |
| ppar080 | c1ccccc1C(=O)C2=C(CCCC2)N[C@@H](C(=O)O)Cc3ccc(cc3)OCc4ccccc4 |

| | |
|---|---|
| ppar081 | N1C(=O)O[C@H](C1=O)CCCc2ccc(cc2)OCc3c(C)oc(n3)-c(c4)ccc(c45)cccc5 |
| ppar082 | CC[C@](C1)(C(=O)O)Oc(c12)ccc(c2)OCCCOc(cc3)c(Cl)cc3C(F)(F)F |
| ppar083 | CC(C)[C@](C1)(C(=O)O)Oc(c12)ccc(c2)OCCCOc(c(Cl)cc3)cc3C(F)(F)F |
| ppar084 | S1SCC[C@H]1CCCCC(=O)N(C)CCCOc(cc2)c(Cl)cc2CC(=O)OC |
| ppar085 | c1ccccc1-c(cc2)ccc2C(\C)=N/OCCOc(cc3)ccc3C[C@@H](C4=O)SC(=O)N4 |
| ppar086 | c1ccccc1-c(cc2)ccc2C(\C)=N/OCCOc(cc3)ccc3C[C@@H](C4=O)SC(=O)N4 |
| ppar087 | CC(C)[C@](C1)(C(=O)O)Oc(c12)ccc(c2)OCCOc3c(Cl)cc(cc3)OC(F)(F)F |
| ppar088 | CC[C@](C1)(C(=O)O)Oc(c12)ccc(c2)OCCCOc3c(Cl)cc(cc3)OC(F)(F)F |
| ppar089 | O=C(O)[C@](C)(C1)Oc(c12)ccc(c2)OCCCOc3c(Cl)cc(cc3)OCC(F)(F)F |
| ppar090 | c1cccc(c1CC(=O)O)OCC[C@H](Oc(c23)cccc2)C(=O)N3CCCCCC(F)(F)C |
| ppar091 | c1ccccc1-c(cn2)ccc2C(\C)=N/OCCOc(cc3)ccc3C[C@@H](C4=O)SC(=O)N4 |
| ppar092 | c1ccccc1-c(nc2)ccc2C(\C)=N/OCCOc(cc3)ccc3C[C@@H](C4=O)SC(=O)N4 |
| ppar093 | n1ccccc1-c(cc2)ccc2C(\C)=N/OCCOc(cc3)ccc3C[C@@H](C4=O)SC(=O)N4 |
| ppar094 | c1ncccc1-c(cc2)ccc2C(\C)=N/OCCOc(cc3)ccc3C[C@@H](C4=O)SC(=O)N4 |
| ppar095 | c1cnccc1-c(cc2)ccc2C(\C)=N/OCCOc(cc3)ccc3C[C@@H](C4=O)SC(=O)N4 |
| ppar096 | c1cccc(c1CC(=O)O)OCC[C@@H](Oc(c23)cccc2)C(=O)N3CCCCCCCCC |
| ppar097 | S1SCC[C@@H]1CCCCC(=O)N(C)CCOc(cc2)ccc2C[C@@H](C3=O)SC(=O)N3 |
| ppar098 | O=C(O)C(C)(C)CCCCN1C(=O)[C@@H](Oc(c12)cccc2)CCOc(c3CC(=O)O)cccc3 |
| ppar099 | SCC[C@@H](S)CCCCC(=O)N(C)CCOc(cc1)ccc1C[C@@H](C2=O)SC(=O)N2 |
| ppar100 | c1ccccc1C(=O)c2c(cccc2)N[C@@H](C(=O)O)Cc3ccc(cc3)OCC4(C)CCCCC4 |
| ppar101 | c1ccccc1-c(cc2)ccc2C(\CC)=N/OCCOc(cc3)ccc3C[C@@H](C4=O)SC(=O)N4 |
| ppar102 | CC[C@](C1)(C(=O)O)Oc(c12)ccc(c2)OCCCOc3c(Cl)cc(cc3)OCC(F)(F)F |
| ppar103 | CC(C)[C@](C1)(C(=O)O)Oc(c12)ccc(c2)OCCCOc3c(Cl)cc(cc3)OC(F)(F)F |
| ppar104 | C1C[C@@](C)(C(=O)O)Oc(c12)cc(cc2)OCCCOc3c(Cl)cc(cc3)OCC(F)(F)F |
| ppar105 | CC[C@](C1)(C(=O)O)Oc(c12)ccc(c2)OCCCOc3c(Cl)cc(cc3)SC(F)(F)F |
| ppar106 | CC[C@](C1)(C(=O)O)Oc(c12)ccc(c2)OCCCOc(c3CCC)ccc(c3)SC(F)(F)F |
| ppar107 | CC(C)(C)[C@](C1)(C(=O)O)Oc(c12)ccc(c2)OCCCOc(cc3)c(Cl)cc3CC(F)(F)F |
| ppar108 | c1ccccc1-c(cc2)ccc2C(\CCC)=N/OCCOc(cc3)ccc3C[C@H](C4=O)SC(=O)N4 |
| ppar109 | N#CC(=C1)C(=O)C(C)(C)[C@@H](CC2)[C@]1(C)C(=CC3=O)[C@]2(C)[C@@](C)(CC4)[C@@H]3[C@H]([C@@]45C(=O)O)CC(C)(C)CC5 |
| ppar110 | c1ccccc1C(=O)c2c(cccc2)N[C@@H](C(=O)O)Cc3ccc(cc3)OCCc(nc4)ccc4CC |
| ppar111 | N1C(=O)S[C@@H](C1=O)c2ccc(cc2)OCCCOc(c3CCC)ccc(c3)Oc4ccc(F)cc4 |
| ppar112 | O=C(O)[C@](C(F)(F)F)(C1)Oc(c12)ccc(c2)OCCCOc(cc3)c(Cl)cc3CC(F)(F)F |
| ppar113 | CCC(C(=O)O)(CC)Cc1ccc(cc1)OCCCOc2c(Cl)cc(cc2)Oc3ccc(F)cc3 |
| ppar114 | COc(c1)cc(OC)cc1\C=C(\C(=O)O)c2ccc(cc2)Oc(cc3)ccc3C[C@@H](C4=O)SC(=O)N4 |
| ppar115 | c1ccccc1C(=N/N)\c2c(cccc2)N[C@@H](C(=O)O)Cc3ccc(cc3)OCCN(C)c4ccccn4 |
| ppar116 | COc(c1)cc(OC)cc1\C=C(\C(=O)OC)c2ccc(cc2)Oc(cc3)ccc3/C=C4/C(=O)NC(=O)S4 |
| ppar117 | COc(c1)cc(OC)cc1C[C@H](C(=O)OC)c2ccc(cc2)Oc(cc3)ccc3/C=C4/C(=O)NC(=O)S4 |
| ppar118 | c1ccccc1[C@H](C(=O)O)Cc2ccc(cc2)OCCCOc3c(Cl)cc(cc3)Oc4ccc(F)cc4 |
| ppar119 | COc(c1)cc(OC)cc1C[C@@H](C(=O)OC)c2ccc(cc2)Oc(cc3)ccc3C[C@@H](C4=O)SC(=O)N4 |
| ppar120 | c1ccccc1Oc(cccc2)c2N[C@@H](C(=O)O)Cc3ccc(cc3)OCCN(C)c(n4)oc(c45)cccc5 |
| ppar121 | c1cccc(C2=O)c1-c(c23)cccc3N[C@@H](C(=O)O)Cc4ccc(cc4)OCCN(C)c(n5)oc(c56)cccc6 |
| ppar122 | c1ccccc1C(=O)c2c(cccc2)C[C@@H](C(=O)O)Cc3ccc(cc3)OCCN(C)c(n4)oc(c45)cccc5 |

| | |
|---|---|
| ppar123 | c1ccccc1C(=O)c2c(cccc2)N[C@@H](C(=O)O)Cc3ccc(cc3)OCCN(C)c(n4)oc(c45)cccc5 |
| ppar124 | c1ccccc1C(=O)c2c(cccc2)O[C@@H](C(=O)O)Cc3ccc(cc3)OCCN(C)c(n4)oc(c45)cccc5 |
| ppar125 | c1ccccc1CC(=O)c2c(cccc2)N[C@@H](C(=O)O)Cc3ccc(cc3)OCCN(C)c(n4)oc(c45)cccc5 |
| ppar126 | c1ccccc1C(=O)c2c(cccc2)S[C@@H](C(=O)O)Cc3ccc(cc3)OCCN(C)c(n4)oc(c45)cccc5 |
| ppar127 | CC(=O)SCC[C@@H](SC(=O)C)CCCCC(=O)N(C)CCOc(cc1)ccc1C[C@@H](C2=O)SC(=O)N2 |
| ppar128 | c1cccc(C2=O)c1C(=O)c(c23)cccc3N[C@@H](C(=O)O)Cc4ccc(cc4)OCCN(C)c(n5)oc(c56)cccc6 |
| ppar129 | O=C(O)CCC(=O)SCC[C@@H](S)CCCCC(=O)N(C)CCOc(cc1)ccc1C[C@@H](C2=O)SC(=O)N2 |
| ppar130 | [NH4]CC(=O)SCC[C@@H](SC(=O)C)CCCCC(=O)N(C)CCOc(cc1)ccc1C[C@@H](C2=O)SC(=O)N2 |
| ppar131 | c1ccccc1S(=O)(=O)c(ccc2)cc2N[C@@H](C(=O)O)Cc3ccc(cc3)OCCN(C)c(n4)oc(c45)cccc5 |
| ppar132 | c1ccccc1C(=O)c2c(cccc2)N[C@@H](C(=O)O)Cc3ccc(cc3)OCC[C@@H](CC4)Oc(c45)c(C)c(C)c(O)c5C |
| ppar133 | [NH4]CC(=O)SCC[C@@H](SC(=O)C[NH4])CCCCC(=O)N(C)CCOc(cc1)ccc1C[C@@H](C2=O)SC(=O)N2 |
| ppar134 | O=C(O)CCC(=O)SCC[C@@H](SC(=O)C)CCCCC(=O)N(C)CCOc(cc1)ccc1C[C@@H](C2=O)SC(=O)N2 |
| ppar135 | O=C(O)CCC(=O)SCC[C@@H](SC(=O)CCC(=O)O)CCCCC(=O)N(C)CCOc(cc1)ccc1C[C@@H](C2=O)SC(=O)N2 |

**Table S2.** Smiles of the 19 PPARγ partial agonists used in the VS validation.

| Title | Smile |
|---|---|
| 14 | O=C(O)[C@H](C)Oc(ccc1)cc1Cc2c(C)n(c(c23)ccc(c3)OC(F)(F)F)C(=O)c4ccc(cc4)OC |
| 15 | O=C(O)[C@H](C)Oc(c1)ccc(Cl)c1Cn(c(c23)cc(cc3)OC(F)(F)F)c(C)c2-c4noc(c45)cc(cc5)OC |
| 16 | CC(C)(C)c1ccc(cc1)-n(c(c23)cccc3)c(C(=O)O)c2Oc(cc4C(F)(F)F)ccc4 |
| 17 | FC(F)(F)c1cc(C(F)(F)F)cc(c1)S(=O)(=O)Nc(cc2-c3cccs3)n(n2)-c4ccc(F)cc4 |
| 1WM0 | c1cc(Cl)cc(Cl)c1C(=O)Nc(c2C(=O)O)ccc(c2)OCN/C=C\C |
| 2G0G | s1cccc1-c2cc(NS(=O)(=O)c(cc3F)ccc3)n(n2)-c4ccc(F)cc4 |
| 2HFP | c1ccccc1S(=O)(=O)NC(=O)c(n(Cc(cc2C(F)(F)F)ccc2)c(c34)cccc4)c3-c5ccc(cc5)OC |
| 2Q61 | c1ccccc1Sc2c(C(=O)O)n(c(c23)ccc(Cl)c3)Cc4ccccc4 |
| 2Q6R | c1ccccc1Sc2c(C(=O)O)n(c(c23)ccc(Cl)c3)Cc4cc(OC)ccc4 |
| amg-131 | c1cc(Cl)cc(Cl)c1C(=O)Nc(c2C(=O)O)ccc(c2)Oc3ncccn3 |
| fk-614 | CCCCCS(=O)(=O)NC(=O)c(c1)ccc(c12)nc(C)n2Cc3c(Cl)cc(Cl)cc3 |
| FMOC | CC(C)C[C@@H](C(=O)O)NC(=O)OCC(c(c-12)cccc2)c3c1cccc3 |
| gw0072 | c1ccccc1CN(Cc2ccccc2)C(=O)C[C@@H](C3=O)S[C@@H](CCCCCCC)N3CCCCc(cc4)ccc4C(=O)O |
| kr-62980 | c1ccccc1C2=C(C(=O)OCC)/C(=[N+]([O-])\CC)c(c23)cc(cc3)OCCN4CCOCC4 |
| Metaglidasen | CC(=O)NCOC(=O)[C@@H](Oc(cc1C(F)(F)F)ccc1)c2ccc(Cl)cc2 |
| nTZDpa | c1ccccc1Sc2c(C(=O)O)n(c(c23)ccc(Cl)c3)Cc4ccc(Cl)cc4 |
| pa-082 | COc(cc1)c(OC)cc1Cc(c(c23)cc(OC)c(c3)OC)ncc2CN(CC4)CCC4c5c(OC)cccc5 |
| pat5a | O=C(S1)NC(=O)/C1=C\c2ccc(cc2)OC[C@H]3CCCN3c4ccccn4 |
| s-26948 | COC(=O)C(C(=O)OC)Cc1ccc(cc1)OCCn2c(=O)sc(c23)cc(cc3)C(=O)c4ccccc4 |

CHAPTER 4

# IDENTIFICATION OF NATURAL EXTRACTS WITH ANTIDIABETIC PROPERTIES THAT CONTAIN PPARγ PARTIAL

## ABSTRACT

Natural extracts have played an important role in the prevention and treatment of diseases and are important sources for drug discovery. However, to be effectively used in these processes, natural extracts must be characterized through the identification of their active compounds and their modes of action. In this paper, from an initial set of 29,779 natural products that are annotated with their natural source and using a previously developed virtual screening procedure (carefully validated experimentally), we have predicted as potential peroxisome proliferators-activated receptor gamma (PPARγ) partial agonists 12 molecules from 11 extracts known to have antidiabetic activity (these molecules are from the plants *Achyrocline satureoides*, *Andrographis paniculata*, *Angelica keiskei*, *Cryptolepis sanguinolenta*, *Harungana madagascariensis*, *Salvia miltiorrhiza* and *Scutellaria baicalensis*; the fungi *Aspergillus terreus* and *Hericium erinaceum*; and the marine species *Dysidea villosa* and *Fucus vesiculosus*). Six of these molecules are similar to molecules with described antidiabetic activity but whose mechanism of action is unknown. Therefore, it is plausible that these 12 molecules could be the bioactive molecules responsible, at least in part, for the antidiabetic activity of the extracts containing them. In addition, we have also identified as potential PPARγ partial agonists 10 molecules from 16 plants with undescribed antidiabetic activity but that are related (*i.e.*, they are from the same genus) to plants with known antidiabetic properties. These plants (*e.g.*, *Annona purpurea, Artocarpus gomezianus, Helichrysum stenopterum, Melicope ptelefolia, Murraya paniculata, Salvia eriophora, Salvia lanigera, Salvia prionitis, Swertia hookeri* and *Tephrosia watsoniana*) represent a new source of potential antidiabetic compounds. None of the 22 molecules that we predict as PPARγ partial agonists show chemical similarity with a group of 211 known PPARγ partial agonists obtained from the literature. Consequently, these molecules are lead-hopping candidates for the development of new PPARγ partial agonists.

## Introduction

Since ancient times, natural products (NPs) have played an important role in the treatment of type 2 diabetes mellitus (T2DM) [1]. Plants are one of the most important sources of antidiabetic compounds. Thus, 656 species from 437 genera, representing 111 plant families, with antidiabetic properties have been identified [1]. The plant families most studied as a result of their confirmed antidiabetic effects include *Leguminoseae*, *Lamiaceae*, *Liliaceae*, *Cucurbitaceae*, *Asteraceae*, *Moraceae*, *Rosaceae*, *Euphorbiaceae* and *Araliaceae* [2].

Although plant extracts have been used for the treatment of T2DM for hundreds of years in India [3,4], China and other parts of the world, more research is needed for the identification of their active compounds and their mode of action. Some of the active principles associated with the antidiabetic activity of plant extracts are alkaloids, saponins, xanthones, flavonoids and nonstarch polysaccharides [1]. Despite the wide array of these active principles with a demonstrated antidiabetic activity, to date, metformin is the only drug approved for treatment of T2DM derived from a medicinal plant [5]. Therefore, the identification of the active compounds and the modes of action from plants traditionally used in the treatment of T2DM is an important issue for the discovery of new antidiabetic drugs and for the validation, standardization and rational use of traditional herbal remedies [1].

Numerous mechanisms of antidiabetic actions have been proposed for several plant extracts [1,6] and some hypotheses relate their effects to the increase of the insulin-stimulated glucose uptake. One target of interest for antidiabetic drugs is peroxisome proliferators-activated receptor gamma (PPARγ). PPARγ is a member of the nuclear receptors superfamily that regulate the gene expression of proteins involved in the control of glucose and lipid metabolism [7]. Indeed, the importance of PPARγ in regulating the insulin sensitivity has motivated research groups in both academia and the pharmaceutical industry to devote increasing efforts toward developing synthetic PPARγ agonists, which could be of therapeutic use in patients affected by T2DM [8]. Thiazolidinediones (TZDs) are one important class of synthetic agonists of PPARγ. TZDs are antidiabetic agents currently used in the treatment of T2DM that target adipose tissue and improve insulin sensitivity. Despite the clinical benefit of these drugs, the use of TZDs has been associated with adverse effects including weight gain, increased adipogenesis, renal fluid retention and possible increased incidence of cardiovascular events [9,10]. Therefore, new PPARγ ligands with enhanced therapeutic efficacy and reduced adverse effects are needed. A promising new class of such ligands is selective PPARγ modulators (*i.e.*, SPPARγMs) [9,10]. These compounds act as partial agonists of PPARγ and display different binding properties in comparison to full agonists [11]. Several plant extracts have been found to increase insulin-stimulated glucose uptake through the action of PPARγ with no or little effect

on adipocyte differentiation [6]. Thus, PPARγ partial agonists from natural extracts are promising candidates for the treatment of T2DM.

Based on the hypothesis that it would be possible to identify PPARγ partial agonists among medicinal extracts previously used as hypoglycemic agents, the goal of the present work was to find natural extracts with known antidiabetic activity that contain at least one molecule that we predict as a PPARγ partial agonist through a virtual screening (VS) workflow that has previously been carefully validated experimentally [12]. Our results provide new information about potential active molecules of natural extracts with antidiabetic properties and their mode of action, *i.e.*, the increase of the insulin-stimulated glucose uptake through the action of PPARγ. We also suggest plants with undescribed antidiabetic activity that may contain PPARγ partial agonists and are related to plants with known antidiabetic activity. These plants represent a potential new source of antidiabetic extracts. In addition, the new PPARγ partial agonists that we have predicted are chemically different from known PPARγ partial agonists and could be used as lead-hopping candidates for the development of new antidiabetic drugs.

## Results and Discussion

**Virtual Screening Description, Validation and Application.** We used a slightly modified version of a VS workflow that was previously developed and validated experimentally [12] to identify PPARγ partial agonists from a large in-house database of compounds. Briefly, the VS used consists of a combination of two pharmacophore modeling methods (*i.e.*, one of them to discard potential PPARγ full agonists and the second one to identify PPARγ partial agonists), a protein-ligand docking and an electrostatic and shape similarity search. The discriminatory power of the VS workflow to identify PPARγ partial agonists was evaluated by applying it to a group of 211 known PPARγ partial agonists obtained from the literature and to 3,122 decoys obtained from the DUD database [13]. See Table 1 for data about how many of these molecules *survived* each VS step. Because we were interested in discovering novel PPARγ partial agonists but not full agonists, we developed an initial structure-based pharmacophore, called the *antipharmacophore*, to exclude possible full agonists. We used this strategy because full agonists present more clearly defined features than partial agonists. Although both types of agonists interact with the ligand-binding domain of PPARγ through several hydrophobic contacts, their mode of binding, and thus their effects, are different [11]. Full agonists are characterized by making a hydrogen-bond network with Ser289, Tyr473, His323 and His449 PPARγ residues, but most partial agonists form a hydrogen bond with Ser342 [11]. In total, 135 known PPARγ partial agonists and 2,204 decoys survived the antipharmacophore step, *i.e.*, they were not identified as potential PPARγ full agonists and served as the input molecules in the next step (Table 1). From the molecules that survived the

antipharmacophore step, 111 known PPARγ partial agonists and 964 decoys were identified as PPARγ partial agonists by our partial agonist pharmacophore (Table 1). To find docking poses that were compatible with the partial agonist pharmacophore, the compounds that had at least one conformer, generated *in vacuo*, that matched with the partial agonist pharmacophore were also docked to the PPARγ structure from 2Q5S. The best docking poses were then matched again to the partial agonist pharmacophore, identifying that 72 out of 111 partial agonists and 382 out 964 decoys that *survived* the previous step have at least one docked pose that simultaneously accomplished the following: (a) compatibility with the PPARγ ligand-binding site; and (b) possession of functional groups that match the 3D location of the sites of the partial agonists pharmacophore. Finally an electrostatic and shape similarity analysis was applied. Using the experimental poses of five known PPARγ partial agonists as queries, 65 out of 72 partial agonists and 102 out 382 decoys were identified as partial agonist candidates by this VS step (Table 1). Overall, our VS workflow identified as partial agonists 65 and 102 out of the initial 211 and 3,122 molecules labeled as partial agonists and decoys, respectively. Therefore, the Enrichment Factor (EF) of the process was 6.15 (a 38.92% of 15.80 that would correspond to the highest possible EF value) and the sensitivity (Se) and the specificity (Sp) were 30.81% and 94.99%, respectively. The high Sp and moderate Se of our procedure reflect the correct assignment of inactive compounds and the loss of potential partial agonists, respectively. However, because of the high number of initial compounds and the difficulties in differentiating partial from full agonists, we preferred a specific, but less sensible, VS workflow. This VS workflow therefore seems adequate to identify molecules with antidiabetic properties that could act as PPARγ partial agonists.

**Table 1**. **Validation and application of the Virtual Screening (VS) workflow.** A dataset of 211 known PPARγ partial agonists and 3,122 decoys extracted from the DUD database were used to validate our VS workflow. Once the VS was validated, it was applied to a dataset of 29,779 natural products (NPs). The numbers represent the number of compounds from each set that *survived* each step when applied sequentially.

| Set of Compounds | Initial Nº of Compounds | Structure-based pharmacophore screening | | | Electrostatic shape similarity analysis |
|---|---|---|---|---|---|
| | | anti pharmacophore | partial agonist pharmacophore | | |
| | | *in vacuo conformations* | *in vacuo conformations* | *docking poses* | |
| Partial Agonists | 211 | 135 | 111 | 72 | 65 |
| Decoys | 3,122 | 2,204 | 964 | 382 | 102 |
| NP database | 29,779 | 21,705 | 2,899 | 935 | 65 |

Once the VS workflow was validated, it was applied to an in-house database formed by 29,779 NPs that contained an annotation of their natural source. After applying the VS workflow described above, a group of 65 PPARγ partial agonist candidates were ultimately identified (see Table 1 for viewing the number of molecules that survived each step of the VS workflow).

**Virtual Screening Hits in Natural Extracts with Known Antidiabetic Activity.** According to the information available in our in-house NP database, the 65 molecules that were predicted by the VS workflow as potential PPARγ partial agonists have been isolated from 74 different natural sources. Interestingly, a systematic bibliographic search of PubMed (http://www.pubmed.org) revealed that 11 out of these 74 natural extracts were described previously as having antidiabetic activity (see Table 2). These 11 extracts contained 12 molecules that we predict to be PPARγ partial agonists, therefore, it is plausible that they could contribute to the observed antidiabetic activity of their corresponding extracts. In fact, a search with SciFinder (http://www.cas.org/products/sfacad) revealed that 6 out of these 12 natural compounds are extremely similar to molecules for which antidiabetic properties have already been described (Table 2 and Figure 1), although no mechanism of action has been suggested for them. This finding validates our methodology and suggests that the mode of action of these molecules could be through PPARγ. The remaining 6 natural compounds not identified previously as antidiabetic molecules represent new molecules with this activity. The most significant compounds found in these 11 antidiabetic extracts will be discussed below:

- The genus name *Salvia* derives from the Latin *salvere* meaning "to save" perhaps referring to the healing properties of plants from this genus. Leafs, roots or flowers from species of *Salvia*, like *Salvia officinalis* [14], *Salvia miltiorrhiza* (a regional Chinese variety) [15], *Salvia fruticosa* [16] and *Salvia lavandulifolia* [17] have been used traditionally worldwide to treat diabetes [1]. The molecules deoxyneocryptotanshinone and miltionone I, which are found in *Salvia miltiorrhiza* extracts (Table 2), were predicted by our VS as PPARγ partial agonists, and they are extremely similar to the main lipophilic diterpene compounds from Danshen (*i.e.*, the dried root of *Salvia miltiorrhiza*), and in particular to tanshinone IIA (see Figure 1A for a comparison of the three structures). Tanshinone IIA enhances low-dose insulin-mediated tyrosine auto-phosphorylation of the insulin receptor β-subunit [18]. Although *Salvia miltiorrhiza* extracts have been shown to have anti-atherosclerotic and antidiabetic properties [15], there is not any evidence that relates the antidiabetic action of the extracts from *Salvia miltiorrhiza* with PPARγ. However, it is known that extracts from the leaves of *Salvia officinalis* activate PPARγ [6]. Deoxyneocryptotanshinone and miltionone I

molecules may be useful for the development of a new class of specific insulin receptor activators that combine this action with the action of PPARγ partial agonists. In addition, we have predicted as PPARγ partial agonists three extra molecules from other extracts of *Salvia* whose species have never been described as antidiabetic: (a) sanigerone from *Salvia lanigera*; (b) 12-hydroxysapriparaquinone from *Salvia prionitis* and *Salvia eriophora* and (c) prionitin from *Salvia prionitis* (Table 3). These molecules are new candidates of PPARγ partial agonists.

