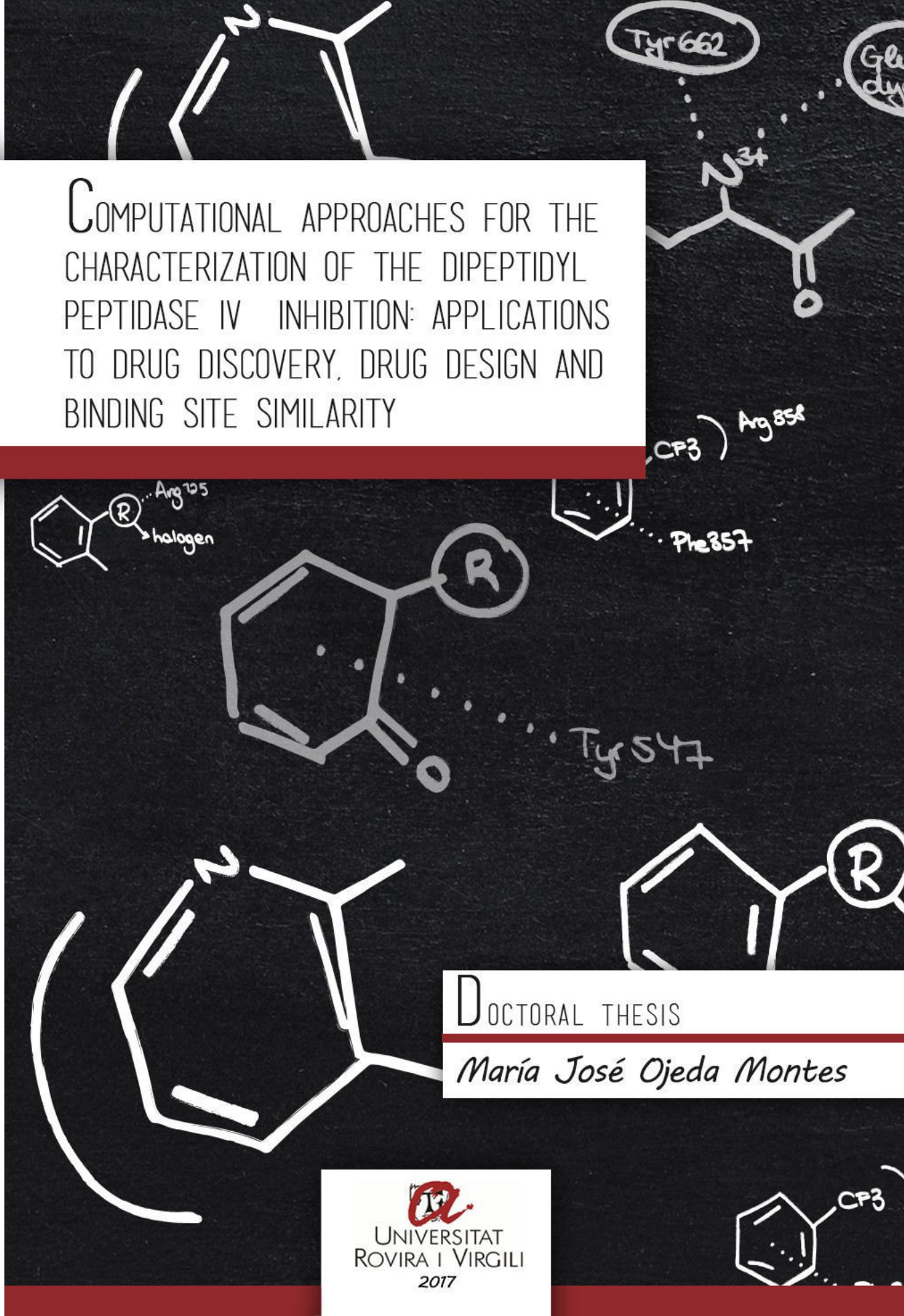
COMPUTATIONAL APPROACHES FOR THE CHARACTERIZATION OF THE DIPEPTIDYL PEPTIDASE IV INHIBITION: APPLICATIONS TO DRUG DISCOVERY, DRUG DESIGN AND BINDING SITE SIMILARITY