- The 2',5,6'-trihydroxy-6,7,8-trimethoxyflavone that is isolated from the roots of four species of plants of the genus *Scutellaria* (*Scutellaria baicalensis*, *Scutellaria adenostegia*, *Scutellaria alpina* and *Scutellaria ramosissima*), was also identified as a PPARγ partial agonist in our VS procedure (Table 2). Extracts from *Scutellaria baicalensis* are prescribed in Kampo medicines in Japan [1], and they are reported to enhance the antidiabetic activity of metformin [19]. Baicalein (5,6,7-trihydroxyflavone), a related compound of the flavone hit, isolated from the roots of *Scutellaria baicalensis* (see Figure 1B for a comparison of both structures), is an α-glucosidase inhibitor [20]. *Scutellaria baicalensis* extracts may therefore contain more than one active component with different modes of antidiabetic action.

- *Cryptolepis sanguinolenta*, a shrub indigenous to West Africa, has been employed by traditional healers in the treatment of various fevers, including malaria [21]. Cryptolepine, an indoloquinolone alkaloid isolated from *Cryptolepis sanguinolenta*, significantly lowers glucose when given orally in a mouse model of diabetes [5], and its antihyperglycemic activity has been demonstrated by several cryptolepine analogs [22]. Cryptolepicarboline is a cryptolepine analog isolated from *Cryptolepis sanguinolenta* [21] (see Figure 1C for a comparison of both structures) that we predict as a PPARγ partial agonist (see Table 2). This result suggests that the increase of glucose uptake caused by cryptolepine and analogous compounds could therefore be mediated by the action of PPARγ.

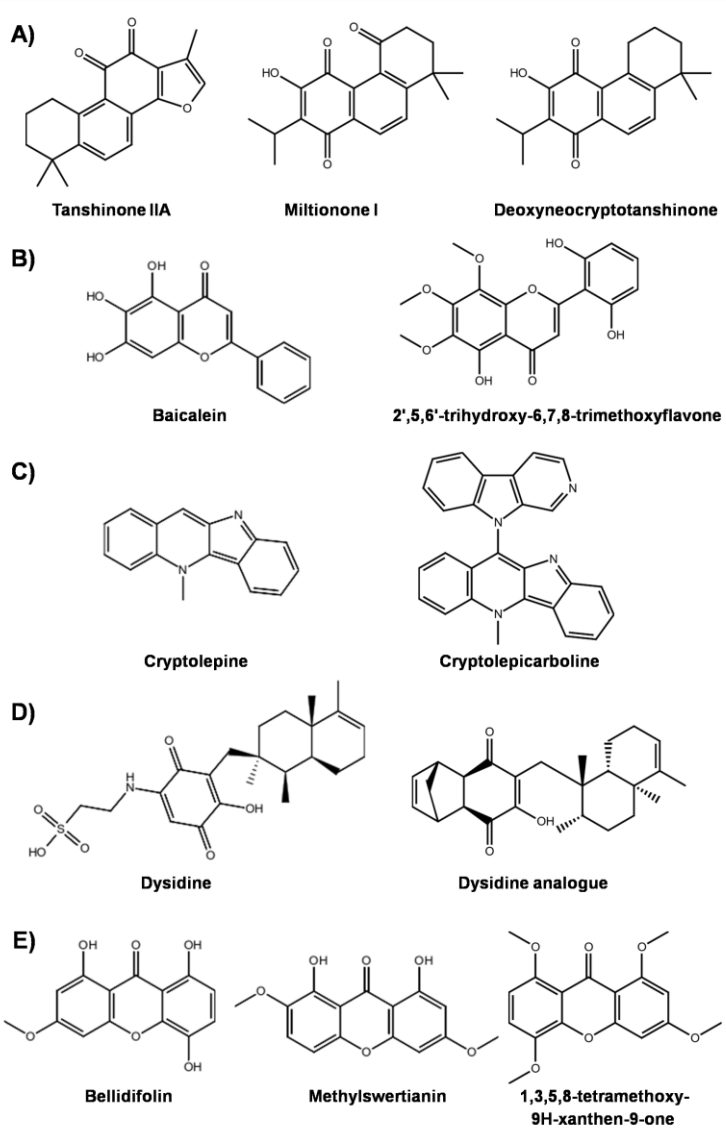


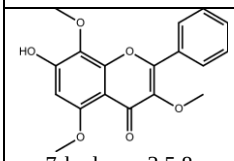
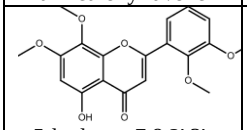
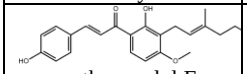
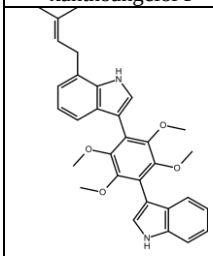
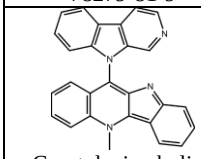

**Figure 1. Chemical comparison between molecules that we predict as PPARγ partial agonists and molecules with described antidiabetic activity.** Each row represents the comparison of the 2D chemical structure between a molecule predicted as PPARγ partial agonists trough our VS workflow and a similar molecule that have been described to present antidiabetic activity.

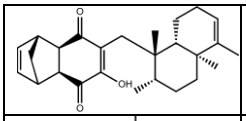
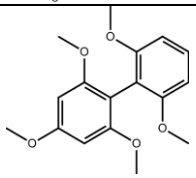
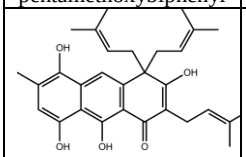
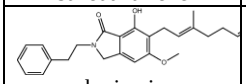
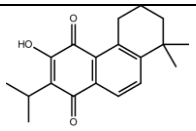
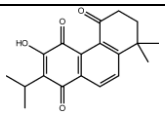
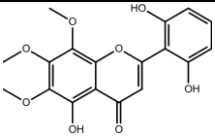
- Dysidine is a sesquiterpene quinone from the marine sponge *Dysidea villosa* that greatly promotes glucose uptake in 3T3-L1 cells and shows strong insulin-sensitizing activity [23]. The results of our VS procedure suggest that an analog of dysidine isolated from *Dysidea villosa* (see Figure 1D for a comparison of both structures) may be a PPARγ partial agonist (Table 2). Although it has been suggested that dysidine exhibits its cellular effects through the activation of the insulin pathway, possibly through the inhibition of protein tyrosine phosphatases [23], it is possible that the mode of action of dysidine and analogous molecules could also be through the action of PPARγ, or that different components of a *Dysidea villosa* extract show antidiabetic activity through different mechanisms. Dysidine and analogous molecules are therefore potential lead compounds for the discovery of new antidiabetic compounds.

- Xanthoangelol F from the Japanese plant *Angelica keiskei* significantly enhances glucose uptake without activating the transactivation activity of PPARγ [24]. This agrees with the results of our VS workflow that suggest that xanthoangelol F acts as a PPARγ partial agonist (Table 2). This compound may therefore belong to the interesting group of PPARγ partial agonists that stimulate glucose uptake without promoting the transactivation activity of PPARγ and avoid some of the problematic side effects of PPARγ full agonists [25].

- The remaining 6 molecules predicted as PPARγ partial agonists through our VS workflow that belong to extracts with described antidiabetic properties (see Table 2) are: (a) 7-hydroxy-3,5,8-trimethoxyflavone from *Achyrocline satureoides*; (b) 5-hydroxy-7,8,2',3'-tetramethoxyflavone from *Andrographis paniculata*; (c) 2,4,6,2',6'-pentamethoxybiphenyl isolated from *Fucus vesiculosus*; (d) hericerin from *Hericium erinaceum*; (e) the molecule with CAS number 78279-81-9 from *Aspergillus terreus*; and (f) bazouanthrone from *Harungana madagascariensis*. Our results suggest that these molecules could be PPARγ partial agonists and that extracts containing these molecules could stimulate glucose uptake through the action of PPARγ. This information is novel and relevant because it is the first time that antidiabetic properties for these molecules have been suggested.

**Table 2.** Natural extracts with described antidiabetic activity that contain one molecule that is predicted to be a PPARγ partial agonist by our virtual screening protocol (identified by their 2D structure and, when available, their common name or CAS number). The bibliographic references for each extract are split in three columns where (a) the fifth column reports papers that describe the purification of each molecule from the corresponding extract; (b) the sixth column reports papers that describe the antidiabetic activity of the corresponding extract; and (c) the seventh column reports papers that describe the antidiabetic activity of the corresponding molecule or similar molecules (when available). The second column represents the number of the cluster that each molecule belongs when they were compared with a group of 211 synthetic PPARγ partial agonists.

| Molecule CAS number or Name | Clus -ter | Extract | Kingdom Family | Ref. Isolation Molecule from Extract | Ref. Antidia betic Extract | Ref. Antidia betic Molecul e |
|---|---|---|---|---|---|---|
|  7-hydroxy-3,5,8-trimethoxyflavone | 8 | *Achyrocline satureoides* | Plantae Asteracea e | [35] | [36] | - |
|  5-hydroxy-7,8,2',3'-tetramethoxyflavone | 8 | *Andrographis paniculata* | Plantae Acanthace ae | [37,38] | [39] | - |
|  xanthoangelol F | 7 | *Angelica keiskei* | Plantae Apiaceae | [40,41] | [21] | [21] |
|  78279-81-9 | 7 | *Aspergillus terreus* | Fungi Trichoco maceae | [42] | [43] | - |
|  Cryptolepicarboline | 24 | *Cryptolepis sanguinolenta* | Plantea Apocynac eae | [17,44] | [19] | [19] |

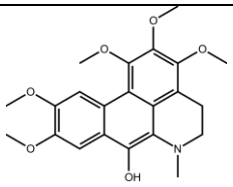
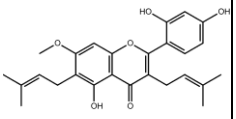
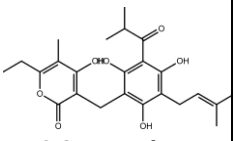
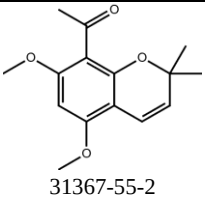
| | | | | | | |
|---|---|---|---|---|---|---|
|  | 1 | *Dysidea villosa* | Animalia Dysideidea | [45] | [20] | [20] |
|  2,4,6,2',6'-pentamethoxybiphenyl | 8 | *Fucus vesiculosus* | Chromalveolata Fucaceae | [46] | [47] | - |
|  bazouanthrone | 3 | *Harungana madagascariensis* | Plantae Hypericaceae | [48] | [49] | - |
|  hericerin | 7 | *Hericium erinaceum* | Fungi Hericiaceae | [50] | [51] | - |
|  deoxyneocryptotanshinone | 2 | *Salvia miltiorrhiza* | Plantae Lamiaceae | [52] | [11] | [14] |
|  miltionone I | 2 | *Salvia miltiorrhiza* | Plantae Lamiaceae | [53] | [11] | [14] |
|  2',5,6'-trihydroxy-6,7,8-trimethoxyflavone | 8 | *Scutellaria baicalensis* | Plantae Lamiaceae | [54] | [15] | [55] |

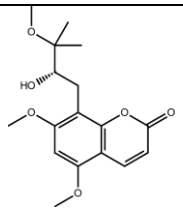
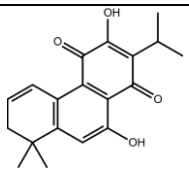
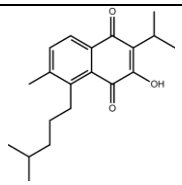
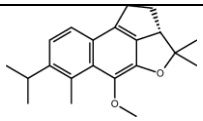
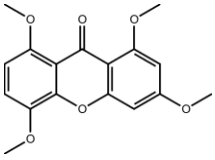
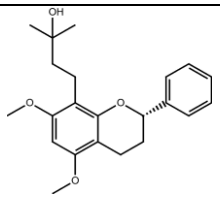
Taking into account the fact that extracts from closer species of the same genus may share a high number of components, we also look for species that contain a molecule that we predict as a PPARγ partial agonist and, although they have been not described previously as antidiabetic, they are related (*i.e.*, they belong to the same genus) to species with known antidiabetic properties. Thus, we identified 10 molecules isolated from 16 different plants, such as *Acradenia franklinii*, *Annona purpurea*, *Artocarpus gomezianus*, *Euodia lunuankenda*, *Evodia elleryana*, *Helichrysum mixtum*, *Helichrysum odoratissimum*, *Helichrysum stenopterum*, *Melicope ptelefolia*, *Melicope simplex*, *Murraya paniculata*, *Salvia eriophora*, *Salvia lanigera*, *Salvia prionitis*,

*Swertia hookeri* and *Tephrosia watsoniana* (Table 3), whose extracts could show antidiabetic properties mediated by the action of PPARγ. For example, the 1,3,5,8-tetramethoxy-9H-xanthen-9-one from *Swertia hookeri* was identified as a PPARγ partial agonist by our VS (see Table 3). However, neither the molecule nor an extract from this species has been identified previously as an antidiabetic agent. Nevertheless, the whole plants of *Swertia japonica* and *Swertia chirayita* have been reported to exhibit hypoglycemic effects by oral administration, and the xanthone constituents, bellidifolin and methylswertianin, have been isolated as active constituents [26-28]. Methylswertianin and bellidifolin are molecules highly similar to the tetramethoxyxanthon from *Swertia hookeri* that we identified as a PPARγ partial agonist (see Figure 1E for a comparison of their structures). Our results therefore suggest that the antidiabetic action of *Swertia* species could be mediated at least in part by PPARγ.

The 22 molecules from Tables 2 and 3 that we predict to be PPARγ partial agonists are also interesting. To compare their chemical structures with known PPARγ partial agonists, we merged their structures with 211 structures of known PPARγ partial agonists obtained from the literature. The resulting set was classified into 26 clusters according to structure similarity. The 22 NP hits of our VS were classified into 12 clusters. None of these clusters contained any of the 211 known PPARγ partial agonists that were previously merged with the VS hits. Thus, these 22 predicted PPARγ partial agonists represent 12 different chemical scaffolds that are different from the ones present in known synthetic PPARγ partial agonists. Therefore, these scaffolds are lead-hoping candidates for searching for new PPARγ partial agonists.

**Table 3.** Natural extracts that contain one molecule predicted to be a PPARγ partial agonist by our VS protocol and that are related to natural extracts that are described to have antidiabetic activity. Table shows the natural extracts (*i.e.*, third column), the VS hits that have been purified from them (identified by their 2D structure and, when available, their common name or CAS number) and that are the related extracts with described antidiabetic activity (*i.e.*, sixth column). The bibliographic references for each extract are split in three columns where (a) the fifth column reports papers that describe the purification of each molecule from the corresponding extract (*i.e.*, ref.1); (b) the seventh column reports papers that describe the antidiabetic activity of the related extract (see sixth column, ref.2); and (c) the eighth column reports papers that describe the antidiabetic activity of the corresponding or similar molecules when available (*i.e.*, ref.3). The second column represents the number of the cluster to which each molecule belongs when they were compared with a group of 211 synthetic PPARγ partial agonists.

| **Molecule CAS number or Name** | **Cluster** | **Extract** | **Kingdom Family** | **Ref. 1** | **Antidiabetic Extract** | **Ref. 2** | **Ref. 3** |
|---|---|---|---|---|---|---|---|
|  7-hydroxydehydro-thalicsimidine | 12 | *Annona purpurea* | Plantae Annonaceae | [56] | *Annona squamosa* | [57] | - |
|  artocarpin | 7 | *Artocarpus gomezianus* | Plantae Moraceae | [58] | *Artocarpus heterophyllus* | [59] | - |
|  6-O-Desmethy-lauricepyron | 4 | *Helichrysum stenopterum* | Plantae Asteraceae | [60] | *Helichrysum plicatum* | [61] | - |
| | | *Helichrysum odoratissimum* | | [62] | | | |
| | | *Helichrysum mixtum* | | [60] | *Helichrysum graveolens* | | |
|  31367-55-2 | 5 | *Melicope ptelefolia* | Plantae Rutaceae | [63] | *Evodia officinalis* | [64] | - |
| | | *Melicope simplex* | Plantae Rutaceae | [65] | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
|  - | 9 | *Murraya paniculata* | Plantae Rutaceae | [66] | *Murraya koeingii* | [67] | - |
|  sanigerone | 2 | *Salvia lanigera* | Plantae Lamiaceae | [68] | *Salvia lavandulifolia* | [13] | - |
|  12-hydroxysapri-paraquinone | 2 | *Salvia prionitis* | Plantae | [69] | *Salvia officinalis* | [10] | - |
| | | *Salvia eriophora* | Plantae Lamiaceae | [70] | | | |
|  prionitin | 11 | *Salvia prionitis* | Plantae Lamiaceae | [71] | *Salvia fruticosa* | [12] | - |
|  1,3,5,8-tetramethoxy-9H-xanthen-9-one | 8 | *Swertia hookeri* | Plantae Gentianaceae | [72] | *Swertia punicea* | [23] | [23] |
| | | | | | *Swertia japonica* | [24] | [73] |
| | | | | | *Swertia chirayita* | [25] | - |
| | | | | | *Swertia paniculata* | [74] | - |
|  nitenin | 10 | *Tephrosia watsoniana* | Plantae Fabaceae | [75] | *Tephrosia purpurea* | [76] | - |

## Conclusion

We have applied an experimentally validated VS workflow based on (a) two structure-based pharmacophores, (b) protein-ligand docking and (c) an electrostatic/shape similarity analysis to identify NPs that may be novel scaffolds for the discovery of new PPARγ partial agonists. Thus, from an initial set of 29,779 NPs that are annotated with their natural source, we predict 22 molecules to be potential PPARγ partial agonists. A subset of 12 of these molecules are present in 11 natural extracts with known antidiabetic activity and 10 of them are present in extracts related (*i.e.*, they are from species of the same genus) to plants with known antidiabetic activity. None of the 22 hits show chemical similarity with 211 known PPARγ partial agonists obtained from the literature and, therefore, are new chemical scaffold candidates for the development of PPARγ partial agonists. Moreover, our results provide a new hypothesis about the active molecules of natural extracts with antidiabetic properties and their mode of action, *i.e.*, the insulin-stimulated glucose uptake is increased through the action of PPARγ. We also suggest plants with undescribed antidiabetic activity that may contain PPARγ partial agonists and are related to plants with known antidiabetic activity. These plants represent a new source of potential antidiabetic extracts. Consequently, our work opens the door to the discovery of new antidiabetic extracts and molecules that can be of use, for instance, in the design of new antidiabetic drugs or functional foods focused towards the prevention/treatment of T2DM.

## Experimental section

**Initial Dataset of Natural Compounds Used.** The initial in-house dataset of natural compounds that was filtered through the VS contained 29,779 compounds annotated with the natural sources from which they were obtained and the bibliographic references that describe how to extract them from each natural source. Moreover, according to the FAF-Drugs2 program [29], all of these molecules (a) show good ADME properties according to the Lipinski rule of five [30] (*i.e.*, only one violation of this rule was allowed) and (b) are not potentially toxic (*i.e.*, they lack "warhead" chelators, frequent hitters, promiscuous inhibitors and other undesirable functional groups). Conformations and sites for the 3D structures of these 29,779 compounds were determined during the generation of the corresponding Phase v3.1 (Schrödinger LLC., Portland, USA; http://www.schrodinger.com) [31] databases with the *Generate Phase Database* graphic front-end. The parameter values used during this conformer generation were the default values, with the exception of the maximum number of conformers per structure, which was increased from 100 (the default value) to 200. The conformer sites were generated with definitions made by adding the ability to consider aromatic rings as hydrophobic groups to the default built-in Phase definitions.

**Virtual Screening Workflow.** The VS workflow used in this work is a slightly modified version of a VS workflow developed previously (that was also validated experimentally) to identify PPARγ partial agonists in chemical databases [12]. Briefly, the VS workflow consisted of several steps that must be applied sequentially (*i.e.*, the output molecules of one step were the input molecules for the next step). Thus, the filters applied (and sorted according their usage) were the following: (1) a structure-based antipharmacophore screening; (2) a structure-based pharmacophore screening (called partial agonist pharmacophore); and (3) an electrostatic/shape similarity analysis (the previously developed VS workflow was altered for the current work with lower threshold values for the electrostatic and shape comparisons; see below for more details). Moreover, all of the PDB files used in that work were superposed with the DeepView v3.7 program (http://spdbv.vital-it.ch/) [32] to ensure that all of them had the same relative orientation. From then on, only the resulting re-oriented coordinates for these PDB files were used during the subsequent structure-based pharmacophore generation and in the steps of the VS workflow where spatial orientation is crucial (*i.e.*, pharmacophore-based searches, protein-ligand docking studies and shape and electrostatic-potential comparisons).

The initial set of compounds was filtered by a structure-based antipharmacophore with the aim of discarding potential PPARγ full agonists. This pharmacophore is formed by 5 sites (two hydrogen-bond acceptors and three hydrophobic sites) that are present in most of the validated 19 complexes of full agonists (where *validated* means that the coordinates for the ligand and the PPARγ active site are reliable according to their corresponding electron density map) and is completed with receptor-based excluded volumes obtained from the PDB file coded as 1FM9. Thus, this filter removed from the sample those molecules that had at least one *in vacuo*-generated conformer that matched at least 4 out of 5 sites of the antipharmacophore. The fitting between the molecules and the pharmacophore was analyzed with Phase v3.1 [31]. The subset of molecules that did not match the antipharmacophore was then used to identify possible partial agonists. To accomplish this task, a second pharmacophore obtained from the common sites of 12 validated complexes between PPARγ, and a partial agonist was used. It consists of one hydrogen-bond acceptor and three hydrophobic sites with receptor-based excluded volumes obtained from the PDB file coded as 2Q5S. Molecules that had at least one *in vacuo*-generated conformer and that matched with the 4 sites of the partial agonist pharmacophore were initially identified as putative PPARγ partial agonists. To find docking poses that were compatible with the partial agonist pharmacophore, those molecules identified as putative PPARγ partial agonists were then docked to the ligand-binding site of 2Q5S. Thus, the best 32 docked poses predicted by the eHiTS v2009 program (SimBioSys Inc., Toronto, Canada) [33] were filtered again with Phase through the partial agonist pharmacophore, using the same filtering options as the first pharmacophore matching, except no re-orientation of the poses was allowed during the search.

The poses that passed the pharmacophore and docking screenings were submitted to an electrostatic/shape similarity analysis, using the PPARγ partial agonists crystallized in the structures 2G0H, 4PRG, 2Q5S, 2FVJ and 2Q6S as a queries. This analysis was performed with EON v2.0.1 (OpenEye Scientific Software, Inc., Santa Fe, New Mexico, USA; http://www.eyesopen.com), using the Electrostatic Tanimoto combo (ET_combo) score as the similarity criteria. The ET_combo score is the sum of two calculations: (a) the Shape Tanimoto (ST) score, which is a quantitative measure of three-dimensional overlap (where 1 corresponds to a perfect overlap, *i.e.*, the same shape) and (b) the Possion-Boltzman Electrostatic Tanimoto (ET_pb) score that compares the electrostatic potential of two small molecules and ranges from 1 (identical potential) to negative values that results from the overlap of positive and negative charges. In this work, we selected the EON thresholds taking into account the results of the comparison between a group of experimental poses for PPARγ partial agonists in their complexes with PPARγ. Applying the five query poses against twelve other experimental poses, the lowest values for the ET_pb score and ST were 0.2 and 0.4, respectively. Therefore, both values were used as thresholds during the VS electrostatic/shape similarity analysis (where the thresholds used in the original VS workflow were 0.3 and 0.5 for ET_pb and ST, respectively [12].

**Virtual Screening Workflow Validation.** The ability of the VS workflow to identify PPARγ partial agonists was tested by applying it to a group of 211 known PPARγ partial agonists obtained from the literature and 3,122 decoys obtained from the DUD database [13]. The structures of the 211 partial agonists were built with ChemDraw Ultra v11.0 (CambridgeSoft Corporation, Cambridge, MA, USA; http://www.cambridgesoft.com/) [34] and were cleaned using LigPrep v2.3 (Schrödinger LLC., Portland, USA; http://www.schrodinger.com). We calculated an EF and values for Se and Sp for the global VS workflow [35]. The EF was obtained as the quotient between the fraction of actives in the sample that *survived* the VS workflow and the fraction of actives that were initially in the sample. The EF therefore represents the ratio of the number of actives actually retrieved by a method compared to the number expected purely by chance. Se describes how well the model correctly identifies active compounds and it is calculated as the ratio between the number of active molecules that *survived* the VS workflow and the number of all active compounds that were initially in the sample. Sp measures the correct assignment of inactive compounds. It is calculated as the ratio between the number of inactive compounds that were discarded in the VS workflow and the number of all of the inactive molecules that were initially in the sample.

**Structural Similarity Analysis.** To obtain new scaffolds for PPARγ partial agonists, the VS hits were merged with the 211 PPARγ partial agonists previously used for validating the VS workflow and clustered with Canvas v1.2 (Schrödinger LLC., Portland, USA; http://www.schrodinger.com). Using a fingerprint precision of 32 bits,

MOLPRINT2D fingerprints [36] were calculated for each molecule, and then a hierarchical clustering, based on Tanimoto similarities, was obtained. The number of clusters obtained was defined using the Kelley criterion [37].

## References

1. Howes M, Simmonds M. (2005) Plants used in the treatment of diabetes. In: Soumyanath A, editor. Traditional Medicines for Modern Times. Press.

2. Bnouham M, Ziyyat A, Mekhfi H, Tahri A, Legssyer A. (2006) Medicinal plants with potential antidiabetic activity - A review of ten years of herbal medicine research (1990-2000). Int J Diabetes & Metabolism 14(1).

3. Modak M, Dixit P, Londhe J, Ghaskadbi S, Paul A, Devasagayam T. (2007) Indian herbs and herbal drugs used for the treatment of diabetes. Journal of Clinical Biochemistry and Nutrition 40(3): 163-173.

4. Grover JK, Yadav S, Vats V. (2002) Medicinal plants of india with anti-diabetic potential. J Ethnopharmacol 81(1): 81-100.

5. Luo J, Fort DM, Carlson TJ, Noamesi BK, nii-Amon-Kotei D, et al. (1998) Cryptolepis sanguinolenta: An ethnobotanical approach to drug discovery and the isolation of a potentially useful new antihyperglycaemic agent. Diabetic Medicine: A Journal of the British Diabetic Association 15(5): 367-374.

6. Christensen KB, Minet A, Svenstrup H, Grevsen K, Zhang H, et al. (2009) Identification of plant extracts with potential antidiabetic properties: Effect on human peroxisome proliferator-activated receptor (PPAR), adipocyte differentiation and insulin-stimulated glucose uptake. Phytother Res 23(9): 1316-1325.

7. Berger J, Moller DE. (2002) The mechanisms of action of PPARs. Annu Rev Med 53: 409-435.

8. Shearer BG, Billin AN. (2007) The next generation of PPAR drugs: Do we have the tools to find them? Biochim Biophys Acta 1771(8): 1082-1093.

9. Feldman PL, Lambert MH, Henke BR. (2008) PPAR modulators and PPAR pan agonists for metabolic diseases: The next generation of drugs targeting peroxisome proliferator-activated receptors? Current Topics in Medicinal Chemistry 8(9): 728-749.

10. Pourcet B, Fruchart JC, Staels B, Glineur C. (2006) Selective PPAR modulators, dual and pan PPAR agonists: Multimodal drugs for the treatment of type 2 diabetes and atherosclerosis. Expert Opin Emerg Drugs 11(3): 379-401.

11. Guasch L, Sala E, Valls C, Blay M, Mulero M, et al. (2011) Structural insights for the design of new PPARgamma partial agonists with high binding affinity and low transactivation activity. J Comput Aided Mol Des.*Accepted*

12. Guasch L, Sala E, Castell-Auvi A, Cedo L, Liedl K, et al. Identification of PPARgamma partial agonists of natural origin (part I): Development of a virtual screening procedure and *in vitro* validation.*Submitted*

13. Huang N, Shoichet BK, Irwin JJ. (2006) Benchmarking sets for molecular docking. J Med Chem 49(23): 6789-6801.

14. Eidi M, Eidi A, Zamanizadeh H. (2005) Effect of salvia officinalis L. leaves on serum glucose and insulin in healthy and streptozotocin-induced diabetic rats. J Ethnopharmacol 100(3): 310-313.

15. Tan Y, Kamal MA, Wang Z, Xiao W, Seale JP, et al. (2011) Chinese herbal extracts (SK0506) as a potential candidate for the therapy of the metabolic syndrome. Clinical Science 120(7): 297-305.

16. Perfumi M, Arnold N, Tacconi R. (1991) Hypoglycemic activity of salvia fruticosa mill from cyprus. J Ethnopharmacol 34(2-3): 135-140.

17. Jimenez J, Risco S, Ruiz T, Zarzuelo A. (1986) Hypoglycemic activity of salvia lavandulifolia. Planta Med (4): 260-262.

18. Jung SH, Seol HJ, Jeon SJ, Son KH, Lee JR. (2009) Insulin-sensitizing activities of tanshinones, diterpene compounds of the root of salvia miltiorrhiza bunge. Phytomedicine 16(4): 327-335.

19. Waisundara VY, Hsu A, Huang D, Tan BK. (2008) Scutellaria baicalensis enhances the anti-diabetic activity of metformin in streptozotocin-induced diabetic wistar rats. Am J Chin Med 36(3): 517-540.

20. Nishioka T, Kawabata J, Aoyama Y. (1998) Baicalein, an alpha-glucosidase inhibitor from scutellaria baicalensis. J Nat Prod 61(11): 1413-1415.

21. Sharaf MHM, Schiff PL, Tackie AN, Phoebe CH, Howard L, et al. (1995) Submicromole structure elucidation: Cryptolepicarbolinea novel dimeric alkaloid from Cryptolepis sanguinolenta. Magn Reson Chem 33(10): 767-778.

22. Bierer DE, Dubenko LG, Zhang P, Lu Q, Imbach PA, et al. (1998) Antihyperglycemic activities of cryptolepine analogues: An ethnobotanical lead structure isolated from cryptolepis sanguinolenta. J Med Chem 41(15): 2754-2764.

23. Zhang Y, Li Y, Guo Y, Jiang H, Shen X. (2009) A sesquiterpene quinone, dysidine, from the sponge dysidea villosa, activates the insulin pathway through inhibition of PTPases. Acta Pharmacol Sin 30(3): 333-345.

24. Enoki T, Ohnogi H, Nagamine K, Kudo Y, Sugiyama K, et al. (2007) Antidiabetic activities of chalcones isolated from a japanese herb, angelica keiskei. J Agric Food Chem 55(15): 6013-6017.

25. Choi JH, Banks AS, Estall JL, Kajimura S, Bostram P, et al. (2010) Anti-diabetic drugs inhibit obesity-linked phosphorylation of PPARgamma by Cdk5. Nature 466(7305): 451-456.

26. Tian L, Bai X, Chen X, Fang J, Liu S, et al. (2010) Anti-diabetic effect of methylswertianin and bellidifolin from swertia punicea hemsl. and its potential mechanism. Phytomedicine: International Journal of Phytotherapy and Phytopharmacology 17(7): 533-539.

27. Basnet P, Kadota S, Shimizu M, Namba T. (1994) Bellidifolin: A potent hypoglycemic agent in streptozotocin (STZ)-induced diabetic rats from swertia japonica. Planta Med 60(6): 507-511.

28. Chandrasekar B, Bajpai MB, Mukherjee SK. (1990) Hypoglycemic activity of swertia chirayita (roxb ex flem) karst. Indian J Exp Biol 28(7): 616-618.

29. Lagorce D, Sperandio O, Galons H, Miteva MA, Villoutreix BO. (2008) FAF-Drugs2: Free ADME/tox filtering tool to assist drug discovery and chemical biology projects. Bioinformatics 9: 396.

30. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 46(1-3): 3-26.

31. Dixon SL, Smondyrev AM, Knoll EH, Rao SN, Shaw DE, et al. (2006) PHASE: A new engine for pharmacophore perception, 3D QSAR model development, and 3D

database screening: 1. methodology and preliminary results. J Comput Aided Mol Des 20(10-11): 647-671.

32. Guex N, Peitsch MC. (1997) SWISS-MODEL and the swiss-PdbViewer: An environment for comparative protein modeling. Electrophoresis 18(15): 2714-2723.

33. Zsoldos Z, Reid D, Simon A, Sadjad SB, Johnson AP. (2007) eHiTS: A new fast, exhaustive flexible ligand docking system. Journal of Molecular Graphics & Modelling 26(1): 198-212.

34. Cousins KR. (2011) Computer review of ChemDraw ultra 12.0. J Am Chem Soc 133(21): 8388.

35. Schuster D, Wolber G. (2010) Identification of bioactive natural products by pharmacophore-based virtual screening. Curr Pharm Des 16(15): 1666-1681.

36. Duan J, Dixon SL, Lowrie JF, Sherman W. (2010) Analysis and comparison of 2D fingerprints: Insights into database screening performance using eight fingerprint methods. Journal of Molecular Graphics & Modelling 29(2): 157-170.

37. Kelley LA, Gardner SP, Sutcliffe MJ. (1996) An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies. Protein Eng 9(11): 1063-1065.

38. Mesquita AAL, Correa DDB, De Padua AP, Guedes MLO, Gottlieb OR. (1986) Flavonoids from four compositae species. Phytochemistry 25(5): 1255-1256.

39. Arredondo MF, Blasina F, Echeverry C, Morquio A, Ferreira M, et al. (2004) Cytoprotection by achyrocline satureioides (lam) D.C. and some of its main flavonoids against oxidative stress. J Ethnopharmacol 91(1): 13-20.

40. Koteswara Rao Y, Vimalamma G, Venkata Rao C, Tzeng Y. (2004) Flavonoids and andrographolides from andrographis paniculata. Phytochemistry 65(16): 2317-2321.

41. Malik A, Jahan N, Muhammad P. (2001) New flavonoid from mentha longifolia. Heterocycles 55(10): 1951.

42. Wibudi A, Kiranadi B, Manalu W, winarto A, Suyono S. (2008) The traditional plant, andrographis paniculata (sambiloto) exhibits insulin-releasing actions in vitro. Acta Medica Indonesiana 40(2): 63-68.

43. Nishimura R, Tabata K, Arakawa M, Ito Y, Kimura Y, et al. (2007) Isobavachalcone, a chalcone constituent of Angelica keiskei induces apoptosis in neuroblastoma. Biol Pharm Bull 30(10): 1878-1883.

44. Matsuura M, Kimura Y, Nakata K, Baba K, Okuda H. (2001) Artery relaxation by chalcones isolated from the roots of angelica keiskei. Planta Med 67(3): 230-235.

45. Arai K, Shimizu S, Yamamoto Y. (1981) Metabolic products of aspregillus terreus. VI. metabolites of the strain IFO 8835. (3). the isolation and chemical structures of colorless metabolites. Chemical & Pharmaceutical Bulletin 29(4): 1005-1012.

46. Dewi RT, Iskandar YM, Hanafi M, Kardono LBS, Angelina M, et al. (2007) Inhibitory effect of koji aspergillus terreus on alpha-glucosidase activity and postprandial hyperglycemia. Pakistan Journal of Biological Sciences 10(18): 3131-3135.

47. Sharaf MHM, Schiff PL, Tackie AN, Phoebe CH, Johnson RL, et al. (1996) The isolation and structure determination of cryptomisrine, a novel indolo[3,2-b]quinoline dimeric alkaloid from Cryptolepis sanguinolenta. J Heterocycl Chem 33(3): 789-797.

48. Stewart M, Fell PM, Blunt JW, Munro MHG. (1997) Avarol and related compounds from the new zealand marine sponge dysidea sp. Aust J Chem 50(4): 341-348.

49. Glombitza K, Lentz G. (1981) Antibiotics from algae--XXVIII: Cleavage of high molecular phlorotannin derivatives from the brown alga fucus vesiculosus L. Tetrahedron 37(22): 3861-3866.

50. Lamela M, Anca J, Villar R, Otero J, Calleja JM. (1989) Hypoglycemic activity of several seaweed extracts. J Ethnopharmacol 27(1-2): 35-43.

51. Lenta BN, Ngouela S, Boyom FF, Tantangmo F, Tchouya GRF, et al. (2007) Anti-plasmodial activity of some constituents of the root bark of \itHarungana madagascariensis LAM. (hypericaceae). Chem Pharm Bull 55(3): 464-467.

52. Agbor GA, Kuate D, Oben JE. (2007) Medicinal plants can be good source of antioxidants: Case study in cameroon. Pakistan Journal of Biological Sciences 10(4): 537-544.

53. Yasuo K, Masahiko N, Hiromitsu N, Takashi H, Atsumi S, et al. (1991) Hericerin, a new pollen growth inhibitor from the mushroom hericium erinaceum. Agric Biol Chem 55(10): 2673-2674.

54. Fujiwara M, Egashira N, Mishima K. (2006) Neuroprotective effect of hericium erinaceum. Foods Food Ingredients J Jpn 211(2): 141-147.

55. Ikeshiro Y, Hashimoto I, Iwamoto Y, Mase I, Tomita Y. (1991) Diterpenoids from salvia miltiorrhiza. Phytochemistry 30(8): 2791-2792.

56. Ikeshiro Y, Mase I, Tomita Y. (1989) Abietane type diterpenoids from salvia miltiorrhiza. Phytochemistry 28(11): 3139-3141.

57. Chemesova II, Iinuma M, Budantsev AL. (1993) Investigation of the flavonoid composition of scutellaria adenostegia. Chemistry of Natural Compounds 29(1): 133-134.

58. Prabhakar PK, Doble M. (2011) Mechanism of action of natural products used in the treatment of diabetes mellitus. Chinese Journal of Integrative Medicine 17(8): 563-574.

59. Chang F, Wei J, Teng C, Wu Y. (1998) Two new 7-dehydroaporphine alkaloids and antiplatelet action aporphines from the leaves of annona purpurea. Phytochemistry 49(7): 2015-2018.

60. Shirwaikar A, Rajendran K, Dinesh Kumar C, Bodla R. (2004) Antidiabetic activity of aqueous leaf extract of annona squamosa in streptozotocin-nicotinamide type 2 diabetic rats. J Ethnopharmacol 91(1): 171-175.

61. Likhitwitayawuid K, Sritularak B, De-Eknamkul Wanchai. (2000) Tyrosinase inhibitors from artocarpus gomezianus. Planta Med 66(3): 275-277.

62. Fernando MR, Thabrew MI, Karunanayake EH. (1990) Hypoglycaemic activity of some medicinal plants in sri-lanka. Gen Pharmacol 21(5): 779-782.

63. Jakupovic J, Kuhnke J, Schuster A, Metwally MA, Bohlmann F. (1986) Phloroglucinol derivatives and other constituents from south african helichrysum species. Phytochemistry 25(5): 1133-1142.

64. Aslan M, Deliorman Orhan D, Orhan N, Sezik E, Yesilada E. (2007) In vivo antidiabetic and antioxidant potential of helichrysum plicatum ssp. plicatum capitulums in streptozotocin-induced-diabetic rats. J Ethnopharmacol 109(1): 54-59.

65. Hänsel R, Cybulski E, Çubuku B, Meriçli AH, Bohlmann F, et al. (1980) Neue pyron-derivate aus helichrysum-arten. Phytochemistry 19(4): 639-644.

66. Kamperdick C, Van NH, Sung TV, Adam G. (1997) Benzopyrans from melicope ptelefolia leaves. Phytochemistry 45(5): 1049-1056.

67. Yeo J, Kang Y, Cho S, Jung M. (2011) Effects of a multi-herbal extract on type 2 diabetes. Chinese Medicine 6: 10.

68. Briggs LH, Locker RH. (1950) 486. chemistry of new zealand melicope species. part IV. constituents of the bark of melicope simplex. J.Chem.Soc. : 2376-2379.

69. Kinoshita T, Wu JB, Ho FC. (1996) Prenylcoumarins from murraya paniculata var. omphalocarpe (rutaceae): The absolute configuration of sibiricin, mexoticin and omphamurin. ChemInform 27(47): 1208-1211.

70. El-lakany AM. (2003) Two new diterpene quinones from the roots of salvia lanigera poir. Pharmazie 58(1): 75-76.

71. Li M, Zhang J, Chen M. (2001) A novel dimeric diterpene from salvia prionitis. J Nat Prod 64(7): 971-972.

72. Ulubelen A, Birman H, Oksüz S, Topçu G, Kolak U, et al. (2002) Cardioactive diterpenes from the roots of Salvia eriophora. Planta Med 68(9): 818-821.

73. Blasko G, Lin LZ, Cordell GA. (1988) Determination of a new tetracyclic diterpene skeleton through selective INEPT spectroscopy. J Org Chem 53(26): 6113-6115.

74. Ghosal S, Biswas K, Jaiswal DK. (1980) Xanthone and flavonol constituents of swertia hookeri. Phytochemistry 19(1): 123-126.

75. Basnet P, Kadota S, Shimizu M, Takata Y, Kobayashi M, et al. (1995) Bellidifolin stimulates glucose uptake in rat 1 fibroblasts and ameliorates hyperglycemia in streptozotocin (STZ)-induced diabetic rats. Planta Med 61(5): 402-405.

76. Negi JS, Singh P, Pant GJn, Rawat MSM. (2010) RP-HPLC analysis and antidiabetic activity of swertia paniculata. Natural Product Communications 5(6): 907-910.

77. Gomez F, Quijano L, Calderon JS, Rodriquez C, Rios T. (1985) Prenylflavans from tephrosia watsoniana. Phytochemistry 24(5): 1057-1059.

78. Pavana P, Manoharan S, Renju GL, Sethupathy S. (2007) Antihyperglycemic and antihyperlipidemic effects of tephrosia purpurea leaf extract in streptozotocin induced diabetic rats. Journal of Environmental Biology 28(4): 833-837.

CHAPTER 5

# IDENTIFICATION OF NOVEL HUMAN DIPEPTIDYL PEPTIDASE-IV INHIBITORS OF NATURAL ORIGIN: VIRTUAL SCREENING AND ACTIVITY ASSAYS

## ABSTRACT

The large scaffold diversity and properties of natural products, such as structural complexity and drug similarity, form the basis of claims that these molecules are ideal starting points for drug design and development. Consequently, there has been great interest in determining whether natural products show biological activity toward protein targets of pharmacological relevance. One target of particular interest is DPP-IV, a serine protease that specifically removes N-terminal dipeptides from substrates containing proline or alanine as the second residue. The most important substrates of DPP-IV are incretins whose production, among other beneficial effects, stimulates insulin biosynthesis and secretion. Incretins have very short half-lives because of their rapid degradation by DPP-IV and, therefore, inhibiting this enzyme prolongs the actions of incretins and improves glucose homeostasis. As a result, DPP-IV inhibitors are of considerable interest to the pharmaceutical industry. The main goals of this study were to (a) use virtual screening to identify potential DPP-IV inhibitors of natural origin and (b) evaluate the reliability of our virtual-screening protocol by experimentally testing the *in vitro* activity of selected natural-product hits.

We predicted that 446 out of the 89,425 molecules present in the natural products subset of the ZINC database would inhibit DPP-IV with good ADMET properties. Notably, when these 446 molecules were merged with 2,571 known DPP-IV inhibitors and the resulting set was classified into 50 clusters according to chemical similarity, there were 12 clusters that contained only natural products for which no DPP-IV inhibitory activity has been previously reported. Nine molecules from 7 of these 12 clusters were then selected for *in vitro* activity testing and 7 out of the 9 molecules were shown to inhibit DPP-IV.

We have demonstrated that our virtual-screening protocol was successful in identifying lead compounds for developing new inhibitors for DPP-IV, a target of great interest in medicinal chemistry.

## Introduction

Type 2 diabetes mellitus (T2DM) is considered to be the "epidemic of the 21st century" and, consequently, the development of new therapies is one of the main challenges in drug discovery today [1]. While current T2DM therapies that increase insulin secretion have proven to have beneficial therapeutic effects, these treatments often suffer from undesirable side effects such as hypoglycemia and weight gain [2]. Therefore, there is a significant unmet medical need for better drugs to treat T2DM.

Recently, the inhibition of human dipeptidyl peptidase-IV (DPP-IV; EC 3.4.14.5) has emerged as a new treatment option for T2DM [3]. This enzyme belongs to the serine protease family and selectively removes N-terminal dipeptides from substrates containing proline or alanine as the second residue. The most important substrates of DPP-IV are incretins, such as glucagon-like peptide-1 (GLP-1) and glucose-dependent insulinotropic polypeptide (GIP) [4]. GLP-1 is released from intestinal L-cells in response to meals and performs the following actions: GLP-1 stimulates insulin biosynthesis and secretion, reduces glucagon release, slows gastric emptying, reduces appetite, and stimulates the regeneration and differentiation of islet B-cells [5]. Alternatively, GIP is produced by the duodenal K-cells and is extensively involved in glucose metabolism by enhancing insulin secretion [6]. Both peptides have very short half-lives because of their rapid degradation by DPP-IV. Inhibiting DPP-IV prolongs the action of GLP-1 and GIP which, in turn, improves glucose homeostasis with a lower risk of hypoglycemia. Consequently, DPP-IV inhibitors are of considerable interest to the pharmaceutical industry [7]. Intense research activities in this area have resulted in the launch of sitagliptin and vildagliptin and the advancement of a few drugs, such as saxagliptin, alogliptin and ABT-279, into pre-registration/phase 3 [2].

The DPP-IV binding site is highly druggable in the sense that tight, specific binding to the enzyme can be achieved with small molecules with drug-like physicochemical properties [8,9]. The different interaction motifs used by these DPP-IV ligands include the catalytic Ser630, the oxyanion hole (formed by Tyr547 and Tyr631), the hydrophobic S1 pocket (formed by Tyr631, Val656, Trp659, Tyr662, Tyr666 and Val711), the P2 region (formed by Arg125 and Asn710) and the N-terminal recognition region (formed by Glu205, Glu206 and Tyr662)[8]. Based on the analysis of the DPP-IV crystal structures [10-16] and interpretation of the structural-activity relationship data, both the lipophilic S1 pocket and the Glu205/Glu206 dyad can be considered as crucial molecular anchors for DPP-IV inhibition [8].

The large scaffold diversity and properties of natural products (NPs), such as structural complexity and drug similarity, makes these molecules ideal starting points for drug design. The main goal of this paper is to apply a virtual screening (VS) protocol to identify NPs with DPP-IV inhibitory activity as well as different scaffolds

relative to known DPP-IV inhibitors that could be used as lead compounds in drug-design. In order to achieve this goal, we first identified complexes between DPP-IV and potent reversible inhibitors of non-peptide nature in the PDB. After validating the fit of the coordinates of binding site residues and inhibitors onto the corresponding electron density map, the validated DPP-IV complexes were overlapped to get the experimental poses of the inhibitor in the same orientation. Subsequently, the relative contribution of the different intermolecular interactions to the protein-ligand binding affinity was quantified to derive structure-based pharmacophores. The resulting energetically optimized pharmacophores were used to derive a structure-based common pharmacophore that contained key intermolecular interactions between DPP-IV and the inhibitors. The exclusion volumes were also determined and added to the pharmacophore. Lastly, the previous structure-based pharmacophore and a VS protocol were used to look for DPP-IV inhibitors in a NPs database [17], and the reliability of the prediction was demonstrated using *in vitro* testing to determine the DPP-IV inhibitory effects of selected VS hits.

**Table 1.** Codes for DPP-IV structures currently available at PDB. Some PDB structures were discarded for the following reasons: (a) the structures were of apo forms without inhibitor, (b) inhibitors were covalently linked with Ser630, (c) inhibitors were of oligopeptide nature, (d) there were no structural factors available in the PDB or (e) the scripts in the EDS failed to produce the map from the structural factors. PDB structures marked with an asterisk (*) have mutations in the enzyme to modify the activity. Only the PDB files from the "**Valid PDB Structures**" section with $IC_{50}$ values ≤ 10 nM (in bold) were used to derive the corresponding structure-based common pharmacophore for DPP-IV inhibition (see Figure 1).

| **Valid PDB Structures** | | | **Discarted PDB Structures** | | | | |
|---|---|---|---|---|---|---|---|
| 1N1M | 2OPH | 2RIP | (a) | (b) | (c) | (d) | (e) |
| **2FJP** | 2OQI | 3C43 | 1J2E | 1TKR * | 1R9N | 1RWQ | 1X70 |
| 2HHA | 2OQV | **3C45** | 1NU6 | 2AJL | 1WCY | 2BUB | 2OAG |
| 2I78 | 2P8S | 3CCC | 1NU8 | 2G5T | 2BGN | 2JID | 3CCB |
| **2IIT** | 2QJR | 3D4L | 1PFQ | 2G5P | 2BGR | | 3EIO |
| **2IIV** | 2QOE | 3F8S | 1R9M | 2G63 | | | |
| 2OGZ | **2QT9** | **3H0C** | 1TK3 | 2I03 | | | |
| 2OLE | **2QTB** | **3HAB** | 1U8E | 2QKY | | | |
| 2ONC | **2RGU** | **3HAC** | 1W1I | 3BJM * | | | |

## Results and discussion

**Common structure-based pharmacophore building and description.** There are currently 54 entries for DPP-IV in the Protein Data Bank (PDB; http://www.pdb.org; see Table 1) [18] but only 10 of those entries correspond to validated complexes of the native enzyme with potent reversible inhibitors of a non-peptide nature (see Figure 1). As a result, only these 10 entries are suitable for deriving reliable structure-based pharmacophores that capture the key intermolecular interactions needed for drugs to inhibit DPP-IV. In order to define a common background for DPP-IV inhibition, we identified features of inhibitors that make the most important contributions to the bioactivity of the ligand by first superposing all 10 PDB files.

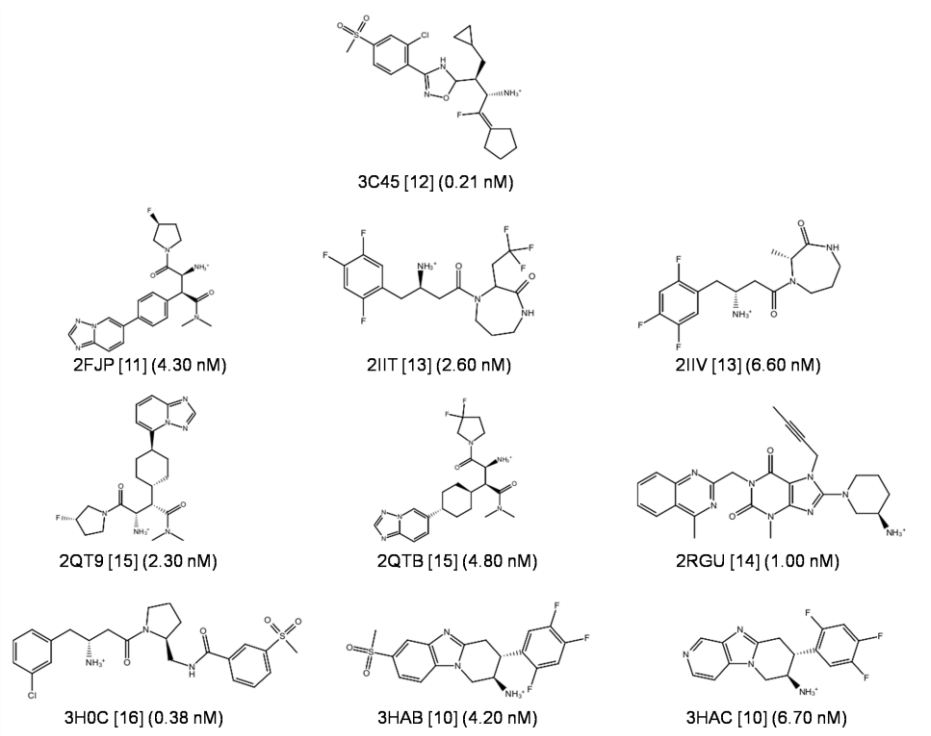



**Figure 1.** Drug-like reversible DPP-IV inhibitors used for the generation of the common structure-based pharmacophore with their corresponding $IC_{50}$ values. The codes of the PDB complexes from which the ligand poses were used are also shown.