María José Ojeda Montes



Computational approaches for the characterization of the dipeptidyl peptidase IV inhibition: applications to drug discovery, drug design and binding site similarity

Doctoral thesis

María José Ojeda Montes

Universitat Rovira i Virgili
2017

María José Ojeda Montes

Computational approaches for the characterization of the Dipeptidyl Peptidase IV inhibition: Applications to drug discovery, drug design and binding site similarity

Doctoral Thesis

Thesis supervised by Dr. Gerard Pujadas Anguiano,
Dr. Santiago Garcia Vallvé and Dra. Cristina Valls Bautista

Cheminformatics & Nutrition Research Group

Biochemistry & Biotechnology Department

UNIVERSITAT ROVIRA i VIRGILI

Tarragona 2017

UNIVERSITAT ROVIRA i VIRGILI

FEM CONSTAR que aquest treball, titulat *"Computational approaches for the characterization of the Dipeptidyl Peptidase IV inhibition: Applications to drug discovery, drug design and binding site similarity"*, que presenta María José Ojeda Montes 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 compleix els requisits per a poder optar a la Menció Internacional de Doctorat.

HACEMOS CONSTAR que el presente trabajo, titulado *"Computational approaches for the characterization of the Dipeptidyl Peptidase IV inhibition: Applications to drug discovery, drug design and binding site similarity"*, que presenta María José Ojeda Montes para la obtención del título de Doctora, ha sido realizado bajo nuestra dirección en el Departamento de Bioquímica y Biotecnología de esta universidad y que cumple los requisitos para poder optar a la Mención Internacional de Doctorado.

WE STATE that the present study, entitled *"Computational approaches for the characterization of the Dipeptidyl Peptidase IV inhibition: Applications to drug discovery, drug design and binding site similarity"*, presented by María José Ojeda Montes for the award of the degree of Doctor, has been carried out under our supervision at the Biochemistry and Biotechnology Department of this university and that this thesis is eligible to apply for the International Doctoral Mention.

Tarragona, 30 de juny del 2017

Els directors de la tesi doctoral

Dr. Gerard Pujadas Anguiano

Dr. Santiago Garcia Vallvé

Dra. Cristina Valls Bautista

A mis padres

Vam fer el que vam saber i és allò que reivindiquem

perquè les dones sempre hem estat en el combat

Neus Català

AGRAÏMENTS

L'assumpte del correu que m'enviava Gerard deia «l'any vinent», emocionada el vaig obrir, llavors no sabia que el doctorat és feina, però sobretot és aprendre, és descobrir i és equivocar-se per tornar a encaminar-ho. Per això, Gerard vull agrair-te especialment haver dipositat la confiança en mi aquell primer dia i tota la resta que han vingut, per la infinita paciència i per contagiar el teu entusiasme per la feina. També agrair al Santi, així com a la resta del grup de Quimioinformàtica i Nutrició per haver-me guiat durant els meus primers passos per aquest laberint científic. També donar les gràcies als grup de recerca de Nutrigenòmica i MoBioFood.

Dr. Gerhard Wolber, I am very grateful for give me the opportunity to work in your group which provided me new bioinformatic perspectives while I was walking around the exotic streets of Berlin.

Agrair als tots els companys de laboratori per posar-li color al doctorat, especialment a l'Aleix per a què continuem programant en equip, a Belén para que sigamos mochileando i sobretot a la Sarah i al Manuel que saben que un bon cafè matiner fa treure els millor dels somriures per a afrontar el dia. Vull agrair a l'Oriol per ensenyar-me tantes formes de lluita i a la meva germana per no deixar de fer-ho.

Y especialmente dar miles de gracias mis padres que aún sin saber muy bien «de qué va eso del doctorado que se te ha alargado tanto tiempo», no han dejado de confiar en mi y permitirme ser la persona que quería.

A tots, moltes gràcies

TABLE OF CONTENTS

SUMMARY

The inhibition of dipeptidyl peptidase-IV (DPP-IV) enzyme has emerged over the last decade as one of the most effective treatments for type II diabetes mellitus with low risk for hypoglycemia and weight gain. Structure-activity relationship analyses and virtual screening protocols have been used to explain how ligands interact with the DPP-IV binding site and to mine large databases of small molecules searching for new DPP-IV inhibitors.