Then, the energetic pharmacophores were derived from the resulting coordinates, and energetically relevant pharmacophore sites were visually inspected for finding common or frequent ones. Figure 2 shows that all 10 pharmacophores have two sites in common (one positive/donor and one hydrophobic/aromatic ring) that often make

the most important contribution to the protein-ligand binding affinity (see data for sites **P/D** and **H/R1** in Table 2). We inferred that these two sites are essential for the inhibition of DPP-IV and considered them to be required in the common structure-based pharmacophore (see Figure 3). Interestingly, previous studies have identified the lipophilic S1 pocket (formed by Tyr631, Val656, Trp659, Tyr662, Tyr666 and Val711) and the Glu205/Glu206 dyad as crucial molecular anchors for inhibition [8,19,20] and, in coherence with this, the mandatory hydrophobic/aromatic ring and positive/donor sites interact with the S1 pocket and Glu205/Glu206, respectively.

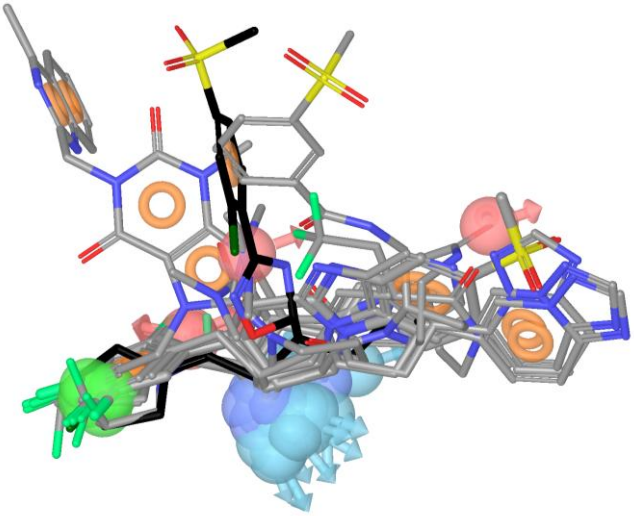


**Figure 2.** The relative location of the experimental poses of the ligands in Figure 1 after DPP-IV superposition. The experimental pose for the most potent inhibitor (*i.e.*, the one at 3C45) is shown in black for reference. For each ligand, the energetically-relevant pharmacophore sites are shown. Light red and light blue spheres represent the acceptor and donor features, respectively. The green spheres and orange torus display the hydrophobic regions and aromatic rings, respectively. Blue spheres represent positively charged regions.

Table 2 also shows that there are two other hydrogen-bond acceptors (**A1** and **A2**) and three hydrophobic/aromatic ring sites (**H/R2**, **H/R3** and **H/R4**) that, although not common to all experimental poses, could increase either protein-ligand binding affinity or drug-specificity. Moreover, it is remarkable that these sites correspond to interactions with other relevant areas from the DPP-IV binding site. For example, the two acceptor sites interact with the main chain carbonyl oxygen from Arg125 and Asn710 at the P2 region, whereas the hydrophobic/aromatic ring sites interact with Tyr547 (at the oxyanion hole), Phe357 and Arg358 among other residues. Therefore, these sites were also included as optional sites in the common structure-based pharmacophore (see Figure 3).

**Table 2**. Site contribution to the energy-optimized pharmacophores obtained from PDB complexes in bold from Table 1. Required and optional sites at the structure-based common pharmacophore are shown in cyan and yellow, respectively. Other sites that are not part of the structure-based common pharmacophore are shown in grey. PDB complexes with the same raw values indicate that the pharmacophore site is shared by these complexes.

| PDB | 2FJP | 2IIT | 2IIV | 2QT9 | 2QTB | 2RGU | 3C45 | 3H0C | 3HAB | 3HAC |
|---|---|---|---|---|---|---|---|---|---|---|
| **P/D** | -4,6 | -4,13 | -4,45 | -4,09 | -4,54 | -1,66 | -4,54 | -4,81 | -4,54 | -4,29 |
| **H/R1** | 0,77 | -1,29 | -1,36 | -1,25 | -1,68 | -0,075 | -0,64 | -1,18 | -1,1 | -1,25 |
| **H/R2** | -0,69 | | | -0,66 | -0,68 | | | | -0,9 | -0,69 |
| **H/R3** | | | | | | -1,94 | | | | |
| **H/R4** | | | | | | | -0,85 | | | |
| **H/R5** | | | | | | -0,56 | | | | |
| **A1** | -0,64 | | | -0,4 | -0,59 | | | | | |
| **A2** | | | -0,62 | | | | | | | |
| **A3** | -0,35 | | | | -0,35 | | | | | |
| **A4** | | | | | | -0,44 | | | | |

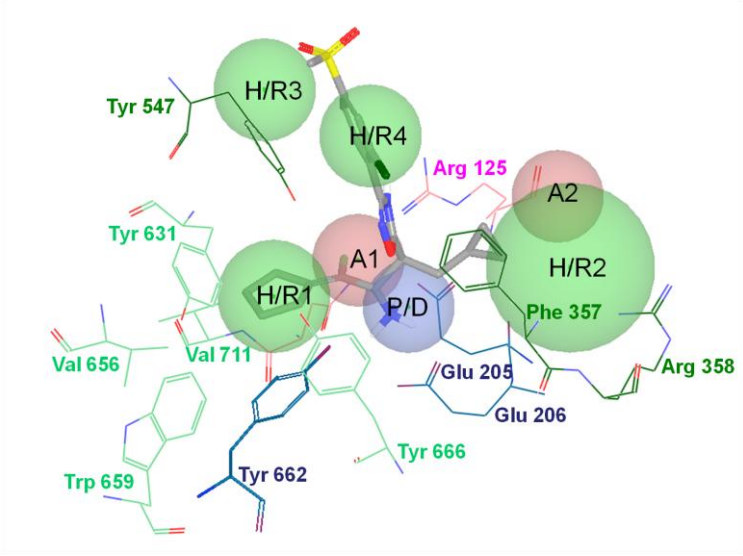



**Figure 3.** The structure-based common pharmacophore derived from the alignment of the poses in Figure 2 and shown in the context of the 3C45 active site. The pharmacophore is formed by two hydrogen-bond acceptors (**A1** and **A2**), one positive/hydrogen-bond donor feature (**P/D**) and 4 hydrophobic/aromatic ring sites (**H/R1**, **H/R2**, **H/R3** and **H/R4**). The associated tolerances (*i.e.*, radii) of the pharmacophore are 1.8Å for **P/D**, **A1** and **A2**, 2.0Å for **H/R1**, **H/R3** and **H/R4** and 3.3Å for **H/R2**. Two out of these seven sites (**P/D** and **H/R1**) are required during pharmacophore-based searches whereas the remaining five are optional. The **P/D** site interacts with the Glu205/Glu206 dyad whereas the **H/R1** site potentially fills the S1 pocket. The

residues are colored according to the type of intermolecular interactions involved. For example, blue residues interact with donor sites, pink residues interact with acceptor sites and green residues are involved in hydrophobic contacts. Light green residues are a part of the S1 pocket.

**VS workflow description and application to the NP subset of the ZINC database.** The VS workflow (see Figure 4) consisted of several sequential steps where the output molecules of one step were the input molecules for the next step and so on. The NP subset of the ZINC database was used as the source of molecules to which our VS schema was applied to search for new DPP-IV inhibitors. Initially, these 89,425 molecules were submitted to an ADME/Tox filter with the FAF-Drugs2 tool [21] aimed at discarding molecules that were either potentially toxic or exhibited poor ADME properties.

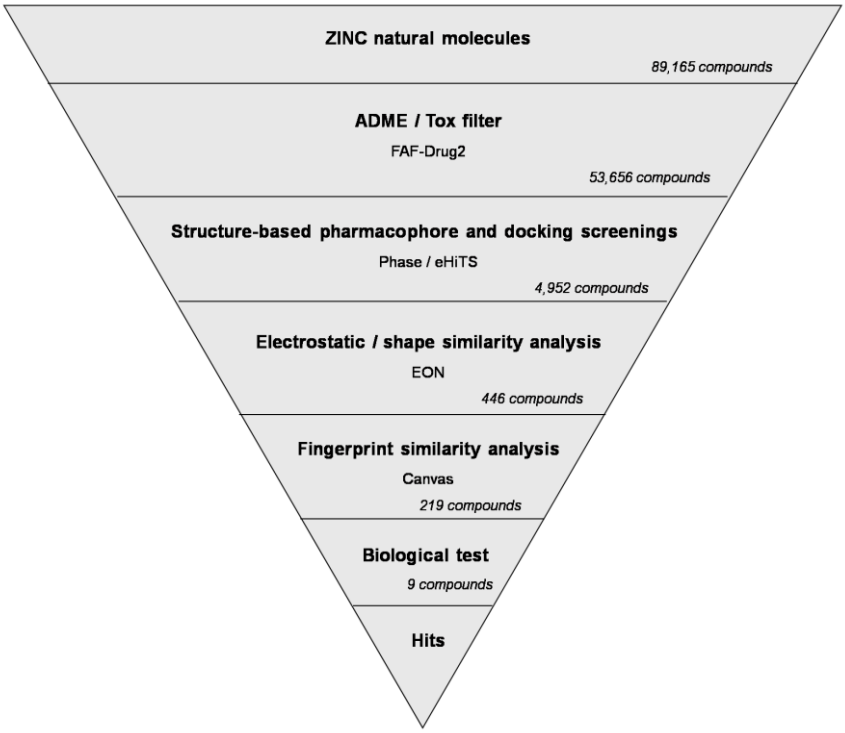



**Figure 4.** Schematic overview of the VS workflow and the procedure used for selecting the VS hits that were tested for DPP-IV inhibitory activity. For the VS, the number of compounds that passed each step and the programs used are showed. For the selection of VS hits for bioactivity testing, the numbers show either how many VS hits are scaffold-hopping candidates for DPP-IV inhibition (*Fingerprint similarity analysis* step) or how many molecules were experimentally tested for bioactivity (*Biological test* step).

Molecules with appropriate ADME/Tox properties were then filtered with Phase through the structure-based common pharmacophore. Ligands with at least one hit in the Phase search were then used in a protein-ligand rigid-docking study and docked onto the ligand binding site of the DPP-IV conformation present in the 3C45 PDB file [12]. In order to find docking poses that were compatible with the pharmacophore, the resulting ligand poses were filtered again with Phase through the structure-based common pharmacophore using the same filtering conditions as in the first Phase run without reorienting the poses (*i.e.*, the score in place option was used). From these two pharmacophore screens, we obtained 4,952 compounds (see Figure 4) with at least one pose that was both compatible with the DPP-IV active site and had functional groups that match the 3D location of the two compulsory sites and at least one of the optional sites of the structure-based common pharmacophore.

Finally, the poses for the 4,952 compounds from the second pharmacophore screen were submitted to a shape and electrostatic-potential comparison with the experimental pose of the DPP-IV inhibitor at the PDB file 3C45 (that has the smallest $IC_{50}$ for all the non-peptide reversible inhibitors found in DPP-IV-inhibitor complexes at the PDB [12]; see Figure 1). The shape and electrostatic-potential comparison identified 446 hit molecules with potential DPP-IV inhibitory activity (see Figure 4).

**Finding new scaffolds of natural origin for DPP-IV inhibitors.** One of the most important challenges of any VS workflow is the ability to find molecules with the required activity but without trivial similarity (in terms of chemical structure) to known active compounds. To determine which of the 446 potential DPP-IV inhibitors predicted by our VS workflow could be considered as new lead molecules, we merged the 446 potential DPP-IV inhibitors with 2,571 known DPP-IV inhibitors that were obtained from the BindingDB database [22]. After calculating the 2D fingerprints of these inhibitors, the resulting set was classified into 50 clusters by means of a hierarchical cluster analysis (data not shown). Notably, 12 out of the 50 clusters obtained consisted exclusively of NPs that were previously unidentified as DPP-IV inhibitors. The 219 molecules that belong to these 12 clusters are scaffold-hopping candidates for DPP-IV inhibition (see Table S1). To prove the reliability of our predictions, we selected 9 molecules (**C1** and **C2** from cluster 30, **C3** from cluster 36, **C4** from cluster 37, **C5** and **C6** from cluster 41, **C7** from cluster 45, **C8** from cluster 49 and **C9** from cluster 50) from 7 of these 12 clusters (see Figure 5) and tested their effects on the DPP-IV activity using an *in vitro* assay.

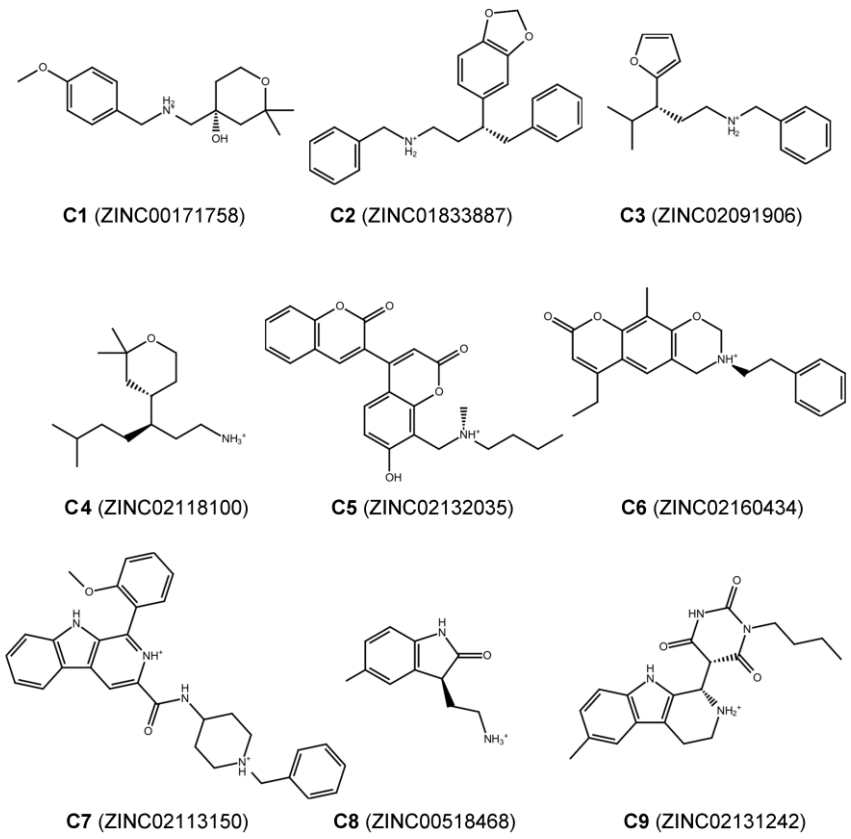



**Figure 5**. Chemical structures and ZINC codes for the 9 molecules selected for experimentally testing whether these compounds exhibited DPP-IV inhibitory activity. The insolubility of **C4** and **C6** prevented these compounds from being assayed for DPP-IV inhibitory activity.

The results of this experiment demonstrated that 7 out of the 9 molecules (**C1**, **C2**, **C3**, **C5**, **C7**, **C8** and **C9**) inhibit DPP-IV (see Figure 6). The remaining molecules, **C4** and **C6**, could not be solubilized, preventing the evaluation of their DPP-IV inhibitory activity. The lack of DPP-IV inhibitory activity for **C5**, **C7** and **C9** at 1mM was also due to insolubility (see Figure 6). Furthermore, Figure 6 shows that from all the tested molecules, **C5** is the most potent inhibitor with an $IC_{50}$ of 61.55 μM (see Figure 7). With the exception of **C1**, which significantly inhibited DPP-IV only at 1 mM, the rest of the molecules significantly inhibit DPP-IV at 0.25 mM (see Figure 6) showing a dose-response effect. Moreover, a SciFinder search (Chemical Abstracts Service, Columbus, Ohio, USA; http://www.cas.org/products/sfacad) of the literature revealed that none of these 7 molecules have been reported as antidiabetic drugs. In fact, no bioactivity has been described for these 7 molecules.

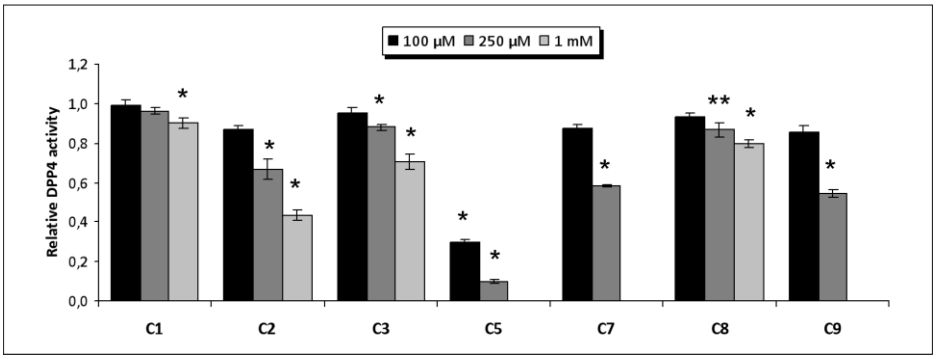



**Figure 6**. Dose-response results for the *in vitro* inhibition of DPP-IV by **C1**, **C2**, **C3**, **C5**, **C7**, **C8** and **C9**. The relative DPP-IV inhibitory activity with or without the selected NPs (vehicle, 1% DMSO) is shown where each column represents the average ± SEM (n=3 or 4). The insolubility of **C5**, **C7** and **C9** in DMSO at 1 mM prevented the measurement of DPP-IV inhibitory activity . *p<0.05 **p<0.1 vs vehicle, T-student.

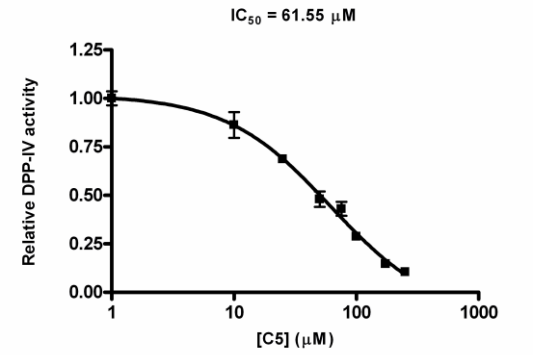



**Figure 7.** DPP-IV inhibitory dose-response curve obtained for **C5** via a competitive binding assay.

**Structural analysis of the inhibition of DPP-IV by C1, C2, C3, C5, C7, C8 and C9.** The docking of **C1**, **C2**, **C3**, **C5**, **C7**, **C8** and **C9** in the DPP-IV binding site of the 3C45 structure demonstrated that these molecules match the structure-based common pharmacophore in the same orientation, sharing the same intermolecular interaction with DPP-IV (see Figures 8 and 9). With the exception of **C7** in which the positive charge of the tertiary amine forms a salt bridge with Glu205/Glu206 (see panel D of Figure 9), all compounds use primary or secondary amines to form hydrogen bond interactions with either Glu205 or Glu206 (see Figures 8 and 9). Additionally, all molecules filled the S1 pocket (partially in the case of **C8,** the smallest molecule) establishing one intermolecular interaction that corresponds to the compulsory **H/R1** site of our common structure-based pharmacophore (see Figure 3). Moreover, it is worthwhile to mention that some molecules could potentially form additional

interactions with DPP-IV. For example, the hydroxyl group of **C1** could hydrogen bond with the side chain of Glu205 in addition to its the typical interaction with Glu206 (see panel A in Figure 9). **C2** could form an additional hydrogen bond between the oxygen of the dioxole moiety and Arg358 (see panel B in Figure 9). **C7** could hydrogen bond with Tyr585 and form additional hydrophobic contacts with Phe357, Ile405, Cys551 and Tyr585, due to its large size (see panel D in Figure 9). **C9** could form two additional hydrogen bonds with Tyr662 and the main chain oxygen of Glu205.

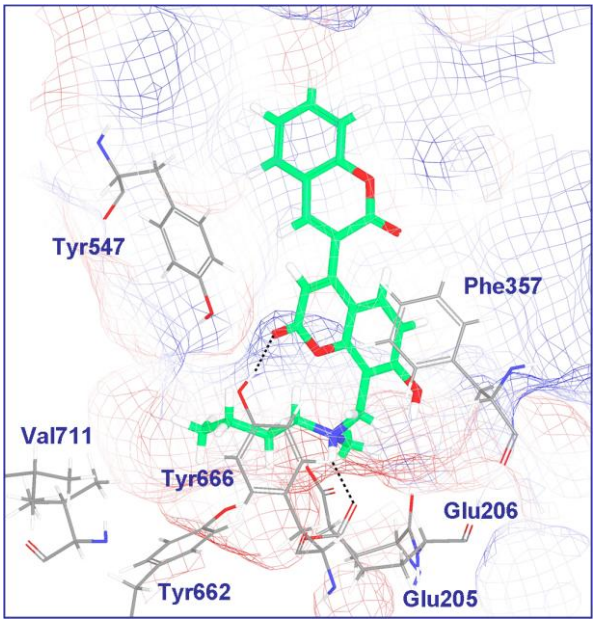



**Figure 8.** Docking pose of compound **C5** at the 3C45 binding site.

Figure 8 shows the best docking pose of **C5** in the DPP-IV binding pocket where its tertiary amine hydrogen bonds with Glu205. The oxygen of the 7-hydroxy-2H-chromen-2-one moiety could also hydrogen bond with Tyr666. The S1 pocket is occupied by the **C5** butyl chain that could form hydrophobic interactions with Tyr662, Tyr666 and Val711. Finally, the chromene rings of **C5** could form π-π interactions with Phe357. Compared to the remaining 6 compounds, the high bioactivity of **C5** seems to be related to the presence of additional intermolecular interactions with DPP-IV. Interestingly, an electrostatic and shape comparison of the 7 poses in Figures 8 and 9 revealed that the molecule with the highest similarity to the 3C45 ligand (with the lowest $IC_{50}$; see Figure 1) is **C5** (results not shown). The ET_combo score for this comparison is 1.050 which corresponds to a shape and electrostatic contribution of 0.628 and 0.422, respectively. Remarkably, the same analysis with **C2** (which shows a

significant bioactivity as DPP-IV inhibitor; see Figure 6), also has a significant ET_combo score of 1.038.

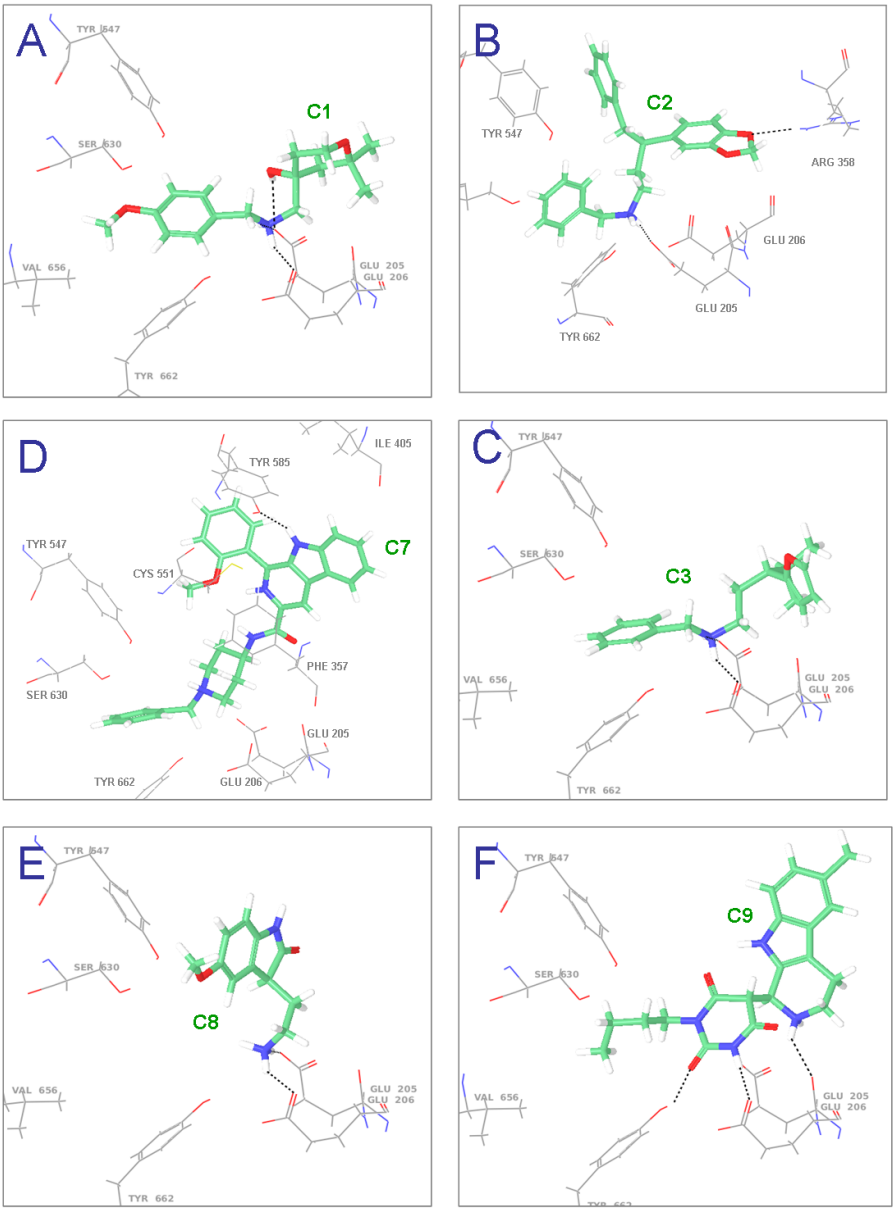



**Figure 9.** Docking poses for **C1**, **C2**, **C3**, **C7**, **C8** and **C9** at the 3C45 binding site.

## Conclusions

The challenge of any VS protocol consists of using *in silico* tools to predict which molecules in a database have the required activity against a specific target. The results of the present study demonstrate that our VS protocol is highly successful in the non-trivial identification of DPP-IV inhibitors with no chemical-structure similarities to known activities. Therefore, scaffold hopping on this target can be achieved. Moreover, this is the first time that anti-diabetic activity has been described for **C1** (*i.e.*, ZINC00171758), **C2** (*i.e.*, ZINC01833887), **C3** (*i.e.*, ZINC02091906), **C5** (*i.e.*, ZINC02132035), **C7** (*i.e.*, ZINC02113150), **C8** (*i.e.*, ZINC00518468) and **C9** (*i.e.*, ZINC02131242).

Although the $IC_{50}$ of the 7 hit molecules indicates their *in vitro* activity is significantly lower than that of most known DPP-IV inhibitors used to derive the structure-based common pharmacophore (see Figure 1), it is important to remark that these molecules can be used as lead compounds for developing more potent inhibitors using structural-activity relationship studies. Furthermore, these 7 molecules were selected based on their commercial availability, cost and purity with the primary goal of testing the performance of our VS protocol. Therefore, it is possible that there are other molecules among the remaining 210 molecules in clusters 10, 29, 30, 36, 37, 38, 40, 41, 44, 45, 49 and 50 (see Table S1) that could be better starting points for the rational drug design of potent and selective DPP-IV inhibitors with chemical scaffolds that are different from **C5**. Remarkably, our work makes a significant contribution to the discovery of DPP-IV inhibitors of natural origin (described, at present, for only few NPs [23-26]) from a quantitative point of view. Moreover, this work is also applicable to screen synthetic molecules databases when looking for antidiabetic activity.

## Experimental section

**Criteria for selecting the 3D structures for DPP-IV complexes used to derive the common structure-based pharmacophore.** Coordinates for complexes between DPP-IV and potent reversible inhibitors were obtained from the PDB with the help of the following information: (a) LigPlot [27] schemes downloaded from the PDBsum website (http://www.ebi.ac.uk/pdbsum/) that were used to confirm the non-peptide and reversible character of the DPP-IV inhibitor present in each complex and; (b) $IC_{50}$ values directly extracted from the literature describing the complexes (only complexes with inhibitors with $IC_{50} \leq 10$ nM were considered). Furthermore, the complexes with at least one mutation in their amino acid sequences were discarded. The reliability of the binding-site residues and inhibitor coordinates was assessed for the remaining complexes by visually inspecting their degree of fitness to the corresponding electron

density map available from the Uppsala Electron Density Server (EDS; http://eds.bmc.uu.se/eds/) [28].

**Superposition of the selected DPP-IV structures.** The coordinates from the PDB complexes that met all the mentioned requirements were superposed with the DeepView v3.7 program (http://spdbv.vital-it.ch/) [29] to have the complexes in the same relative orientation. Only the resulting re-oriented coordinates for these PDB files were used during the subsequent structure-based pharmacophore generation and in the steps of the VS workflow (*i.e.*, pharmacophore-based searches, protein-ligand docking studies and shape and electrostatic-potential comparisons) where spatial orientation is crucial.

**Common structure-based pharmacophore for DPP-IV inhibition.** Energetic structure-based pharmacophores were built from the superposed coordinates of the previously selected complexes by means of the Glide-based procedure developed by Schrödinger (Schrödinger LLC., Portland, USA; http://www.schrodinger.com) [30]. According to this procedure, pharmacophore sites are ranked based on the Glide XP energies with the advantage that each contribution to the protein-ligand interactions is quantified. Therefore, energetically favorable features can be incorporated into the pharmacophore with preference over energetically weaker features. The resulting individual energetic pharmacophores were used for the construction of a common structure-based pharmacophore for DPP-IV reversible inhibition. This pharmacophore consists on two compulsory sites (one positive/donor and one hydrophobic/aromatic ring) whereas the remaining acceptor and hydrophobic/aromatic ring sites are optional. The associated tolerances for the different sites are 1.8Å for **P/D**, **A1** and **A2**, 2.0Å for **H/R1**, **H/R3** and **H/R4** and 3.3Å for **H/R2**. The pharmacophore was completed with receptor-based excluded volumes that schematically represent the location of the DPP-IV residues that form the binding pocket by applying the ***Receptor-Based Excluded Volumes*** graphic front-end from Phase v3.1 (Schrödinger LLC., Portland, USA; http://www.schrodinger.com)[31] to the PDB file 3C45. The ***Sphere filters*** parameter values were set to the following criteria: (a) ignoring receptor atoms whose surfaces were within 0.25 Å of ligand surface; and (b) limit excluded volume shell thickness to 10 Å. Otherwise, the remaining parameter values used were the default values.

**Ligand selection for VS purposes.** Ligands for VS purposes were downloaded from the Natural Products subset of the ZINC database [17]. This dataset contains 89,165 commercially available natural products and natural-product derivatives, making the dataset suitable for experimentally testing the success of a VS workflow.

**ADME/Tox filter.** The ADME/Tox filter was carried out with the FAF-Drugs2 tool [21]. The drug-like properties of a compound were evaluated using the Lipinski rule [32]. The Lipinski rule is based on a set of property values , such as the number of

hydrogen-bond donors and acceptors, the molecular weight and the logP, that were derived from drugs with good ADME characteristics. Molecules that adhere to the Lipinski rule are expected to be active in humans after oral admission. Only one violation of this rule was allowed. Molecules containing toxic groups were filtered using the 204 substructures for "warhead" chelators, frequent hitters, promiscuous inhibitors and other undesirable functional groups available in the FAF-Drugs2 tool [21].

**Ligand setup.** The 3D structures of the ligands for VS purposes were incorporated into LigPrep v2.3 (Schrödinger LLC., Portland, USA; http://www.schrodinger.com) and improved by cleaning. The cleaning process was carried out using the following parameters: (a) the force field used was OPLS 2005; (b) all possible ionization states at pH 7.0±2.0 were generated with Ionizer; (c) the desalt option was activated; (d) tautomers were generated for all ionization states at pH 7.0±2.0; (e) chiralities were determined from the 3D structure; and (f) one low-energy ring conformation per ligand was generated. Conformations and sites for the resulting ligand structures were determined during the generation of the corresponding Phase [31] databases with the ***Generate Phase Database*** graphic front-end. Default parameter values were used during this conformer generation with the exception of the maximum number of conformers per structure, which increased from 100 (the default value) to 200. The conformer sites were generated with definitions made by adding the ability to consider aromatic rings as hydrophobic groups to the default built-in Phase definitions.

**Structure-based pharmacophore screening.** The initial filtering through the structure-based common pharmacophore was performed with Phase v3.1 using the following steps: (a) search in the conformers database, (b) do not score in place the conformers into the structure-based common pharmacophore (*i.e.*, allow reorientation of the conformers to determine if they match the pharmacophore or not), (c) match the two compulsory sites of the structure-based common pharmacophore and at least one of the optional sites, (d) do not have a preference for partial matches involving more sites and (e) use the excluded volumes from the structure-based common pharmacophore. Default values were used for the rest of the options and parameter values used during this search. For the second pharmacophore screening, the same filtering options of the first pharmacophore matching were applied with the exception that now no re-orientation of the poses was allowed during the search (*i.e.*, the *score in place* option was used) because it was performed by using docked poses.

**Protein-ligand docking during the VS.** During the VS, the protein-ligand docking was performed with eHiTS v2009 (SimBioSys Inc., Toronto, Canada; http://www.simbiosys.ca/ehits) [33]. The receptor was considered to be a rigid body and the ligands as flexible such that free rotation was allowed around the single bonds of the ligand. Default docking conditions were selected with the exception of the size

of the sides of the cubic box encompassing the DPP-IV binding site, which was increased from 10 Å to 15 Å.

**Electrostatic and shape similarity screening.** The software EON v2.0.1 (OpenEye Scientific Software, Inc., Santa Fe, New Mexico, USA; http://www.eyesopen.com) determines the electrostatic potentials of two compounds and consequently calculates the Electrostatic Tanimoto combo score (ET_combo). The ET_combo is the sum of the Shape Tanimoto (ST) and the Poisson-Boltzman Electrostatic Tanimoto scores. The Shape Tanimoto (ST) score is a quantitative measure of three-dimensional overlap where 1 corresponds to a perfect overlap (*i.e.*, the same shape) [34]. The Poisson-Boltzman Electrostatic Tanimoto score (ET_pb) compares the electrostatic potential of two small molecules where 1 corresponds to identical potentials and negative values correspond to the overlap of positive and negative charges [35]. Only those molecules that have both ET_pb and ST score values higher than 0.623 and 0.244, respectively, were selected and visualized with VIDA v4.0.3 (OpenEye Scientific Software, Inc., Santa Fe, New Mexico, USA; http://www.eyesopen.com). These threshold values were chosen after analyzing which ET pb and ST score values are obtained when the DPP-IV inhibitor in PDB file 3C45 is compared with the experimental poses of the rest of the inhibitors from which the common pharmacophore was derived (see Figure 1).

**Hit selection for further experimental assays on DPP-IV activity.** The molecules that survived the electrostatics/shape similarity filter were merged with 2,571 known inhibitors obtained from the BindingDB database [22], and then clustered using Canvas v1.2 (Schrödinger LLC., Portland, USA; http://www.schrodinger.com). MOLPRINT2D fingerprints [36], using a fingerprint precision of 32 bits, were calculated for each molecule and then hierarchical clustering, based on Tanimoto similarities, was performed resulting in 50 clusters. Nine compounds from 7 of the 12 clusters exclusively formed by NPs that were previously unidentified as DPP-IV inhibitors were selected based on their commercial availability, cost and purity (≥ 92%) for *in vitro* assays of DPP-IV inhibitory activity. These compounds were ZINC00171758 (*i.e.*, **C1**), ZINC01833887 (*i.e.*, **C2**), ZINC02091906 (*i.e.*, **C3**), ZINC02118100 (*i.e.*, **C4**), ZINC02132035 (*i.e.*, **C5**) and ZINC02160434 (*i.e.*, **C6**), ZINC02113150 (*i.e.*, **C7**), ZINC00518468 (*i.e.*, **C8**) and ZINC02131242 (*i.e.*, **C9**), which were all purchased from InterBioScreen, Ltd (http://www.ibscreen.com).

***In vitro* assay of the effect of selected compounds on the DPP-IV activity.** The *DPP-IV Drug Discovery Kit-AK499* (Enzo Life Sciences International, Inc.) was used to conduct DPP-IV inhibition assays. Briefly, 10 µL of each compound were added to commercial recombinant human DPP-IV. Stock solutions of the assayed compound were made in DMSO and diluted in buffer (50 mM Tris-HCl) to final concentrations

ranging from 10-1000 µM in the assay. The final concentration of DMSO in the assay was 1%. After 10 minutes of incubation at 37 ºC, the reaction was initiated by the addition of the fluorimetric substrate H-Gly-Pro-AMC. Fluorescence was measured continuously for 30 minutes at Ex: 380 nm/Em: 460 nm in a microplate reader. At least three independent assays were performed, each with two technical replicates. A standard DPP-IV inhibitor (P32/98 from Biomol, Germany) served as positive control.

**$IC_{50}$ calculation.** $IC_{50}$ was determined using GraphPad Prism v4.0 for Windows (GraphPad Software, San Diego CA, USA; http://www.graphpad.com) by fitting the experimental data from the *in vitro* assay to a nonlinear regression function using a four-parameter logistic equation.

**Docking of Novel DPP-IV Ligands** Docking studies of DPP-IV inhibitors **C1**, **C2**, **C3**, **C5**, **C7**, **C8** and **C9** were performed with the software Glide v5.6 (Schrödinger LLC., Portland, USA; http://www.schrodinger.com) using the DPP-IV protein coordinates that can be found using the 3C45 PDB code. The binding site was defined using the default options of the *Receptor Grid Generation* panel. Standard-precision (SP) docking was used to screen the ligands. The flexible docking mode was selected such that Glide internally generated conformations during the docking process. No constraints were selected for docking. Each docking run recorded at most ten poses per ligand that survived the post-docking minimization. The best docking poses for the novel DPP-IV ligands were selected by not only considering the docking scores but also by taking into account the results of the visual inspection of all docking poses that was performed with Maestro v9.2 (Schrödinger LLC., Portland, USA; http://www.schrodinger.com).

## References

1. Green BD, Flatt PR, Bailey CJ. (2006) Dipeptidyl peptidase IV (DPP IV) inhibitors: A newly emerging drug class for the treatment of type 2 diabetes. Diabetes & Vascular Disease Research: Official Journal of the International Society of Diabetes and Vascular Disease 3(3): 159-165.

2. Havale SH, Pal M. (2009) Medicinal chemistry approaches to the inhibition of dipeptidyl peptidase-4 for the treatment of type 2 diabetes. Bioorganic & Medicinal Chemistry 17(5): 1783-1802.

3. Yazbeck R, Howarth GS, Abbott CA. (2009) Dipeptidyl peptidase inhibitors, an emerging drug class for inflammatory disease? Trends Pharmacol Sci 30(11): 600-607.

4. Mentlein R, Gallwitz B, Schmidt WE. (1993) Dipeptidyl-peptidase IV hydrolyses gastric inhibitory polypeptide, glucagon-like peptide-1(7-36)amide, peptide histidine methionine and is responsible for their degradation in human serum. European Journal of Biochemistry FEBS 214(3): 829-835.

5. Brubaker PL, Drucker DJ. (2004) Minireview: Glucagon-like peptides regulate cell proliferation and apoptosis in the pancreas, gut, and central nervous system. Endocrinology 145(6): 2653-2659.

6. Meier JJ, Nauck MA, Schmidt WE, Gallwitz B. (2002) Gastric inhibitory polypeptide: The neglected incretin revisited. Regul Pept 107(1-3): 1-13.

7. Demuth H, McIntosh Christopher H S, Pederson RA. (2005) Type 2 diabetes--therapy with dipeptidyl peptidase IV inhibitors. Biochim Biophys Acta 1751(1): 33-44.

8. Kuhn B, Hennig M, Mattei P. (2007) Molecular recognition of ligands in dipeptidyl peptidase IV. Current Topics in Medicinal Chemistry 7(6): 609-619.

9. Zettl H, Schubert-Zsilavecz Manfred, Steinhilber D. (2010) Medicinal chemistry of incretin mimetics and DPP-4 inhibitors. ChemMedChem 5(2): 179-185.

10. Edmondson SD, Mastracchio A, Cox JM, Eiermann GJ, He H, et al. (2009) Aminopiperidine-fused imidazoles as dipeptidyl peptidase-IV inhibitors. Bioorganic & Medicinal Chemistry Letters 19(15): 4097-4101.

11. Edmondson SD, Mastracchio A, Mathvink RJ, He J, Harper B, et al. (2006) (2S,3S)-3-amino-4-(3,3-difluoropyrrolidin-1-yl)-N,N-dimethyl-4-oxo-2-(4-[12,4]triazolo[1,5-a]-pyridin-6-ylphenyl)butanamide: A selective alpha-amino amide dipeptidyl peptidase IV inhibitor for the treatment of type 2 diabetes. J Med Chem 49(12): 3614-3627.

12. Edmondson SD, Wei L, Xu J, Shang J, Xu S, et al. (2008) Fluoroolefins as amide bond mimics in dipeptidyl peptidase IV inhibitors. Bioorganic & Medicinal Chemistry Letters 18(7): 2409-2413.

13. Biftu T, Feng D, Qian X, Liang G, Kieczykowski G, et al. (2007) (3R)-4-[(3R)-3-amino-4-(24,5-trifluorophenyl)butanoyl]-3-(2,2,2-trifluoroethyl)-1,4-diazepan-2-one, a selective dipeptidyl peptidase IV inhibitor for the treatment of type 2 diabetes. Bioorganic & Medicinal Chemistry Letters 17(1): 49-52.

14. Eckhardt M, Langkopf E, Mark M, Tadayyon M, Thomas L, et al. (2007) 8-(3-(R)-aminopiperidin-1-yl)-7-but-2-ynyl-3-methyl-1-(4-methyl-quinazolin-2-ylmethyl)-37-dihydropurine-2,6-dione (BI 1356), a highly potent, selective, long-acting, and orally bioavailable DPP-4 inhibitor for the treatment of type 2 diabetes. J Med Chem 50(26): 6450-6453.

15. Kaelin DE, Smenton AL, Eiermann GJ, He H, Leiting B, et al. (2007) 4-arylcyclohexylalanine analogs as potent, selective, and orally active inhibitors of dipeptidyl peptidase IV. Bioorganic & Medicinal Chemistry Letters 17(21): 5806-5811.

16. Nordhoff S, Cerezo-Galvez Silvia, Deppe H, Hill O, LÃ³pez-Canet Meritxell, et al. (2009) Discovery of beta-homophenylalanine based pyrrolidin-2-ylmethyl amides and sulfonamides as highly potent and selective inhibitors of dipeptidyl peptidase IV. Bioorganic & Medicinal Chemistry Letters 19(15): 4201-4203.

17. Irwin JJ, Shoichet BK. (2005) ZINC--a free database of commercially available compounds for virtual screening. Journal of Chemical Information and Modeling 45(1): 177-182.

18. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, et al. (2002) The protein data bank. Acta Crystallographica Section D 58(Pt 6 No 1): 899-907.

19. Al-Masri Ihab M, Mohammad MK, Taha MO. (2008) Discovery of DPP IV inhibitors by pharmacophore modeling and QSAR analysis followed by in silico screening. ChemMedChem 3(11): 1763-1779.

20. Ward RA, Perkins TDJ, Stafford J. (2005) Structure-based virtual screening for low molecular weight chemical starting points for dipeptidyl peptidase IV inhibitors. J Med Chem 48(22): 6991-6996.

21. Lagorce D, Sperandio O, Galons H, Miteva MA, Villoutreix BO. (2008) FAF-Drugs2: Free ADME/tox filtering tool to assist drug discovery and chemical biology projects. BMC Bioinformatics 9: 396.

22. Liu T, Lin Y, Wen X, Jorissen RN, Gilson MK. (2007) BindingDB: A web-accessible database of experimentally determined protein-ligand binding affinities. Nucleic Acids Res 35(Database issue): D198-201.

23. Zhang S, Lu W, Liu X, Diao Y, Bai F, et al. (2011) Fast and effective identification of the bioactive compounds and their targets from medicinal plants via computational chemical biology approach. Med.Chem.Commun. 2(6): 471-477.

24. Al-masri IM, Mohammad MK, Tahaa MO. (2009) Inhibition of dipeptidyl peptidase IV (DPP IV) is one of the mechanisms explaining the hypoglycemic effect of berberine. J Enzyme Inhib Med Chem 24(5): 1061-1066.

25. Akiyama T, Abe M, Harada S, Kojima F, Sawa R, et al. (2001) Sulphostin, a potent inhibitor for dipeptidyl peptidase IV from streptomyces sp. MK251-43F3. J Antibiot (Tokyo) 54(9): 744-746.

26. Pascual I, Lopéz A, Gómez H, Chappé M, Saroyán A, et al. (2007) Screening of inhibitors of porcine dipeptidyl peptidase IV activity in aqueous extracts from marine organisms. Enzyme Microb Technol 40(3): 414.

27. Wallace AC, Laskowski RA, Thornton JM. (1995) LIGPLOT: A program to generate schematic diagrams of protein-ligand interactions. Protein Eng 8(2): 127-134.

28. Kleywegt GJ, Harris MR, Zou JY, Taylor TC, Wählby A, et al. (2004) The uppsala electron-density server. Acta Crystallographica.Section D, Biological Crystallography 60(Pt 12 Pt 1): 2240-2249.

29. Guex N, Peitsch MC. (1997) SWISS-MODEL and the swiss-PdbViewer: An environment for comparative protein modeling. Electrophoresis 18(15): 2714-2723.

30. Salam NK, Nuti R, Sherman W. (2009) Novel method for generating structure-based pharmacophores using energetic analysis. Journal of Chemical Information and Modeling 49(10): 2356-2368.

31. Dixon SL, Smondyrev AM, Knoll EH, Rao SN, Shaw DE, et al. (2006) PHASE: A new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. methodology and preliminary results. Journal of Computer-Aided Molecular Design 20(10-11): 647-671.

32. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 46(1-3): 3-26.

33. Zsoldos Z, Reid D, Simon A, Sadjad SB, Johnson AP. (2007) eHiTS: A new fast, exhaustive flexible ligand docking system. J Mol Graph Model 26(1): 198-212.

34. Rush TS, Grant JA, Mosyak L, Nicholls A. (2005) A shape-based 3-D scaffold hopping method and its application to a bacterial protein-protein interaction. J Med Chem 48(5): 1489-1495.

35. Naylor E, Arredouani A, Vasudevan SR, Lewis AM, Parkesh R, et al. (2009) Identification of a chemical probe for NAADP by virtual screening. Nature Chemical Biology 5(4): 220-226.

36. Duan J, Dixon SL, Lowrie JF, Sherman W. (2010) Analysis and comparison of 2D fingerprints: Insights into database screening performance using eight fingerprint methods. Journal of Molecular Graphics & Modelling 29(2): 157-170.

## Supporting information

**Table S1.** Predicted scaffold-hopping candidates for DPP-IV inhibition. This table shows ZINC codes for the 219 hit molecules predicted to inhibit DPP-IV that belong exclusively to clusters containing NPs that were previously unidentified as DPP-IV inhibitors. The best results of the shape and electrostatic-potential comparisons for each hit molecule with the ligand of 3C45 crystallized structure are shown. The Tanimoto values for the comparison between the electrostatic potentials of the molecules (using an outer dielectric of 80) are shown in the ET_PB columns. Furthermore, the values for the comparison between shapes are shown in the ET_Shape columns. The sum of the ET_PB and ET_Shape values is reported in the Combo columns. Hits from each cluster are sorted according to their decreasing combo value. ZINC00171758 and ZINC01833887 (cluster 30), ZINC02091906 (cluster 36), ZINC02118100 (cluster 37), ZINC02132035 and ZINC02160434 (cluster 41), ZINC02113150 (from cluster 45), ZINC00518468 (cluster 49) and ZINC02131242 (cluster 50) were tested in an *in vitro* assay to validate the success rate of our predictions (in bold in Table S1). Due to the insolubility, ZINC02118100 (cluster 37) and ZINC02160434 (cluster 41) could not be tested.