The present doctoral thesis has been therefore focused on: **(a)** the characterization of DPP-IV inhibition in order to suggest how virtual screening protocols may be improved either to favor the identification of potent and selective DPP-IV inhibitors or to look for new lead molecules; **(b)** the design of a computational strategy suitable for identifying new lead compounds with very low (or no) similarity to known actives in purchasable databases; **(c)** the demonstration that, at least partly, the described antidiabetic effect of different *Ephedra* species extracts is the result of the DPP-IV inhibitory bioactivity by ephedrine and the ephedrine-derivatives found in these extracts and **(d)** the analysis of the physico-chemical features shared by the DPP-IV and β_2-adrenergic receptors binding sites and their comparison in order to evaluate if small molecules with dual bioactivity as DPP-IV inhibitors and β-blockers are possible.

It is noteworthy that our work provides a new hypothesis about the cardioprotective effect associated with DPP-IV inhibition and opens the door to a single treatment focused toward type II diabetes mellitus and cardiovascular diseases involved in the metabolic syndrome.

RESUM

La inhibició de l'enzim dipeptidil peptidasa IV (DPP-IV) ha emergit durant les últimes dècades com un dels tractaments més efectius per a la diabetis mellitus tipus II gràcies al seu baix risc hipoglucèmic i al manteniment del pes corporal. Els estudis d'anàlisi de relació estructura-activitat i els protocols de cribratge virtual s'han fet servir per explicar com els lligands interactuen amb el lloc d'unió de la DPP-IV i cercar en extenses bases de dades de compostos de baix pes molecular per tal de trobar nous inhibidors de DPP-IV.

Per tant, la tesi doctoral s'ha centrat en: **(a)** la caracterització de la inhibició de DPP-IV amb l'objectiu de suggerir com els cribratges virtuals podrien ser millorats per a afavorir la identificació d'inhibidors de DPP-IV potents i selectius o bé per cercar noves molècules de partida; **(b)** el disseny d'una estratègia computacional adequada per identificar nous compostos de partida en bases de dades de molècules comercials que presentin baixa (o nul·la) similitud amb els actius existents; **(c)** la demostració que almenys de forma parcial, l'efecte antidiabètic descrit per a extractes de diferents espècies d'*Ephedra* és el resultat de l'activitat inhibitòria de DPP-IV per part dels compostos d'efedrina i derivats d'efedrina trobats en aquests mateixos extractes; i **(d)** l'anàlisi de les característiques fisico-químiques compartides pels llocs d'unió de DPP-IV i del receptor adrenèrgic β_2 i comparar-los amb l'objectiu d'avaluar si és possible que un lligand pugui presentar activitat dual com a inhibidor de DPP-IV i β-bloquejant.

És important destacar que el nostre treball aporta una nova hipòtesi sobre l'efecte cardiosaludable associat a la inhibició de DPP-IV i obre la porta al disseny d'un únic tractament dirigit simultàniament per a la diabetis mellitus tipus II i les malalties cardiovasculars, ambdues involucrades en la síndrome metabòlica.

RESUMEN

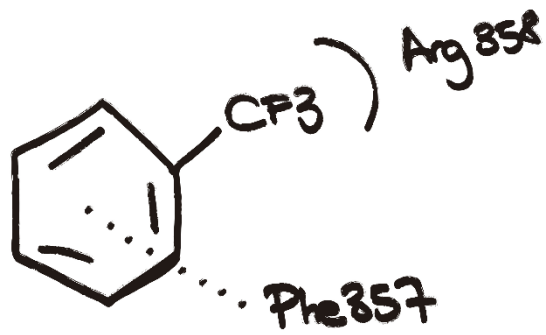
La inhibición de la enzima dipeptidil peptidasa IV (DPP-IV) ha surgido durante las últimas décadas como uno de los tratamientos más efectivos para la diabetes mellitus tipo II gracias a su bajo riesgo hipoglucémico y al mantenimiento del peso corporal. Los estudios de análisis de relación estructura-actividad y los protocolos de cribado virtual se han usado para explicar cómo los ligandos interactúan con el lugar de unión de la DPP-IV y buscar en extensas bases de datos de compuestos de bajo peso molecular para identificar nuevos inhibidores de DPP-IV.