| Molecule | Cluster | ET PB | ET Combo | ET Shape |
|---|---|---|---|---|
| ZINC08624216 | 10 | 0.670 | 1.299 | 0.629 |
| ZINC08624227 | 10 | 0.664 | 1.270 | 0.606 |
| ZINC12603399 | 10 | 0.660 | 1.120 | 0.460 |
| ZINC08624212 | 10 | 0.679 | 1.063 | 0.384 |
| ZINC12603352 | 10 | 0.647 | 1.039 | 0.393 |
| ZINC12603382 | 10 | 0.628 | 1.009 | 0.381 |
| ZINC08636132 | 10 | 0.737 | 1.007 | 0.270 |
| ZINC04235298 | 29 | 0.732 | 1.113 | 0.381 |
| ZINC03850486 | 29 | 0.629 | 1.011 | 0.383 |
| ZINC01826519 | 30 | 0.691 | 1.390 | 0.699 |
| ZINC01826518 | 30 | 0.716 | 1.254 | 0.538 |
| ZINC02111476 | 30 | 0.747 | 1.252 | 0.505 |
| ZINC02095165 | 30 | 0.702 | 1.237 | 0.535 |
| ZINC02111654 | 30 | 0.680 | 1.221 | 0.541 |
| ZINC04026917 | 30 | 0.766 | 1.174 | 0.408 |
| ZINC04027780 | 30 | 0.745 | 1.167 | 0.422 |
| ZINC00171740 | 30 | 0.764 | 1.166 | 0.402 |
| ZINC02111107 | 30 | 0.689 | 1.159 | 0.470 |
| ZINC01833886 | 30 | 0.703 | 1.157 | 0.454 |
| **ZINC00171758** | **30** | **0.629** | **1.151** | **0.522** |
| ZINC02098064 | 30 | 0.736 | 1.136 | 0.400 |
| ZINC01744114 | 30 | 0.733 | 1.135 | 0.402 |
| ZINC01831741 | 30 | 0.666 | 1.133 | 0.467 |
| ZINC00288906 | 30 | 0.661 | 1.130 | 0.469 |
| ZINC01767546 | 30 | 0.688 | 1.123 | 0.435 |
| **ZINC01833887** | **30** | **0.686** | **1.123** | **0.437** |
| ZINC01831742 | 30 | 0.660 | 1.121 | 0.461 |
| ZINC02101423 | 30 | 0.687 | 1.120 | 0.433 |
| ZINC04043691 | 30 | 0.720 | 1.119 | 0.398 |
| ZINC01815758 | 30 | 0.626 | 1.117 | 0.490 |
| ZINC00171737 | 30 | 0.770 | 1.113 | 0.344 |
| ZINC00526001 | 30 | 0.669 | 1.113 | 0.444 |
| ZINC04073289 | 30 | 0.699 | 1.109 | 0.411 |
| ZINC02111026 | 30 | 0.626 | 1.102 | 0.476 |
| ZINC00171675 | 30 | 0.681 | 1.101 | 0.420 |
| ZINC04073208 | 30 | 0.662 | 1.091 | 0.428 |
| ZINC00177015 | 30 | 0.760 | 1.089 | 0.330 |
| ZINC00526000 | 30 | 0.646 | 1.089 | 0.443 |
| ZINC02101849 | 30 | 0.697 | 1.089 | 0.392 |
| ZINC01810429 | 30 | 0.690 | 1.079 | 0.389 |
| ZINC04038362 | 30 | 0.721 | 1.075 | 0.353 |
| ZINC01767542 | 30 | 0.731 | 1.073 | 0.342 |
| ZINC02111059 | 30 | 0.644 | 1.065 | 0.421 |
| ZINC02095167 | 30 | 0.655 | 1.049 | 0.394 |
| ZINC01867331 | 30 | 0.669 | 1.047 | 0.378 |
| ZINC01815759 | 30 | 0.682 | 1.044 | 0.362 |
| ZINC00177014 | 30 | 0.672 | 1.041 | 0.369 |
| ZINC00171760 | 30 | 0.654 | 1.040 | 0.386 |
| ZINC01713519 | 30 | 0.655 | 1.033 | 0.378 |
| ZINC01801287 | 30 | 0.631 | 1.023 | 0.392 |
| ZINC02103306 | 30 | 0.636 | 1.019 | 0.384 |
| ZINC00206548 | 30 | 0.664 | 1.005 | 0.341 |
| ZINC01801292 | 30 | 0.629 | 0.993 | 0.364 |
| ZINC04073422 | 30 | 0.635 | 0.939 | 0.304 |
| ZINC08623233 | 30 | 0.626 | 0.890 | 0.263 |
| ZINC02104434 | 36 | 0.779 | 1.260 | 0.481 |
| ZINC02147315 | 36 | 0.792 | 1.206 | 0.414 |
| ZINC02091908 | 36 | 0.711 | 1.202 | 0.490 |
| **ZINC02091906** | **36** | **0.703** | **1.196** | **0.493** |
| ZINC02126019 | 36 | 0.808 | 1.196 | 0.389 |
| ZINC02126020 | 36 | 0.791 | 1.163 | 0.372 |
| ZINC04088482 | 36 | 0.692 | 1.117 | 0.425 |
| ZINC02104432 | 36 | 0.741 | 1.108 | 0.367 |
| ZINC02154701 | 36 | 0.684 | 1.093 | 0.409 |
| ZINC02157136 | 36 | 0.722 | 1.060 | 0.337 |
| ZINC04085398 | 36 | 0.682 | 1.059 | 0.377 |
| ZINC04089120 | 36 | 0.644 | 1.042 | 0.398 |
| ZINC02098258 | 36 | 0.691 | 1.041 | 0.350 |
| ZINC02103154 | 36 | 0.624 | 1.019 | 0.394 |
| ZINC04084395 | 36 | 0.694 | 1.017 | 0.323 |
| ZINC08635889 | 36 | 0.688 | 1.007 | 0.319 |
| ZINC04089122 | 36 | 0.678 | 1.005 | 0.327 |
| ZINC04086470 | 36 | 0.655 | 0.928 | 0.273 |
| ZINC02103096 | 37 | 0.783 | 1.290 | 0.507 |
| ZINC02118102 | 37 | 0.808 | 1.230 | 0.421 |
| ZINC02118103 | 37 | 0.770 | 1.213 | 0.443 |
| **ZINC02118100** | **37** | **0.799** | **1.200** | **0.401** |
| ZINC02118098 | 37 | 0.766 | 1.192 | 0.427 |
| ZINC00206863 | 37 | 0.713 | 1.181 | 0.468 |
| ZINC00206867 | 37 | 0.757 | 1.167 | 0.410 |
| ZINC00244210 | 37 | 0.700 | 1.166 | 0.466 |
| ZINC00201903 | 37 | 0.718 | 1.150 | 0.431 |
| ZINC01791910 | 37 | 0.641 | 1.139 | 0.498 |
| ZINC04046504 | 37 | 0.777 | 1.123 | 0.346 |
| ZINC00201907 | 37 | 0.736 | 1.121 | 0.385 |
| ZINC00244221 | 37 | 0.770 | 1.117 | 0.347 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| ZINC02098134 | 37 | 0.683 | 1.117 | 0.433 | ZINC02159373 | 41 | 0.724 | 1.107 | 0.383 |
| ZINC04046709 | 37 | 0.745 | 1.114 | 0.369 | ZINC00934318 | 41 | 0.655 | 1.103 | 0.448 |
| ZINC00206865 | 37 | 0.768 | 1.106 | 0.337 | ZINC02124654 | 41 | 0.710 | 1.103 | 0.392 |
| ZINC04045984 | 37 | 0.644 | 1.092 | 0.448 | ZINC02159335 | 41 | 0.682 | 1.096 | 0.414 |
| ZINC02111169 | 37 | 0.735 | 1.082 | 0.347 | ZINC02122586 | 41 | 0.681 | 1.089 | 0.407 |
| ZINC04084607 | 37 | 0.673 | 1.070 | 0.397 | ZINC02123323 | 41 | 0.683 | 1.089 | 0.406 |
| ZINC02098138 | 37 | 0.648 | 1.054 | 0.406 | ZINC02135809 | 41 | 0.713 | 1.084 | 0.371 |
| ZINC04046507 | 37 | 0.787 | 1.052 | 0.264 | ZINC02123238 | 41 | 0.642 | 1.074 | 0.432 |
| ZINC00244212 | 37 | 0.716 | 1.051 | 0.335 | ZINC02124894 | 41 | 0.624 | 1.068 | 0.443 |
| ZINC00374801 | 37 | 0.732 | 1.034 | 0.301 | ZINC02123503 | 41 | 0.651 | 1.066 | 0.415 |
| ZINC02108985 | 37 | 0.673 | 1.015 | 0.342 | ZINC02112810 | 41 | 0.662 | 1.061 | 0.399 |
| ZINC04045983 | 37 | 0.743 | 1.013 | 0.270 | ZINC02115714 | 41 | 0.676 | 1.049 | 0.373 |
| ZINC04085193 | 37 | 0.633 | 1.009 | 0.375 | ZINC02120102 | 41 | 0.634 | 1.046 | 0.412 |
| ZINC00244216 | 37 | 0.678 | 1.006 | 0.328 | ZINC02117282 | 41 | 0.629 | 1.043 | 0.414 |
| ZINC04090129 | 37 | 0.658 | 0.964 | 0.306 | ZINC02112838 | 41 | 0.644 | 1.027 | 0.383 |
| ZINC00526580 | 37 | 0.724 | 0.968 | 0.244 | **ZINC02132035** | **41** | **0.690** | **0.983** | **0.293** |
| ZINC03852206 | 37 | 0.632 | 0.967 | 0.335 | ZINC00037965 | 41 | 0.639 | 0.973 | 0.333 |
| ZINC05396688 | 37 | 0.663 | 1.261 | 0.598 | ZINC02119782 | 41 | 0.627 | 0.972 | 0.345 |
| ZINC08295759 | 37 | 0.669 | 1.160 | 0.491 | ZINC02131421 | 41 | 0.640 | 0.963 | 0.324 |
| ZINC05439072 | 37 | 0.715 | 1.097 | 0.382 | **ZINC02160434** | **41** | **0.658** | **0.963** | **0.306** |
| ZINC05396078 | 37 | 0.624 | 1.096 | 0.472 | ZINC08298071 | 44 | 0.677 | 1.151 | 0.474 |
| ZINC03841441 | 37 | 0.704 | 1.073 | 0.369 | ZINC08300451 | 44 | 0.641 | 1.116 | 0.475 |
| ZINC03842017 | 37 | 0.682 | 1.073 | 0.392 | ZINC00920376 | 44 | 0.670 | 1.097 | 0.427 |
| ZINC05397647 | 37 | 0.707 | 1.069 | 0.362 | ZINC08623374 | 44 | 0.690 | 1.075 | 0.385 |
| ZINC05433763 | 37 | 0.660 | 1.055 | 0.395 | ZINC00978757 | 44 | 0.668 | 1.070 | 0.402 |
| ZINC03841985 | 37 | 0.690 | 1.034 | 0.344 | ZINC12602191 | 44 | 0.645 | 1.051 | 0.406 |
| ZINC05399378 | 37 | 0.624 | 1.033 | 0.410 | ZINC08298054 | 44 | 0.632 | 1.037 | 0.406 |
| ZINC05433915 | 37 | 0.675 | 1.032 | 0.357 | ZINC08300458 | 44 | 0.638 | 1.021 | 0.383 |
| ZINC05398720 | 37 | 0.643 | 1.029 | 0.386 | ZINC08254146 | 44 | 0.632 | 0.959 | 0.328 |
| ZINC05409706 | 37 | 0.700 | 1.001 | 0.301 | ZINC08254520 | 44 | 0.635 | 0.918 | 0.282 |
| ZINC05399865 | 37 | 0.663 | 0.974 | 0.312 | ZINC08300478 | 44 | 0.647 | 0.918 | 0.271 |
| ZINC05410357 | 37 | 0.630 | 0.971 | 0.341 | ZINC00940005 | 45 | 0.728 | 1.238 | 0.510 |
| ZINC04270597 | 37 | 0.634 | 0.970 | 0.336 | ZINC03847574 | 45 | 0.660 | 1.184 | 0.524 |
| ZINC05440675 | 37 | 0.657 | 0.928 | 0.271 | ZINC00703857 | 45 | 0.667 | 1.115 | 0.447 |
| ZINC05433913 | 37 | 0.631 | 0.898 | 0.267 | ZINC01628259 | 45 | 0.641 | 1.098 | 0.457 |
| ZINC02138459 | 41 | 0.737 | 1.223 | 0.485 | ZINC03846504 | 45 | 0.630 | 1.059 | 0.429 |
| ZINC02117429 | 41 | 0.755 | 1.210 | 0.455 | ZINC01679777 | 45 | 0.741 | 1.039 | 0.298 |
| ZINC00407890 | 41 | 0.692 | 1.186 | 0.494 | ZINC03847575 | 45 | 0.683 | 1.028 | 0.345 |
| ZINC02122938 | 41 | 0.659 | 1.183 | 0.524 | ZINC02122310 | 45 | 0.663 | 1.012 | 0.349 |
| ZINC02114557 | 41 | 0.723 | 1.176 | 0.453 | **ZINC02113150** | **45** | **0.626** | **0.965** | **0.339** |
| ZINC00132662 | 41 | 0.701 | 1.155 | 0.454 | ZINC03846506 | 45 | 0.641 | 0.942 | 0.301 |
| ZINC02125016 | 41 | 0.703 | 1.135 | 0.432 | ZINC03846608 | 45 | 0.649 | 0.936 | 0.286 |
| ZINC02124939 | 41 | 0.649 | 1.132 | 0.483 | ZINC04267198 | 45 | 0.625 | 0.904 | 0.279 |
| ZINC02159206 | 41 | 0.705 | 1.122 | 0.417 | ZINC00518521 | 49 | 0.753 | 1.150 | 0.397 |
| ZINC02125175 | 41 | 0.642 | 1.109 | 0.467 | ZINC01824117 | 49 | 0.729 | 1.088 | 0.359 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| ZINC00518288 | 49 | 0.786 | 1.071 | 0.285 | ZINC01322729 | 50 | 0.637 | 1.119 | 0.482 |
| **ZINC00518468** | **49** | **0.726** | **1.070** | **0.344** | ZINC00281064 | 50 | 0.635 | 1.103 | 0.468 |
| ZINC00518287 | 49 | 0.737 | 1.023 | 0.286 | ZINC00276052 | 50 | 0.649 | 1.099 | 0.450 |
| ZINC01760845 | 49 | 0.662 | 1.003 | 0.342 | ZINC05954633 | 50 | 0.663 | 1.098 | 0.436 |
| ZINC00518513 | 49 | 0.702 | 0.993 | 0.291 | ZINC01824688 | 50 | 0.626 | 1.095 | 0.470 |
| ZINC01790050 | 49 | 0.673 | 0.963 | 0.290 | ZINC00386493 | 50 | 0.670 | 1.090 | 0.420 |
| ZINC03736221 | 50 | 0.646 | 1.245 | 0.599 | ZINC00386492 | 50 | 0.634 | 1.089 | 0.454 |
| ZINC00526756 | 50 | 0.747 | 1.240 | 0.493 | ZINC03736224 | 50 | 0.637 | 1.083 | 0.446 |
| ZINC00281070 | 50 | 0.681 | 1.231 | 0.550 | ZINC00756618 | 50 | 0.674 | 1.078 | 0.403 |
| ZINC02138329 | 50 | 0.640 | 1.231 | 0.591 | ZINC00060358 | 50 | 0.669 | 1.073 | 0.404 |
| ZINC01322395 | 50 | 0.684 | 1.229 | 0.546 | ZINC01825160 | 50 | 0.672 | 1.046 | 0.374 |
| ZINC02131423 | 50 | 0.673 | 1.214 | 0.541 | ZINC00978630 | 50 | 0.649 | 1.021 | 0.372 |
| ZINC00526446 | 50 | 0.678 | 1.213 | 0.535 | ZINC02135841 | 50 | 0.675 | 1.012 | 0.337 |
| ZINC02128903 | 50 | 0.628 | 1.176 | 0.549 | ZINC02096126 | 50 | 0.654 | 0.996 | 0.342 |
| ZINC03736220 | 50 | 0.721 | 1.173 | 0.452 | ZINC02128246 | 50 | 0.671 | 0.995 | 0.324 |
| ZINC02114987 | 50 | 0.704 | 1.141 | 0.436 | ZINC01838604 | 50 | 0.659 | 0.992 | 0.333 |
| ZINC00386494 | 50 | 0.677 | 1.133 | 0.456 | **ZINC02131242** | **50** | **0.644** | **0.990** | **0.346** |
| ZINC00131414 | 50 | 0.803 | 1.129 | 0.326 | ZINC06624000 | 50 | 0.648 | 0.989 | 0.341 |
| ZINC00188073 | 50 | 0.653 | 1.128 | 0.475 | ZINC02102133 | 50 | 0.631 | 0.976 | 0.345 |
| ZINC02100978 | 50 | 0.661 | 1.127 | 0.466 | ZINC02131243 | 50 | 0.638 | 0.944 | 0.306 |
| ZINC00189593 | 50 | 0.680 | 1.126 | 0.446 | ZINC06624017 | 50 | 0.655 | 0.925 | 0.270 |
| ZINC04028994 | 50 | 0.637 | 1.123 | 0.486 | | | | | |

CHAPTER 6

# IDENTIFICATION OF NATURAL EXTRACTS WITH POTENTIAL ANTIDIABETIC PROPERTIES THAT CONTAIN DPP-IV INHIBITORS

## ABSTRACT

Natural extracts play an important role in traditional medicines for the treatment of diabetes mellitus and are also an essential resource for new drug discovery. Dipeptidyl peptidase IV (DPP-IV) inhibitors are potential candidates for the treatment of type 2 diabetes mellitus, and the effectiveness of certain antidiabetic extracts of natural origin could be, at least partially, explained by the inhibition of DPP-IV.

Using an initial set of 29,779 natural products that are annotated with their natural source and an experimentally validated virtual screening procedure previously developed in our lab [1], we have predicted 12 potential DPP-IV inhibitors from 12 different plant extracts (i.e., *Ephedra alata*, *Ephedra distachya*, *Erythrina variegata*, *Galega orientalis*, *Haloxylon salicornicum*, *Pennisetum typhoideum*, *Rauwolfia serpentina*, *Rauwolfia vomitoria*, *Scoparia dulcis*, *Tecoma stans*, *Vinca major* and *Vitis vinifera*) that are known to have antidiabetic activity. Seven of these molecules are identical or similar to molecules with described antidiabetic activity (although their role as DPP-IV inhibitors has not been suggested as an explanation for their bioactivity). Therefore, it is plausible that these 12 molecules could be responsible, at least in part, for the antidiabetic activity of these extracts through their inhibitory effect on DPP-IV. In addition, we also identified as potential DPP-IV inhibitors 6 molecules from 6 different plants with no described antidiabetic activity but that share the same genus as plants with known antidiabetic properties. Moreover, none of the 18 molecules that we predicted as DPP-IV inhibitors exhibits chemical similarity with a group of 2,342 known DPP-IV inhibitors.

Our study identified 18 potential DPP-IV inhibitors in 18 different plant extracts (12 of these plants have known antidiabetic properties, whereas, for the remaining 6, antidiabetic activity has been reported for other plant species from the same genus). Moreover, none of the 18 molecules exhibits chemical similarity with a large group of known DPP-IV inhibitors. In conclusion, our study identified (a) 18 lead-hopping candidates for the development of new DPP-IV inhibitors (corresponding to 13 different chemical scaffolds) and (b) 6 plants with previously undescribed antidiabetic activity that could be new sources for antidiabetic extracts.

## Introduction

Medical plants play an important role in the management of type 2 diabetes mellitus (T2DM) by delaying the development of diabetic complications and correcting metabolic abnormalities [2]. Traditional plant-based remedies have been and are being used by T2DM patients around the world (*e.g.*, patients belonging to the Chinese [3], Indian [4] and Mexican [5] populations), and many scientific studies have confirmed the benefits of medicinal plants with hypoglycemic effects on these patients [6-9]. Furthermore, during the past few years, some of the new bioactive drugs isolated from hypoglycemic plants have been demonstrated to have antidiabetic activity with greater efficacy than synthetic oral hypoglycemic agents used in clinical therapy regimens [10].

The most commonly studied hypoglycemic plants are *Opuntia streptacantha*, *Trigonella foenum-graecum*, *Momordica charantia*, *Ficus bengalensis*, *Polygala senega* and *Gymnema sylvestre* [10]. Despite their long tradition of use worldwide, few of these plants have been tested in modern, large-scale, clinical-type trials to determine their efficacies. However, it is clear that more research needs to be undertaken on these and other medicinal plants with hypoglycemic effects because, in most cases, the bioactive compounds and their modes of action still remain unclear.

Numerous mechanisms of antidiabetic action have been proposed for extracts of the previous mentioned plants, some of them relate to their ability to stimulate insulin secretion [11]. Regarding the stimulation of insulin secretion, one target of interest for the antidiabetic action of these extracts is the serine protease dipeptidyl peptidase-IV (DPP-IV; EC 3.4.14.5) because the inhibition of DPP-IV has been shown to be an appropriate treatment for T2DM [12]. DPP-IV specifically removes N-terminal dipeptides from substrates containing proline or alanine as the second residue, transforming them into inactive or even antagonistic species. The most important substrates of DPP-IV are incretins, such as glucagon-like peptide-1 (GLP-1) and glucose-dependent insulinotropic polypeptide (GIP), which stimulates insulin secretion [13]. Incretin hormones are intestinal hormones that are released in response to nutrient ingestion and that potentiate the glucose-induced insulin response. Therefore, GLP-1 stimulates insulin biosynthesis and secretion, reduces glucagon release, slows gastric emptying, reduces appetite, and stimulates the regeneration and differentiation of islet B-cells [14]. On the other hand, GIP is involved in glucose metabolism by enhancing insulin secretion [15]. Both peptides have short half-lives because of their rapid degradation by DPP-IV. Therefore, inhibiting DPP-IV prolongs the action of GLP-1 and GIP, which, in turn, improves glucose homeostasis with a low risk of hypoglycemia [12].

The first DPP-IV inhibitor on the market was sitagliptin (by Merck & Co.) [16], which was followed by the structurally similar vildagliptin (by Novartis) [17] and saxagliptin (by Bristol-Myers Squibb and AstraZeneca) [18]. The efficacy and safety profile of DPP-IV inhibitors have been promising and advantageous to date. In contrast to sulfonylureas and other antidiabetic drugs, DPP-IV inhibitors do not have an intrinsic risk of inducing hypoglycemia, and they are body-weight neutral. Their tolerability profile is good, and no specific adverse reactions have been reported [12].

The DPP-IV binding site is highly druggable in the sense that tight, specific binding to the enzyme can be achieved with small molecules with drug-like physicochemical properties [19,20]. The two key binding-site areas for the intermolecular interaction of DPP-IV and reversible inhibitors of non-peptide nature are the lipophilic S1 pocket (formed by Tyr631, Val656, Trp659, Tyr662, Tyr666 and Val711) and the negatively charged Glu205/206 pair [20]. We have recently used coordinates from complexes between DPP-IV and potent reversible inhibitors of non-peptide nature to derive a structure-based common pharmacophore that defines a common background for DPP-IV inhibition [1]. This pharmacophore is part of a virtual screening (VS) workflow that also includes protein-ligand docking studies and shape and electrostatic-potential comparisons [1]. We have applied this VS workflow to the 89,425 molecules present in the natural products subset of the ZINC database (http://wiki.bkslab.org/index.php/Natural_products_database), and we found that 446 of these molecules would inhibit DPP-IV with good ADMET properties. Notably, when these 446 molecules were merged with 2,342 known DPP-IV inhibitors, and the resulting set was classified into 50 clusters according to chemical similarity, there were 12 clusters that contained only natural products for which no DPP-IV inhibitory activity has been previously reported. Nine molecules from 7 of these 12 clusters were then selected for *in vitro* activity testing, and 7 out of the 9 molecules were shown to inhibit DPP-IV (the remaining two molecules could not be solubilized, preventing the evaluation of their DPP-IV inhibitory activity) [1].

The goal of the present work was to identify natural extracts with known antidiabetic activity that contain at least one molecule that we predict to be a DPP-IV inhibitor through a slightly modified version of the VS workflow described above [1]. Therefore, in this study, we provide new information about the active molecules in some natural extracts with antidiabetic properties and suggest that, at least in part, the mode of action of these molecules involves stimulating insulin secretion through the inhibition of DPP-IV. We also provide a list of plants with no previously described antidiabetic activity that may contain DPP-IV inhibitors and that are related to plants with known antidiabetic activity. These plants represent a new source of potential antidiabetic extracts. In addition, the new DPP-IV inhibitors that we identified are chemically different from known DPP-IV inhibitors, and therefore, they could be used as lead-hopping candidates for the development of new antidiabetic drugs.

## Results and discussion

**Virtual screening description and application.** We used a slightly modified version of a VS workflow that was previously developed and experimentally validated [1] to identify DPP-IV inhibitors in a large in-house database of natural products (NPs) annotated with their natural source.

The VS workflow (see Figure 1) consisted of several sequential steps in which the output molecules of one step were the input molecules for the next step and so on. Central in this workflow is one structure-based common pharmacophore that captures the key intermolecular interactions needed for drugs to inhibit DPP-IV; this pharmacophore is formed by 2 mandatory sites (*i.e.*, one positive/donor and one hydrophobic/aromatic ring) and 5 optional sties (*i.e.*, two hydrogen-bond acceptors and three hydrophobic/aromatic ring sites). Both mandatory sites interact with crucial molecular anchors for DPP-IV inhibition (*i.e.*, the hydrophobic/aromatic ring interacts with the hydrophobic S1 pocket at the DPP-IV binding site, and the positive/donor site interacts with the Glu205/Glu206 dyad [20]). Briefly, the VS workflow consists of (1) comparing ligand conformers to the common pharmacophore by allowing reorientation of the conformers to determine if they match the pharmacophore; (2) using ligands with at least one hit in the previous filter in a protein-ligand rigid-docking study and docking them onto the ligand binding site of the DPP-IV conformation present in the 3C45 PDB file; (3) comparing the resulting docking conformations to the structure-based common pharmacophore without reorienting the poses; and (4) submitting the poses that were hits in the previous filter to a shape and electrostatic-potential comparison with the experimental pose of the DPP-IV inhibitor in the PDB file 3C45 (the previously developed VS workflow [1] was altered for the current work to use slightly lower threshold values for the electrostatic and shape comparisons).

This VS protocol was applied to an in-house database of 29,779 NPs with appropriate ADME/Tox properties. The first filter found that 10,883 molecules in our database have at least one conformer that after proper reorientation, matches the pharmacophore (see Figure 1). Only 332 out of these 10,883 molecules have docked conformations that without reorientation, are able to match the pharmacophore (see Figure 1). This reduction is useful because it discards those molecules that are predicted to bind in a non-productive way to the DPP-IV binding site. Finally, the later filter (*i.e.*, electrostatic and shape similarity screening) aims to smooth differences in chemical structures and translate them into criteria important for their intermolecular interactions with the ligand-binding site. This filter has been reported to be a valuable VS tool for the discovery of novel scaffolds [21], and in our case, it was applied to rescore the 332 hits that survived the previous filter. Consequently, only those molecules that had at least one docked pose that met the following conditions in the

comparisons made by EON were predicted to inhibit DPP-IV: (a) *ET_pb* ≥ 0.468 and (b) ST ≥ 0.237. Interestingly, the fact that DPP-IV inhibitors (a) have a significant positive electrostatic potential in the region that interacts with the Glu205/Glu206 dyad (see Figure 1) and (b) that this ligand area matches the mandatory positive/donor site justifies the dominance of the electrostatic contribution over the shape contribution in the selected thresholds. Finally, the VS workflow identified 84 molecules with potential DPP-IV inhibitory activity (see Figure 1).

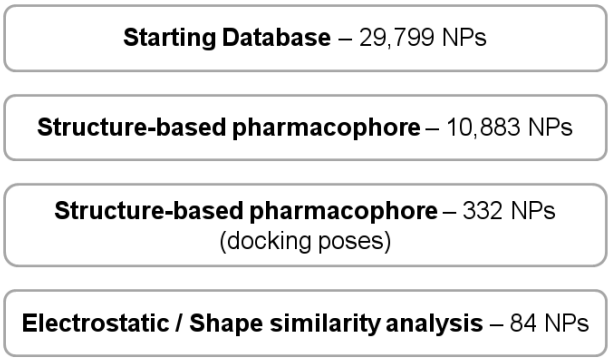



**Figure 1.** The VS workflow used in the present work. The data beside each VS step show the number of molecules that survived it.

**Virtual screening hits in natural extracts with known antidiabetic activity.** According to the information available in our in-house NPs database, the 84 molecules that were predicted by the VS workflow as potential DPP-IV inhibitors have been isolated from 139 different natural sources. Interestingly, a systematic bibliographic search of PubMed (http://www.pubmed.org) revealed that the extracts of 12 out of these 139 natural sources have been reported to exhibit antidiabetic activity (see Table 1). Moreover, among these 12 sources we found 12 VS hits that may, through their role as DPP-IV inhibitors, contribute to the observed antidiabetic activity of their corresponding extracts (see Table 1). In fact, a search using SciFinder (http://www.cas.org/products/sfacad) revealed that 6 out of these 12 natural compounds correspond (or are similar in chemical structure) to molecules with known antidiabetic properties (see Table 1). This finding further validates our methodology (in addition to the experimental validation of the original VS workflow [1]) and suggests that the mode of action of these molecules could involve DPP-IV inhibition. The remaining 6 natural compounds not previously reported to have antidiabetic properties represent new molecules that may also exhibit this bioactivity. In the next paragraphs, the most significant compounds found in these 12 antidiabetic extracts are discussed:

- N-Nororientaline (CAS number 29079-44-5, see Table 1) has been isolated from five plant species from *genus Erythrina* (*i.e.*, *Erythrina variegata* [22], *Erythrina crystagalli*, *Erythrina indica*, *Erythrina poeppigiana* [22], and *Erythrina arborescens*) and was identified as a DPP-IV inhibitor by our VS procedure. Extracts from *Erythrina variegata*, *Erythrina senegalensis*, *Erythrina addisoniae*, *Erythrina abyssinica*, and *Erythrina mildbraedii* are reported to enhance antidiabetic activity [23-27]. Isoprenylated flavonoids isolated from *Erythrina mildbraedii* and prenylflavonoids isolated from the *Erythrina senegalensis* roots have been described as inhibitors of two other proteins frequently targeted in T2DM treatment (*i.e.*, protein tyrosine phosphatase-1B [25,26] for the former class of molecules and acyl CoA:diacylglycerol acyltransferase [24] for the latter class). All of these flavonoid compounds are chemically related to our hit. Therefore, the antidiabetic action of extracts from these plants may be the result of more than one bioactive component and mode of action.

- Tecostamine, which is found in *Tecoma stans*, has hypoglycemic properties similar to those of tecomine [28]. Our results therefore suggest that the hypoglycemic properties of tecostamine could be mediated by the inhibition of DPP-IV. In addition, the *Tecoma stans* aqueous extract posses at least four antidiabetic-related activities (*i.e.*, intestinal α-glucosidase inhibition, post-prandial antihyperglycemic, hypocholesterolemic and hypotriglyceridemic effects). Therefore, although it is possible that most of these activities are exerted by the phenolic compounds present in the *Tecoma stans* aqueous extract, bioguided studies are necessary to confirm this hypothesis [29].

- Epinephrine (also known as adrenaline), which is found in *Scoparia dulcis*, has been reported to improve hypoglycemia [30]. *Scoparia dulcis* has been described as a folk-medicinal plant and has been traditionally used as a remedy for diabetes mellitus in India and for hypertension in Taiwan [31]. From Indian *Scoparia dulcis*, an antidiabetic compound named amellin was also isolated and characterized by Nath [32].

- From the same family as epinephrine, two additional compounds were predicted by our VS to be potential DPP-IV inhibitors. These compounds are (+)-pseudoephedrine (CAS number 90-82-4) and (-)-ephedrine (CAS number 299-42-3), and both have also been reported to have hypoglycemic activity [33]. Both molecules are found in several *Ephedra* species (*Ephedra alata* [34], *Ephedra distachya* [35], *Ephedra equisetina* [36], *Ephedra gerardiana* [35], *Ephedra shennungiana* [37], *Ephedra sinica*, *Ephedra vulgaris* and *Ephedra pflanze*). However, *Ephedra distachya* and *Ephedra alata* are the only species reported to have antidiabetic properties [33,38]; therefore, we

proposed that the remaining *Ephedra* species may be new sources of antidiabetic extracts.

- The molecules ajmaline and isosandwichine, which are enantiomers with the same CAS number, 509-37-5, are found in several *Rauwolfia* species. *Rauwolfia vomitoria* has been investigated for the content of alkaloids, especially those alkaloids with hypotensive and anti-inflammatory properties [39] in addition to the antidiabetic properties. *Rauwolfia serpentina* is also used as antidiabetic extract [40]. The remaining species that contain these molecules, such as *Rauwolfia canescens* [41], *Rauwolfia degeneri*[42], *Rauwolfia densiflora* [43], *Rauwolfia heterophylla* [44], *Rauwolfia indecora*, *Rauwolfia obscura* [45] and *Rauwolfia tetraphylla* [42], are putative antidiabetic extracts.

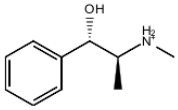
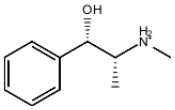
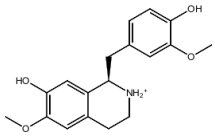
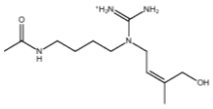
- Serpinine (CAS number 509-38-6) is isolated from *Vinca major* and several *Rauwolfia* species (*Rauwolfia obscura* [45], *Rauwolfia tetraphylla*, *Rauwolfia serpentina* [46] and *Rauwolfia sellowii* [1]), belongs to the same cluster as ajmaline and isosandwichine (*i.e.*, cluster 78). Therefore, they share similar chemical structures and natural sources. Moreover, *Vinca major* organic leaf extract strongly stimulates glucose utilization [47].

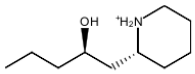
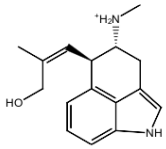
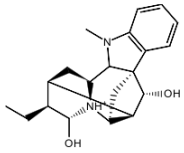
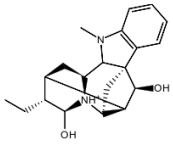
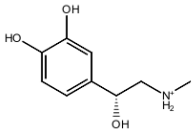
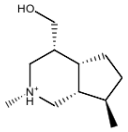
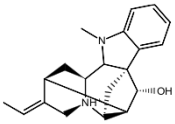
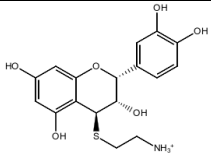
- One interesting hit predicted to be a DPP-IV inhibitor is an epicatechin derivate that is found in *Vitis vinifera*. An antihyperglycemic effect in streptozotocin-induced diabetic rats and insulinomimetic activity in insulin-sensitive cell lines have been described for grape seed procyanidin extracts (GSPE) [48]. In addition, it has been demonstrated that oligomeric procyanidins from GSPE activate the insulin receptor by interacting with and inducing the phosphorylation of the insulin receptor and that this interaction leads to increased glucose uptake [49]. Moreover, several epicatechin derivates have been reported to have antidiabetic properties (the most studied of which is epigallocatechin gallate [50]). Some findings demonstrate that epigallocatechin gallate may be a novel, plant-derived compound capable of reducing the risk of type 1 diabetes [51]. Therefore, the DPP-IV inhibition induced by this epicatechin derivate may contribute to the antihyperglycemic effect of GSPE [48].

- The remaining 3 molecules predicted to be DPP-IV inhibitors through our VS workflow and that are found in extracts with described antidiabetic properties are hydroxysmirnovine from *Galega orientalis* [52], (-)-halosaline (CAS number 26648-71-5) from *Haloxylon salicornicum* [53] and isochanoclavin-(I) (CAS number 1150-44-3) from *Pennisetum typhoideum* [54] (see Table 1). Our results suggest that these molecules could be DPP-IV inhibitors and

that extracts containing these molecules could potentiate the glucose-induced insulin response by prolonging the half-lives of GLP-1 and GIP incretins, due to the inhibition of DPP-IV. This information is novel and relevant, as no mechanism that explains the antidiabetic properties of these extracts has been previously suggested.

**Table 1.** Natural extracts with reported antidiabetic activity that contain molecules that were predicted to be DPP-IV inhibitors by our VS protocol. The first column shows the 2D structure of each molecule and, when available, the corresponding common name and/or CAS number. The second column shows the number of the cluster in which the corresponding molecule was classified when its structure was compared with those of a group of 2,342 known DPP-IV inhibitors. The third column shows the scientific name of one of the sources in which the antidiabetic activity has been reported (rows in that table are alphabetically sorted based on this column). Bibliographic references for each molecule are divided into three columns in which (a) the first column presents papers that describe the purification of the molecule from the corresponding extract; (b) the second column lists papers that describe the antidiabetic activity of the corresponding extract; and (c) the third column lists papers, when available, that describe the antidiabetic activity of the corresponding molecule or one that is very similar to it.

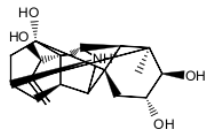
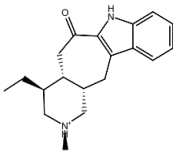
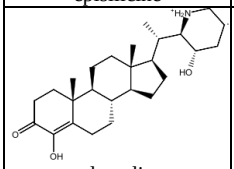
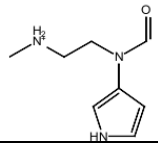
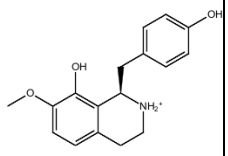
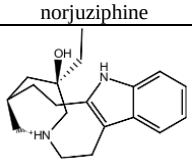
| **Molecule CAS number or Name** | **Cluster** | **Extract** | **Ref. Isolation Molecule from Extract** | **Ref. Antidiabetic Extract** | **Ref. Antidiabetic Molecule** |
|---|---|---|---|---|---|
|  (+)-pseudoephedrine (90-82-4) | 86 | *Ephedra alata* | [34] | [38] | [33] |
|  (-)-ephedrine (299-42-3) | 86 | *Ephedra distachya* | [35] | [33] | [33] |
|  N-nororientalin (29079-44-5) | 89 | *Erythrina variegata* | [22] | [23] | [24-26] |
|  hydroxysmirnovine | 31 | *Galega orientalis* | [52] | [67] | |

| | | | | | |
|---|---|---|---|---|---|
|  (-)-halosaline (26648-71-5) | 79 | *Haloxylon salicornicum* | [53] | [38] | |
|  isochanoclavin-(I) (1150-43-2) | 17 | *Pennisetum typhoideum* | [54] | [68] | |
|  ajmaline (509-37-5) | 78 | *Rauwolfia serpentina* | [46] | [40] | |
|  isosandwichine (509-37-5) | 78 | *Rauwolfia vomitoria* | [69] | [39] | |
|  epinephrine (51-43-4) | 86 | *Scoparia dulcis* | [70] | [31] | [30] |
|  tecostanine | 98 | *Tecoma stans* | [71] | [29] | [28] |
|  serpinine (509-38-6) | 78 | *Vinca major* | [54] | [47] | |
|  epicatechin derivate | 91 | *Vitis vinifera* | [72] | [48] | [50] |

**Virtual screening hits in natural extracts with no described antidiabetic activity.** Taking into account that extracts from closely related species of the same *genus* may share a high number of components, we also determined if any of our VS hits that were isolated from plants with no described antidiabetic activity belong to the same *genus* as species with known antidiabetic properties. We identified 6 molecules isolated from 6 different plants, *Aconitum japonicum*, *Ervatamia officinalis*, *Solanum nudum*, *Solanum sodomaeum*, *Stephania cepharantha* and *Tabernaemontana eglandulosa* (see Table 2), that meet these criteria. The related species with described antidiabetic properties are *Aconitum carmichaelii* [33], *Aconitum moschatum* [2], *Aconitum violaceum* [2], *Ervatamia microphylla* [55], *Solanum lycocarpum* [56], *Solanum nigrum* [57], *Solanum xanthocarpum* [58], *Stephania hernandifolia* [59], *Stephania glabra* [60], *Stephania tetrandra* [61], and *Tabernaemontana divaricata* [55]. Therefore, it is plausible to hypothesize that these 6 plants could also have antidiabetic properties mediated, at least partially, by the inhibition of DPP-IV.

**Finding new scaffolds of natural origin for DPP-IV inhibitors.** The 18 molecules in Tables 1 and 2 that were predicted to be DPP-IV inhibitors are of interest. To quantify the number of new scaffolds for DPP-IV inhibitors that were identified in our study, we merged the 18 VS hits with 2,342 known DPP-IV inhibitors, and the resulting set was classified according to structural similarity into 99 clusters (results not shown). Interestingly, the 18 hits were classified in 13 clusters (see Tables 1 and 2) that do not contain known DPP-IV inhibitors (results not shown). Thus, these 18 predicted DPP-IV inhibitors correspond to 13 different chemical scaffolds that are unrelated to those present in known DPP-IV inhibitors, and, consequently, these new scaffolds could be used either in lead-hopping experiments to identify new DPP-IV inhibitors or in structure-activity studies to identify natural-product derivates with stringer DPP-IV inhibition activity than the original NPs from which they are derived.

**Table 2.** Natural extracts with no described antidiabetic activity (but from the same *genus* as plants with extracts with described antidiabetic activity) that contain molecules that are predicted to be DPP-IV inhibitors by our VS protocol. The first column shows the 2D structure of each molecule and, when available, the corresponding common name or CAS number. The second column shows the number of the cluster in which the corresponding molecule was classified when its structure was compared with those of a group of 2,342 known DPP-IV inhibitors. The third column lists the natural extracts from which the VS hits have been purified (rows in that table are alphabetically sorted based on this column). The fourth column lists the papers that describe the purification of the each molecule from the corresponding extract. The fifth column shows which are the extracts from the same *genus* where the antidiabetic activity has been described. Finally, the last column lists papers that describe the antidiabetic activity of the corresponding extract.

| Molecule CAS number or Name | Cluster | Extract | Ref. Isolation Molecule from Extract | Antidiabetic Extract | Ref. Antidiabetic Extract |
|---|---|---|---|---|---|
|  30373-79-6 | 96 | *Aconitum japonicum* | [73] | *Aconitum carmichaelii* | [74] |
| | | | | *Aconitum moschatum* | [2] |
| | | | | *Aconitum violaceum* | [2] |
|  episilicine | 21 | *Ervatamia officinalis* | [75] | *Ervatamia microphylla* | [55] |
|  solanudine | 97 | *Solanum nudum* | [76] | *Solanum lycocarpum* | [56] |
| | | | | *Solanum nigrum* | [57] |
|  | 2 | *Solanum sodomaeum* | [77] | *Solanum xanthocarpum* | [58] |
|  norjuziphine | 89 | *Stephania cepharantha* | [78] | *Stephania hernandifolia* | [59] |
| | | | | *Stephania glabra* | [60] |
| | | | | *Stephania tetrandra* | [61] |
|  19637-92-4 | 20 | *Tabernaemontana eglandulosa* | [79] | *Tabernaemontana divaricata* | [55] |

## Conclusions

In a previous study [1], we developed a VS workflow that was able to distinguish successfully molecules that inhibit DPP-IV and molecules that do not inhibit this enzyme. We experimentally demonstrated that our VS protocol was able to identify DPP-IV inhibitors that **(a)** were not structurally related to any known molecule that inhibits DPP-IV and **(b)** have never been reported to have antidiabetic activity. In the

present work, we applied a slightly modified version of this VS workflow to an in-house database of 29,779 NPs annotated with their corresponding natural source(s). From this initial set of NPs, our VS procedure identified as potential DPP-IV inhibitors 84 hit molecules that have been isolated from 96 different natural extracts. Interestingly, after an exhaustive bibliographic search, our results demonstrate that we are able to predict (a) 12 DPP-IV inhibitors that are present in 12 plant extracts with known antidiabetic activity and (b) 6 DPP-IV inhibitors that are present in 6 different plants species with no described antidiabetic activity but that share the same *genus* as plants with known antidiabetic properties (consequently, it could be suggested that these plants represent a potential new source of antidiabetic extracts). Moreover, none of these 18 hits exhibits chemical similarity with 2,342 known DPP-IV inhibitors, and, therefore, it is expected that a significant number of these hits could be lead-hopping candidates for the development of new DPP-IV inhibitors. At this point, it is also interesting to note that the analysis of the chemical structures of these 18 NP hits revealed that the majority of them are alkaloids containing basic nitrogen atoms (essential for proper interaction with the Glu205/Glu206 dyad). Moreover, our results provide a new hypothesis about the mechanisms by which, at least partially, these 12 extracts exert their antidiabetic effects (*i.e.*, improving glucose homeostasis by prolonging the activity of GLP-1 and GIP through DPP-IV inhibition).

Lastly, we predicted that there are 77 other extracts with no described antidiabetic activity that contain at least one out of 65 VS hits. Consequently, our work opens the door to the discovery of new antidiabetic extracts of natural origin that could be of use, for example, in the design of functional foods aimed at preventing/treating T2DM. Therefore, the characterization of such extracts merits further attention, and such work is currently underway.

## Experimental section

**Initial dataset of natural compounds used.** The database of NPs that was screened by the VS workflow contains 29,779 NPs from different origins (*e.g.*, plants, and fungi) with appropriate ADME/Tox properties and no chiral ambiguities. An important characteristic of this database is that each molecule is annotated with (a) the natural sources from which it has been obtained and (b) the bibliographic references that describe how to extract the relevant molecule from each natural source. The 3D structures of the molecules in this NP database were processed with LigPrep v2.3 (Schrödinger LLC., Portland, USA; http://www.schrodinger.com) using the following parameters: (a) the force field used was OPLS 2005; (b) all possible ionization states at pH 7.0±2.0 were generated with Ionizer; (c) the desalt option was activated; (d) tautomers were generated for all ionization states at pH 7.0±2.0; (e) chiralities were determined from the 3D structure; and (f) one low-energy ring conformation per ligand was generated. Conformations and sites for the resulting ligand structures were

determined during the generation of the corresponding Phase [62] databases with the ***Generate Phase Database*** graphic front-end. Default parameter values were used during this conformer generation process, with the exception of the maximum number of conformers per structure, which was increased from 100 (the default value) to 200. The conformer sites were generated with definitions made by adding to the default built-in Phase definitions the ability to consider aromatic rings as hydrophobic groups.

**Virtual screening workflow.** The VS workflow used in this work is the same as that described previously [1], except that the conditions of the last filter (*i.e.*, the shape and electrostatic potential comparison) were slightly different, as outlined below.

The VS protocol used a structure-based pharmacophore that was built by (1) selecting from the PDB those reliable complexes of human DPP-IV and potent inhibitors of non-peptide nature (*i.e.*, $IC_{50} \leq 10$ nM) that bind reversibly to the enzyme; (2) using their corresponding DPP-IV coordinates to guide the superposition of the remaining PDB files (the resulting re-oriented coordinates for these PDB files were also used in the pharmacophore-based searches, protein-ligand docking studies and shape and electrostatic-potential comparisons of the VS workflow); (3) using the resulting coordinates to derive the corresponding energetic structure-based pharmacophores; and (4) building the common structure-based pharmacophore for reversible DPP-IV inhibition by prioritizing energetically favorable features over energetically weaker ones. The resulting pharmacophore consists of two compulsory sites (one positive/donor and one hydrophobic/aromatic ring) and five optional sites (*i.e.*, two acceptor sites and three hydrophobic/aromatic ring sites) and was completed with receptor-based excluded volumes that schematically represent the location of the DPP-IV residues that form the binding pocket in the PDB file 3C45.

The first step of the VS workflow uses the common structure-based pharmacophore to screen the conformer database with Phase v3.1 (Schrödinger LLC., Portland, USA; http://www.schrodinger.com) and allows the reorientation of the conformers to determine if they match the pharmacophore. Only those ligands with at least one conformer that matches the two compulsory sites of the common pharmacophore and at least one of the optional sites (without sterically colliding with the excluded volumes) *survive* this VS step. These ligands were docked onto the binding pocket of the PDB file 3C45 with eHiTS v2009 (SimBioSys Inc., Toronto, Canada; http://www.simbiosys.ca/ehits) [63] using default docking conditions, with the exception of the length of the sides of the cubic box encompassing the DPP-IV binding site, which was increased from 10 Å to 15 Å. Next, the resulting docked poses were again filtered with Phase through the common pharmacophore using the same filtering conditions as in the first Phase run (with the exception that no reorientation of the docked poses was allowed during the search). Thus, only docking poses compatible with the pharmacophore *survived* this filter. Finally, in the last step of the

VS protocol, the poses that were hits in this second pharmacophore screen were submitted to a shape and electrostatic potential comparison with the DPP-IV inhibitor present in the 3C45 PDB file to rescore the hits. This comparison was completed with EON v2.0.1 (OpenEye Scientific Software, Inc., Santa Fe, New Mexico, USA; http://www.eyesopen.com) and used the Electrostatic Tanimoto combo (ET_combo) score as the similarity criterion. The ET_combo score is the sum of two calculations: (a) the Shape Tanimoto (ST) score, which is a quantitative measure of three-dimensional overlap (where 1 corresponds to a perfect overlap; *i.e.* the same shape), and (b) the Poisson-Boltzmann Electrostatic Tanimoto (ET_pb) score, which compares the electrostatic potential of two small molecules and ranges from 1 (identical potential) to a negative value that results from the overlap of positive and negative charges. In this work, we determined the EON scores by comparing the inhibitor in 3C45 with the inhibitors found in a group of 24 PDB complexes of DPP-IV and reversible drugs that inhibit this protein (*i.e.*, 1N1M, 20GZ, 2FJP, 2HHA, 2I78, 2IIT, 2IIV, 2OLE, 2ONC, 2OQI, 2OQV, 2QOE, 2QT9, 2QTB, 2RGU, 2RIP, 3C43, 3CCC, 3D4L, 3F8S, 3H0C, 3HAB, and 3HAC). This comparison of the experimental poses of DPP-IV inhibitors yielded threshold values for DPP-IV inhibition that were slightly lower than the ones used in the original VS workflow [1] ($ET\_pb \geq 0.468$ and $ST \geq 0.237$ instead of $ET\_pb \geq 0.623$ and $ST \geq 0.244$), probably because the original threshold values were obtained exclusively using potent DPP-IV inhibitors (*i.e.*, $IC_{50} \leq 10$ nM); in this study, this condition was relaxed slightly. After rescoring with EON, only those NPs with at least one conformation with $ET\_pb \geq 0.468$ and $ST \geq 0.237$ relative to 3C45's inhibitor were considered to be DPP-IV inhibitor candidates.

**Structural similarity analysis.** The molecules that survived the electrostatics/shape similarity filter were merged with 2,342 known DPP-IV inhibitors obtained from the BindingDB database [64] and then clustered using Canvas v1.2 (Schrödinger LLC., Portland, USA; http://www.schrodinger.com). MOLPRINT2D fingerprints [65], using a fingerprint precision of 32 bits, were calculated for each molecule, and hierarchical clustering, based on Tanimoto similarities, was subsequently obtained. The number of clusters obtained was defined using the Kelley criterion [66], corresponding to a Tanimoto coefficient of 0.775 in this case.

## References

1. Guasch L, Ojeda MJ, González-Abuín N, Sala E, Ceretó A, et al. (submitted) Identification of novel human dipeptidyl peptidase-IV inhibitors of natural origin (part I): Virtual screening and activity assay. .

2. Howes M, Simmonds M. (2005) Plants used in the treatment of diabetes. In: Soumyanath A, editor. Traditional Medicines for Modern Times. Press.

3. Yin J, Zhang H, Ye J. (2008) Traditional chinese medicine in treatment of metabolic syndrome. Endocr Metab Immune Disord Drug Targets 8(2): 99-111.

4. Modak M, Dixit P, Londhe J, Ghaskadbi S, Paul A Devasagayam T. (2007) Indian herbs and herbal drugs used for the treatment of diabetes. J Clin Biochem Nutr 40(3): 163-173.

5. Andrade-Cetto Adolfo, Heinrich M. (2005) Mexican plants with hypoglycaemic effect used in the treatment of diabetes. J Ethnopharmacol 99(3): 325-348.

6. Haque N, Salma U, Nurunnabi T, Uddin M, Jahangir M, et al. (2011) Management of type 2 diabetes mellitus by lifestyle, diet and medicinal plants. Pak J Biol Sci 14(1): 13-24.

7. Malviya N, Jain S, Malviya S. (2010) Antidiabetic potential of medicinal plants. Acta Pol Pharm 67(2): 113-118.

8. Prabhakar PK, Doble M. (2008) A target based therapeutic approach towards diabetes mellitus using medicinal plants. Curr Diabetes Rev 4(4): 291-308.

9. Qi L, Liu E, Chu C, Peng Y, Cai H, et al. (2010) Anti-diabetic agents from natural products--an update from 2004 to 2009. Curr Top Med Chem 10(4): 434-457.

10. Bnouham M, Ziyyat A, Mekhfi H, Tahri A, Legssyer A. (2006) Medicinal plants with potential antidiabetic activity - A review of ten years of herbal medicine research (1990-2000). Int J Diabetes Metab 14(1).

11. Prabhakar PK, Doble M. (2011) Mechanism of action of natural products used in the treatment of diabetes mellitus. Chin J Integr Med 17(8): 563-574.

12. Havale SH, Pal M. (2009) Medicinal chemistry approaches to the inhibition of dipeptidyl peptidase-4 for the treatment of type 2 diabetes. Bioorg Med Chem 17(5): 1783-1802.

13. Mentlein R, Gallwitz B, Schmidt W. (1993) Dipeptidyl-peptidase IV hydrolyses gastric inhibitory polypeptide, glucagon-like peptide-1(7-36)amide, peptide histidine methionine and is responsible for their degradation in human serum. Eur J Biochem 214(3): 829-835.

14. Brubaker PL, Drucker DJ. (2004) Minireview: Glucagon-like peptides regulate cell proliferation and apoptosis in the pancreas, gut, and central nervous system. Endocrinology 145(6): 2653-2659.

15. Meier JJ, Nauck MA, Schmidt WE, Gallwitz B. (2002) Gastric inhibitory polypeptide: The neglected incretin revisited. Regul Pept 107(1-3): 1-13.

16. Kim D, Wang L, Beconi M, Eiermann GJ, Fisher MH, et al. (2005) (2R)-4-oxo-4-[3-(trifluoromethyl)-5,6-dihydro[1,2,4]triazolo[4,3-a]pyrazin-7(8H)-yl]-1-(24,5-trifluorophenyl)butan-2-amine: A potent, orally active dipeptidyl peptidase IV inhibitor for the treatment of type 2 diabetes. J Med Chem 48(1): 141-151.

17. Villhauer EB, Brinkman JA, Naderi GB, Burkey BF, Dunning BE, et al. (2003) 1-[[(3-hydroxy-1-adamantyl)amino]acetyl]-2-cyano-(S)-pyrrolidine: A potent, selective, and orally bioavailable dipeptidyl peptidase IV inhibitor with antihyperglycemic properties. J Med Chem 46(13): 2774-2789.

18. Augeri DJ, Robl JA, Betebenner DA, Magnin DR, Khanna A, et al. (2005) Discovery and preclinical profile of saxagliptin (BMS-477118): A highly potent, long-acting, orally active dipeptidyl peptidase IV inhibitor for the treatment of type 2 diabetes. J Med Chem 48(15): 5025-5037.

19. Zettl H, Schubert-Zsilavecz Manfred, Steinhilber D. (2010) Medicinal chemistry of incretin mimetics and DPP-4 inhibitors. ChemMedChem 5(2): 179-185.

20. Kuhn B, Hennig M, Mattei P. (2007) Molecular recognition of ligands in dipeptidyl peptidase IV. Current Topics in Medicinal Chemistry 7(6): 609-619.

21. Sala E, Guasch L, Iwaszkiewicz J, Mulero M, Salvado M, et al. (2011) Identification of human IKK-2 inhibitors of natural origin (part I): Modeling of the IKK-2 kinase domain, virtual screening and activity assays. PloS One 6(2): e16903.

22. Ito K, Haruna M, Furukawa H. (1975) [Studies on the erythrina alkaloids. X. alkaloids of several erythrina plants from singapore]. Yakugaku Zasshi 95(3): 358-362.

23. Kumar A, Lingadurai S, Shrivastava TP, Bhattacharya S, Haldar PK. (2011) Hypoglycemic activity of erythrina variegata leaf in streptozotocin-induced diabetic rats. Pharm Biol 49(6): 577-582.

24. Oh WK, Lee CH, Seo JH, Chung MY, Cui L, et al. (2009) Diacylglycerol acyltransferase-inhibitory compounds from erythrina senegalensis. Arch Pharm Res 32(1): 43-47.