Por lo tanto, la presente tesis doctoral se ha centrado en: **(a)** la caracterización de la inhibición de DPP-IV con el objetivo de sugerir cómo los cribados virtuales podrían mejorarse para favorecer la identificación de inhibidores de DPP-IV potentes y selectivos o bien como buscar nuevas moléculas de partida; **(b)** el diseño de una estrategia computacional adecuada para identificar nuevos compuestos de partida en bases de datos de moléculas comerciales que presenten baja (o nula) similitud con los activos existentes; **(c)** la demostración de que al menos de forma parcial, el efecto antidiabético descrito para los extractos de diferentes especies de *Ephedra* es el resultado de la actividad inhibitoria de DPP-IV por parte de las moléculas de efedrina y derivados de ésta encontrados en estos mismos extractos; y **(d)** el análisis de las características fisico-químicas compartidas por los lugares de unión de DPP-IV y del receptor adrenérgico β_2 y compararlos con el objetivo de evaluar si es posible que un ligando pueda presentar actividad dual como inhibidor de DPP-IV y β-bloqueante.

Es importante destacar que nuestro trabajo aporta una nueva hipótesis sobre el efecto cardiosaludable asociado a la inhibición de DPP-IV y abre la puerta al diseño de un único tratamiento dirigido simultáneamente para la diabetes mellitus tipo II y las enfermedades cardiovasculares, ambas involucradas en el síndrome metabólico.

LIST OF ABBREVIATIONS

ADME/tox	Administration, distribution, metabolism, excretion and toxicity
B_2-AR	B_2 adrenergic receptor
DPP-IV	Dipeptidyl peptidase IV
DPP8	Dipeptidyl peptidase 8
DPP9	Dipeptidyl peptidase 9
EFSA	European Food Safety Authority
FDA	Food and Drug Administration
GDM	Gestational diabetes mellitus
GIP	Gastric inhibitory polypeptide; glucose-dependent insulinotropic polypeptide
GLP-1	Glucagon-like peptide-1
HTS	High throughput screening
IC_{50}	Inhibitory concentration causing 50% reduction in enzyme activity
IDF	International Diabetes Federation
PDB	Protein Data Bank
SAR	Structure-activity relationship
T1DM	Type I diabetes mellitus
T2DM	Type II diabetes mellitus
VS	Virtual screening



Introduction

Diabetes is a chronic disease that occurs when either the pancreas is not able to produce insulin or the body cannot make good use of the insulin it produces. Insulin is a hormone made by β-cells of the pancreas that regulates the metabolism of glucose in blood promoting the absorption into the fat, liver and skeletal muscle cells [1]. Abnormally high glucose levels, known as hyperglycemia, are associated with different type of diabetes (i.e., type I diabetes mellitus [T1DM], type II diabetes mellitus [T2DM] and gestational diabetes [GDM]). Currently, 415 million adults are suffering from diabetes according to data provided by the IDF Diabetes Atlas (2015) and it is expected that this number will rise to 642 million by 2040 in low and middle income countries [1]. Consequently, diabetes becomes one of the greatest healthcare challenges of the 21st century. T2DM accounts for at least 90% of all cases of diabetes and is characterized by insulin resistance (*i.e.*, a gradual decline in insulin sensitivity or secretion) caused mainly by lifestyle behaviors associated with overweight or obesity. Therefore, although oral drugs are required, the treatment is initially managed through exercise and diet.

The high incidence in many countries of the onset T2DM is due to the sedentary lifestyles which has resulted a worldwide epidemic. People with diabetes have an increased risk of developing a number of serious health complications. Consistently, high blood glucose levels can lead to affect cardiovascular diseases, kidney failure and in more severe cases, blindness or limb amputation [1]. During last years, different studies have highlighted the relationship between T2DM and cardiovascular risk factors (such as high blood pressure, coronary artery disease and stroke) [2,3]. Both pathogenesis belong to the group of metabolic abnormalities known as metabolic syndrome which is also associated to other risk factors such as abdominal obesity, atherogenic dyslipidemia and prothrombotic state [3].