25. Bae EY, Na M, Njamen D, Mbafor JT, Fomum ZT, et al. (2006) Inhibition of protein tyrosine phosphatase 1B by prenylated isoflavonoids isolated from the stem bark of erythrina addisoniae. Planta Med 72(10): 945-948.

26. Na M, Jang J, Njamen D, Mbafor JT, Fomum ZT, et al. (2006) Protein tyrosine phosphatase-1B inhibitory activity of isoprenylated flavonoids isolated from erythrina mildbraedii. J Nat Prod 69(11): 1572-1576.

27. Nguyen P, Nguyen T, Dao T, Kang H, Ndinteh D, et al. (2010) AMP-activated protein kinase (AMPK) activation by benzofurans and coumestans isolated from erythrina abyssinica. J Nat Prod 73(4): 598-602.

28. Hammouda Y, Rashid K, Amer S. (1964) Hypoglycaemic properties of tecomine and tecostamine. J Pharm Pharmacol 16: 833-834.

29. Aguilar-Santamaria L, Ramirez G, Nicasio P, Alegria-Reyes C, Herrera-Arellano A. (2009) Antidiabetic activities of tecoma stans (L.) juss. ex kunth. J Ethnopharmacol 124(2): 284-288.

30. Ly TT, Hewitt J, Davey RJ, Lim EM, Davis EA, et al. (2011) Improving epinephrine responses in hypoglycemia unawareness with real-time continuous glucose monitoring in adolescents with type 1 diabetes. Diabetes Care 34(1): 50.

31. Latha M, Pari L, Sitasawad S, Bhonde R. (2004) Scoparia dulcis, a traditional antidiabetic plant, protects against streptozotocin induced oxidative stress and apoptosis in vitro and in vivo. J Biochem Mol Toxicol 18(5): 261-272.

32. Nath CM, Chakrabority KM, Brahmachari DH. (1945) Investigations on the new antidiabetic principle (amellin); its role in the reduction of acetone bodies and the

increase of alkali reserve of the blood of diabetics. Ann Biochem Exp Med 5(3): 101-104.

33. Konno C, Mizuno T, Hikino H. (1985) Isolation and hypoglycemic activity of ephedrans A, B, C, D and E, glycans of ephedra distachya herbs. Planta Med (2): 162-163.

34. Black OF, Kelly JW. (1927) Pseudo ephedrine from Ephedra alata. Amer. Jour. Pharm. 99: 748-751.

35. Grue-Sorensen G, Spenser ID. (1989) The biosynthesis of ephedrine. Can J Chem 67: 998-1009.

36. Osadchii SA, Shults EE, Polukhina EV, Shakirov MM, Vasilevskii SF, et al. (2007) Study of alkaloids of the siberian and altai flora 14. synthesis of alkaloid-based tertiary N-(3-arylprop-2-ynyl)amines. Russian Chemical Bulletin 56(6): 1261-1267.

37. Gilg E, Schürholf P. (1930) Die ephedrinhaltigen stammpflanzen der "ma-huang"-droge. Arch Pharm (Weinheim) 233-239.

38. Shabana MM, Mirhom YW, Genenah AA, Aboutabl EA, Amer HA. (1990) Study into wild egyptian plants of potential medicinal activity. ninth communication: Hypoglycaemic activity of some selected plants in normal fasting and alloxanised rats. Arch Exp Veterinarmed 44(3): 389-394.

39. Campbell J, Mortensen A, Molgaard P. (2006) Tissue lipid lowering-effect of a traditional nigerian anti-diabetic infusion of rauwolfia vomitoria foilage and citrus aurantium fruit. J Ethnopharmacol 104(3): 379-386.

40. Benzi G, Villa R, Dossena M, Vercesi L, Gorini A, et al. (1984) Cerebral and cerebellar metabolic changes induced by drugs during the recovery period after profound hypoglycemia. Farmaco Sci 39(1): 44-56.

41. Gosh et al. (1958) Isolation of serpine and ajmaline from the root of rouvolfia Canescens. Naturwissenschaften 45(15): 365.

42. Gorman M, Neuss N, Djerassi C, Kutney JP, Scheuer PJ. (1957) Alkaloid studies–XIX: Alkaloids of some hawaiian *rauwolfia* species: The structure of sandwicine and its interconversion with ajmaline and ajmalidine. Tetrahedron 1(4): 328-337.

43. Chatterjee A, Talapatra S. (1955) Alkaloids of the roots of *rauwolfia densiflora* benth. and hook,*rauwolfia perakensis* King and gamble, *rauwolfia canescens* Linn. and *rauwolfia serpentina* benth. Naturwissenschaften 77: 3551-3553.

44. Hochstein F, Murai K, Boegemann W. (1955) Alkaloids of rauwolfia heterophylla. J Am Chem Soc 77(13): 3551-3554.

45. Roland M. (1959) [The alkaloids of rouwolfia obscura K. schum]. J Pharm Belg 14: 347-364.

46. Sheludko Y, Gerasimenko I, Kolshorn H, Stackigt J. (2002) New alkaloids of the sarpagine group from rauvolfia serpentina hairy root culture. J Nat Prod 65(7): 1006-1010.

47. van de Venter M, Roux S, Bungu LC, Louw J, Crouch NR, et al. (2008) Antidiabetic screening and scoring of 11 plants traditionally used in south africa. J Ethnopharmacol 119(1): 81-86.

48. Pinent M, Blay M, Bladé MC, Salvadó MJ, Arola L, et al. (2004) Grape seed-derived procyanidins have an antihyperglycemic effect in streptozotocin-induced diabetic rats and insulinomimetic activity in insulin-sensitive cell lines. Endocrinology 145(11): 4985-4990.

49. Montagut G, Onnockx S, Vaque M, Blade C, Blay M, et al. (2010) Oligomers of grape-seed procyanidin extract activate the insulin receptor and key targets of the insulin signaling pathway differently from insulin. J Nutr Biochem 21(6): 476-481.

50. Song E, Hur H, Han M. (2003) Epigallocatechin gallate prevents autoimmune diabetes induced by multiple low doses of streptozotocin in mice. Arch Pharm Res 26(7): 559-563.

51. Fu Z, Zhen W, Yuskavage J, Liu D. (2011) Epigallocatechin gallate delays the onset of type 1 diabetes in spontaneous non-obese diabetic mice. Br J Nutr 105(8): 1218-1225.

52. Benn MH, Shustov G, Shustova L, Majak W, Bai Y, et al. (1996) Isolation and characterization of two guanidines from galega orientalis lam. cv. gale (fodder galega). J Agric Food Chem 44(9): 2779-2781.

53. Michel K, Sandberg F, Haglid F, Norin T. (1967) Alkaloids of haloxylon salicornicum (moq.-tand.) boiss. Acta Pharm Suec 4(2): 97-116.

54. Brack A. (1962) Verlauf der Alkaloidbildung durch den Clavicepsstamm von Pennisetum typhoideum Rich. in saprophytischer Kultur. 54. Mitteilung über Mutterkornalkaloide. Arch Pharm Pharm Med Chem 295(7): 510-515.

55. Fujii M, Takei I, Umezawa K. (2009) Antidiabetic effect of orally administered conophylline-containing plant extract on streptozotocin-treated and goto-kakizaki rats. Biomed Pharmacother 63(10): 710-716.

56. Yoshikawa M, Nakamura S, Ozaki K, Kumahara A, Morikawa T, et al. (2007) Structures of steroidal alkaloid oligoglycosides, robeneosides A and B, and antidiabetogenic constituents from the brazilian medicinal plant solanum lycocarpum. J Nat Prod 70(2): 210-214.

57. Villaseñor I, Lamadrid M. (2006) Comparative anti-hyperglycemic potentials of medicinal plants. J Ethnopharmacol 104(1-2): 129-131.

58. Kar DM, Maharana L, Pattnaik S, Dash GK. (2006) Studies on hypoglycaemic activity of solanum xanthocarpum schrad. & wendl. fruit extract in rats. J Ethnopharmacol 108(2): 251-256.

59. Mosihuzzaman M, Nahar N, Ali L, Rokeya B, Khan AK, et al. (1994) Hypoglycemic effects of three plants from eastern himalayan belt. Diabetes Res 26(3): 127-138.

60. Semwal DK, Rawat U, Semwal R, Singh R, Singh GJP. (2010) Anti-hyperglycemic effect of 11-hydroxypalmatine, a palmatine derivative from stephania glabra tubers. J Asian Nat Prod Res 12(2): 99-105.

61. Tsutsumi T, Kobayashi S, Liu YY, Kontani H. (2003) Anti-hyperglycemic effect of fangchinoline isolated from stephania tetrandra radix in streptozotocin-diabetic mice. Biol Pharm Bull 26(3): 313-317.

62. Dixon SL, Smondyrev AM, Knoll EH, Rao SN, Shaw DE, et al. (2006) PHASE: A new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. methodology and preliminary results. J Comput Aided Mol Des 20(10-11): 647-671.

63. Zsoldos Z, Reid D, Simon A, Sadjad SB, Johnson AP. (2007) eHiTS: A new fast, exhaustive flexible ligand docking system. J Mol Graph Model 26(1): 198-212.

64. Liu T, Lin Y, Wen X, Jorissen RN, Gilson MK. (2007) BindingDB: A web-accessible database of experimentally determined protein-ligand binding affinities. Nucleic Acids Res 35(Database issue): D198-201.

65. Duan J, Dixon SL, Lowrie JF, Sherman W. (2010) Analysis and comparison of 2D fingerprints: Insights into database screening performance using eight fingerprint methods. J Mol Graph Model 29(2): 157-170.

66. Kelley LA, Gardner SP, Sutcliffe MJ. (1996) An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies. Protein Eng 9(11): 1063-1065.

67. Vuksan V, Sievenpiper JL. (2005) Herbal remedies in the management of diabetes: Lessons learned from the study of ginseng. Nutr Metab Cardiovasc Dis 15(3): 149-160.

68. Shukla K, Narain JP, Puri P, Gupta A, Bijlani RL, et al. (1991) Glycaemic response to maize, bajra and barley. Indian J Physiol Pharmacol 35(4): 249-254.

69. Ronchetti F, Russo G, Bombardelli E, Bonati A. (1971) A new alkaloid from rauwolfia vomitoria. Phytochemistry 10(6): 1385.

70. Phan MG, Phan TS, Matsunami K, Otsuka H. (2006) Chemical and biological evaluation on scopadulane-type diterpenoids from scoparia dulcis of vietnamese origin. Chem Pharm Bull 54(4): 546-549.

71. Costantino L, Raimondi L, Pirisino R, Brunetti T, Pessotto P, et al. (2003) Isolation and pharmacological activities of the tecoma stans alkaloids. Farmaco 58(9): 781-785.

72. Torres JL, Bobet R. (2001) New flavanol derivatives from grape (vitis vinifera) byproducts. antioxidant Aminoethylthioâˆ'Flavan-3-ol conjugates from a polymeric waste fraction used as a source of flavanols. J Agric Food Chem 49(10): 4627-4634.

73. Takayama H, Okazaki T, Yamaguchi K, Aimi N, Haginiwa J, et al. (1988) Structure of two new diterpene alkaloids, 3-epi-ignavinol and 2,3-dehydrodelcosine. Chem Pharm Bull (Tokyo) 36(8): 3210-3212.

74. Konno C, Murayama M, Sugiyama K, Arai M, Murakami M, et al. (1985) Isolation and hypoglycemic activity of aconitans A, B, C and D, glycans of aconitum carmichaeli roots. Planta Med (2): 160-161.

75. Zhang H, Wang X, Lin L, Ding J, Yue J. (2007) Indole alkaloids from three species of the ervatamia genus: E. officinalis, E. divaricata, and E. divaricata gouyahua. J Nat Prod 70(1): 54-59.

76. Usubillaga A. (1988) Solanudine, a steroidal alkaloid from solanum nudum. Phytochemistry 27(9): 3031.

77. El Sayed KA, Hamann MT, Abd El-Rahman HA, Zaghloul AM. (1998) New pyrrole alkaloids from solanum sodomaeum. J Nat Prod 61(6): 848-850.

78. Kashiwaba N, Morooka S, Ono M, Toda J, Suzuki H, et al. (1997) Alkaloidal constituents of the leaves of stephania cepharantha cultivated in japan: Structure of cephasugine, a new morphinane alkaloid. Chem Pharm Bull (Tokyo) 45(3): 545-548.

79. Beek TAV, Verpoorte R, Svendsen AB. (1984) Alkaloids of tabernaemontana eglandulosa. Tetrahedron 40(4): 737.

## SUMMARIZING DISCUSSION

This doctoral thesis has focused on the identification of new natural compounds as antidiabetic agents. In that sense, the nuclear receptor peroxisome proliferator-activated receptor γ (PPARγ) and the enzyme dipeptidyl peptidase-IV (DPP-IV) have been shown to be appropriate targets for antidiabetic drugs, and therefore, the development of compounds that target these molecules has great potential as a successful therapeutic approach in the treatment of type 2 diabetes mellitus (T2DM). PPARγ regulates the gene expression of proteins that are involved in glucose and lipid metabolism and increases insulin sensitivity. Thiazolidinediones (TZDs) have been approved by the American Food and Drug Administration (FDA) as antidiabetic drugs that operate by activating PPARγ. However, a side effect of TZDs is weight gain, which is caused by increasing adipocyte differentiation and fatty acid storage. Thus, partial agonists of PPARγ with decreased adipose tissue side effects are being investigated for their utility T2DM therapy. On the other hand, DPP-IV removes N-terminal dipeptides from substrates containing proline or alanine at the penultimate position. GLP-1 and GIP are important active substrates of DPP-IV that have a stimulating effect on insulin secretion in a meal-dependent manner. As DPP-IV is responsible for the rapid degradation of GLP-1 and GIP levels in plasma, and the idea of using DPP-IV inhibitors to increase the half-life of these hormones and prolong its beneficial effects has been pursued as a new potential therapeutic approach to the treatment of T2DM. Although PPARγ partial agonists and DPP-IV inhibitors have different modes of action, they share the same purpose of improving glucose homeostasis. Thus, we propose two ways to treat diabetes through activating PPARγ and inhibiting DPP-IV. A co-therapy with PPARγ partial agonists and DPP-IV inhibitors might be more effective in some cases.

PPARγ and DPP-IV have been extensively studied, and there is now a considerable amount of information related to these targets, including a large number of crystal structures. These crystal structures have been useful for applying structure-based approaches and obtaining a more accurate view of how the ligands interact in the binding site of these targets. However, some crucial aspects of these targets, for example, the difference between partial and full PPARγ agonists, remain unsolved.

The application of 3D-QSAR techniques to PPARγ agonists has been useful in predicting the activity of new PPARγ agonists (chapter 1) and discovering the essential features that ideal PPARγ agonists require to have an antidiabetic effect without the side effects associated with PPARγ full agonists (chapter 2). These features include a low transactivation activity with a high binding affinity to inhibit the

phosphorylation of PPARγ at Ser273. Our models (chapter 2) suggest that effective PPARγ partial agonists should have a hydrophobic moiety and an acceptor site with an appropriate conformation to interact with arm II and to establish a hydrogen bond with Ser342. Although interactions with arm I increase the binding affinity, this region should be avoided in order to decrease the transactivation activity of potential PPARγ partial agonists. These interactions primarily consist of the development of a hydrogen bond network with the PPARγ residues Ser289, Tyr473, His323 and His449. This hydrogen bond network is typical of PPARγ full agonists.

Moreover, we have applied a virtual screening technique to identify the most promising candidates from natural product databases in order to focus experimental efforts by eliminating molecules that do not possess the required features. Following protocols similar to those used by pharmaceutical industries in drug discovery, we have successfully developed two virtual screening (VS) workflows to identify new PPARγ partial agonists (chapters 3 and 4) and DPP-IV inhibitors (chapters 5 and 6). Our VS workflows are similar, but in each case, we have adapted them to the specific ligand-binding site of the corresponding target. Some of the techniques that are used in those workflows include ADMET filters, pharmacophore modeling, protein-ligand docking and electrostatic-shape similarity analysis. Our results have demonstrated that the combination of different computational techniques within the same virtual screening workflow generates better results. It is worth noting that once we obtained the hits from our VS workflow, we merged these hits with known PPARγ agonists or DPP-IV inhibitors. This step is important for identifying compounds with novel scaffolds and eliminating trivial PPARγ partial agonists or DPP-IV inhibitors that are very similar to compounds with these activities. In our case, we performed this analysis by generating the fingerprint of an entire set of molecules and then clustering these molecules based on the tanimoto co-efficient. Therefore, our results provide a wide range of starting points for the development of antidiabetics drugs.

Both VS workflows have been validated. The PPARγ workflow was validated using both a in silico analysis (using a set of actives and a set of decoys and then calculating certain metrics, such as enrichment factor, selectivity and specificity) and in vitro analyses that include a PPAR ligand-binding competitive assay, a dual-luciferase reporter assay, adipogenic activity and a glucose uptake assay. The high specificity and the moderate selectivity of our procedure reflect the correct assignment of inactive compounds and the loss of potential PPARγ partial agonists (false negatives), respectively. However, because of the high number of initial compounds and the difficulties in differentiating partial and full PPARγ agonists, we preferred a very specific but less sensible VS workflow. In terms of sensitivity, the partial agonist pharmacophores proved to be the most useful step whereas in terms of specificity and enrichment factor, the most useful step was the electrostatic/shape similarity analysis. Therefore, the combination of different steps seems adequate to obtain a VS workflow

that combines the best elements of each method. Ten compounds that represent ten new chemical scaffolds for PPARγ partial agonists were selected for in vitro biological testing, but two of them were not assayed because of solubility problems. Five of the remaining eight compounds were confirmed as PPARγ partial agonists. These compounds bound to PPARγ, did not stimulate or only moderately stimulated the transactivation activity of PPARγ, did not induce the adipogenesis of preadipocyte cells and stimulated the insulin-induced glucose uptake of adipocytes. These results demonstrate that our VS procedure is able to identify novel scaffolds for PPARγ partial agonists (chapters 3 and 4).

The DPP-IV VS workflow has been validated using an enzymatic assay. Seven out of the nine molecules selected were shown to inhibit DPP-IV. The remaining molecules could not be solubilized and tested. Although the IC50 of the seven hit molecules indicates that their in vitro activity is significantly lower than most known DPP-IV inhibitors, it is important to note that these molecules can be used as lead compounds for developing more potent inhibitors using structural-activity relationship studies. Thus, the results of the present study have demonstrated that our VS protocol is highly successful in the nontrivial identification of DPP-IV inhibitors with no chemical-structure similarities to known activities (chapters 5 and 6).

In our VS workflows, we used two different databases of natural products or derivatives of natural products. The first one is a subset of the ZINC database that contains compounds that are commercially available. This database enabled us to check the reliability of our predictions, because we were able to purchase some of the hit compounds that were generated by our VS workflow and experimentally test their activity. The second database that we used contains molecules that have been reported to be purified from a natural source (independent of the stage of commercialization). The reason why we screened this second database is the potential applicability of our results to the design of new functional foods. In the field of functional foods, the repertoire of molecules to be screened is restricted to those found in nature. Therefore, VS techniques can be used to identify undescribed bioactivities for known natural molecules and, subsequently, increase the number of ingredients that can be used in functional food development. At this point, it is important to note that while the discovery of new bioactive compounds is very important, the identification of the natural source from which these compounds are derived is equally important, because it makes little sense to chemically synthesize a natural product for their use in a functional food. In fact, the use of a natural extract rich in the desired compound or compounds seems to be the ideal solution. Therefore, not only have we been able to identify natural products as antidiabetic agents, but we have also been able to discover natural extracts that contain these antidiabetic agents.

Thus, using the previously developed and validated VS workflows and applying tolerant conditions in the last filter (electrostatic and shape similarity analysis), we have predicted 22 PPARγ partial agonists (chapter 4) and 19 DPP-IV inhibitors (chapter 6) from extracts known to have antidiabetic activity or from related plants with undescribed antidiabetic activity (i.e., they are from the same genus as plants with known antidiabetic properties). Our results provide new hypothesis about the active molecules of natural extracts with antidiabetic properties and the mechanism of action for PPARγ activation or DPP-IV inhibition. Some of these molecules have been isolated from the plants Achyrocline satureoides, Salvia miltiorrhiza, Scutellaria baicalensis, Rauwolfia serpentina and Vitis vinifera, among other natural extracts. We also identified plants with undescribed antidiabetic activity that may contain PPARγ partial agonists and DPP-IV inhibitors that are related to plants with known antidiabetic activity. Some examples of these plants include Aconitum japonicum, Annona purpurea, Helichrysum stenopterum, Solanum nudum and Swertia hookeri. These plants represent a new source of potential antidiabetic extracts. Furthermore, none of the molecules that we predicted show chemical similarity with known PPARγ partial agonists or known DPP-IV inhibitors obtained from the literature or databases. Consequently, these molecules are lead-hopping candidates for the development of new antidiabetic drugs.

In conclusion, this thesis provides a detailed and effective methodology for the identification of natural products that act as antidiabetic agents and a valuable set of natural compounds and extracts that act as PPARγ partial agonists or DPP-IV inhibitors.

# CONCLUSIONS

**1.** Using a structure-based docking strategy for aligning the molecules, highly predictive 3D-QSAR models have been developed for PPARγ full agonists. These 3D-QSAR models may be useful for the prediction of the activities of new PPARγ full agonists and also to derive some structural insights for the improvement of the bioactivities of known PPARγ agonists.

**2.** Effective PPARγ partial agonists should have a hydrophobic moiety and an acceptor site with an appropriate conformation to interact with the arm II of the LBD of PPARγ and to establish a hydrogen bond with Ser342 or an equivalent residue.

**3.** Despite the fact that interactions with the arm I the LBD of PPARγ increase the binding affinity of PPARγ agonists, this region should be avoided in order to decrease the transactivation activity of potential PPARγ partial agonists.

**4.** We have developed and validated a virtual screening (VS) workflow based on two structure-based 3D pharmacophores (one to exclude potential PPARγ full agonists), protein/ligand docking and electrostatic/shape similarity analysis to identify compounds with a high chance of being effective PPARγ partial agonists and that this bioactivity is not trivial because their chemical structure does not resemble known PPARγ partial agonists.

**5.** From an initial set of 89,165 natural products and natural product derivatives, using our VS workflow, 135 compounds were defined as potential PPARγ partial agonists with good ADME properties. Ten compounds that represent ten new chemical scaffolds for PPARγ partial agonists were selected for *in vitro* biological testing, but two of them were not assayed due to solubility problems. Five out of the remaining eight compounds were confirmed as PPARγ partial agonists as they bind to PPARγ, do not or only moderately stimulate the transactivation activity of PPARγ, do not induce adipogenesis of preadipocyte cells and stimulate the insulin-induced glucose uptake of adipocytes.

**6.** The compound ZINC02128851 was the best PPARγ partial agonists that we obtained. This compound may be used for lead-optimization and the subsequent design of more potent and safe antidiabetic drugs.

**7.** We have identified 12 molecules from known antidiabetic natural extracts that could be the bioactive molecules responsible, at least in part, of the antidiabetic activity of the extracts. These molecules are from the plants *Achyrocline satureoides*, *Andrographis paniculata*, *Angelica keiskei*, *Cryptolepis sanguinolenta*, *Harungana madagascariensis*, *Salvia miltiorrhiza* and *Scutellaria baicalensis*; the fungi *Aspergillus terreus* and *Hericium erinaceum*; and the marine species *Dysidea villosa* and *Fucus vesiculosus*.

**8.** We have also identified as potential PPARγ partial agonists 10 molecules from 16 plants with undescribed antidiabetic activity but related, *i.e.*, they are from the same genus, to plants with known andiabetic properties. These plants (like *Annona purpurea*, *Artocarpus gomezianus*, *Helichrysum stenopterum*, *Melicope ptelefolia*, *Murraya paniculata*, *Salvia eriophora*, *Salvia lanigera*, *Salvia prionitis*, *Swertia hookeri*, *Tephrosia watsoniana*) represent a new source of potential antidiabetic compounds.

**9.** We have provided new hypothesis about these 22 active molecules of natural extracts with antidiabetic properties and their mode of action, i.e., the increase of the insulin-stimulated glucose uptake trough the action of PPARγ. These molecules would act as PPARγ partial agonists and as none of these molecules show chemical similarity with 211 synthetic PPARγ partial agonists, they are lead-hopping candidates for the development of new antidiabetic drugs.

**10.** We have developed and validated a VS workflow that is able to enrich novel scaffolds for DPP-IV inhibitors that could be useful for drug development in the area of type 2 diabetes.

**11.** We have predicted that 446 NPs present in ZINC database can act as potential DPP-IV inhibitors. From 9 of the resulting VS hits, 7 showed binding affinity to DPP-IV in a competitive binding assay.

**12.** The compound ZINC02132035 showed the most significant inhibition of DPP-IV together with a dose-response effect. This compound may be used for lead-optimization to design more potent and safe antidiabetic drugs.

**13.** We have predicted as potential DPP-IV inhibitors 13 molecules from known antidiabetic natural extracts that could be the bioactive molecules responsible, at least in part, of the antidiabetic activity of the extracts. These molecules are from the plants *Ephedra alata*, *Ephedra distachya*, *Erythrina variegata*, *Galega orientalis*, *Haloxylon salicornicum*, *Pennisetum typhoideum*, *Rauwolfia serpentina*, *Rauwolfia vomitoria*, *Scoparia dulcis*, *Tecoma*, *Vinca major* and *Vitis vinifera*.

**14.** We have also identified as potential DPP-IV inhibitors 6 molecules from plants with undescribed antidiabetic activity but related, i.e., they are from the same genus, to plants with known andiabetic properties. These plants (like *Aconitum japonicum*, *Ervatamia officinalis*, *Solanum nudum*, *Solanum sodomaeum*, *Stephania cepharantha* and *Tabernaemontana eglandulosa*) represent a new source of potential antidiabetic compounds.

**15.** We have provided new hypothesis about these 18 active molecules of natural extracts with antidiabetic properties and their mode of action, *i.e.*, stimulate the insulin secretion trough the inhibition of DPP-IV. These molecules would act as DPP-IV inhibitors and as none of these molecules show chemical similarity with 2,342 known DPP-IV inhibitors, they are lead-hopping candidates for the development of new antidiabetic drugs.

**16.** Our work also opens the door to the discovery of new natural compounds and/or extracts that can be of use in the design of functional foods focused toward the prevention or treatment of type 2 diabetes mellitus.