The pathophysiology of T2DM produces different molecular abnormalities which include adipocyte insulin resistance, reduced incretin secretion/sensitivity, increased glucagon secretion, enhanced renal glucose reabsorption and brain insulin resistance (see Figure 1) [4,5]. Incretins (*i.e.*, gastric inhibitory polypeptide [GIP] and glucagon-like peptide 1 [GLP-1]) are secreted from enteroendocrine cells in response to oral nutrient ingestion for stimulating glucose-induced insulin secretion from pancreatic β-cells. The insulinotropic action (*i.e.*, incretin effect which refers to the augmentation of insulin secretion after oral administration of glucose compared with intravenous glucose administration at matched glucose levels [6]) of

the two incretin hormones is markedly impaired in diabetic patients as a possible consequence of the pancreatic resistance to the effects of incretins on insulin secretion and/or decreased secretion of GLP-1 (mainly in the postprandial late phase) [7,8]. Thus, while GLP-1 and GIP are held responsible for approximately 50–70% of the postprandial insulin response in healthy individuals, their contribution to the overall insulin response after oral glucose ingestion may amount to less than 20% in patients with T2DM [7]. Unfortunately, both GLP-1 and GIP are rapidly inactivated by dipeptidyl peptidase IV (DPP-IV) which is a ubiquitous serine protease that cleaves the first two residues at the N-terminal side of the active hormones rendering them inactive [9] (see Figure 2). Therefore, two approaches have been successfully employed to enhance the incretin effect in patients suffering T2DM: **(a)** GLP-1 receptor agonists and **(b)** DPP-IV inhibitors. Moreover, clinical trials have indicated that incretin-based therapies may exert cardiovascular benefits such as blood pressure and lipids [10–12]. GLP-1 administration demonstrated to increase nitric oxide and cardiac glucose uptake in models of cardiac ischemia and to reduce infarct size and to improve left ventricular function in models of ischemia/reperfusion injury [13].

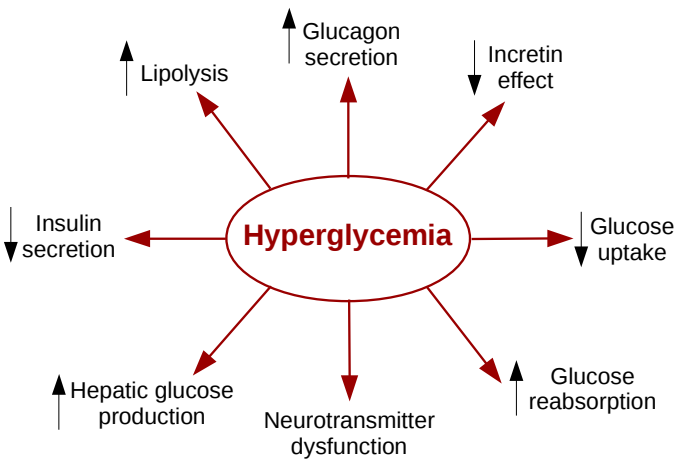


Figure 1. Multiple pathophysiological abnormalities contribute to impaired glucose homeostasis in T2DM. Adapted from [4].

The enzyme DPP-IV (*i.e.*, also known as CD26) is a 766-amino acid membrane glycoprotein with a large extracellular domain, a transmembranous domain and a short cytoplasmic domain [13]. DPP-IV is widely distributed in multiple tissues throughout the body including

the small intestines, kidney, hepatocytes, and endothelial cells and T-cells. Therefore, DPP-IV has catalytic functions as an N-terminal dipeptidase and cleaves a diverse range of proteins (*e.g.*, chemokines, neuropeptides or regulatory peptides) and also non-catalytic functions involved in immune regulation [13]. However, the inhibition of this enzyme has a potential role on the T2DM treatment, heart failure, renal impairment, anti-inflammatory effect and even on Alzheimer disease [14–16].

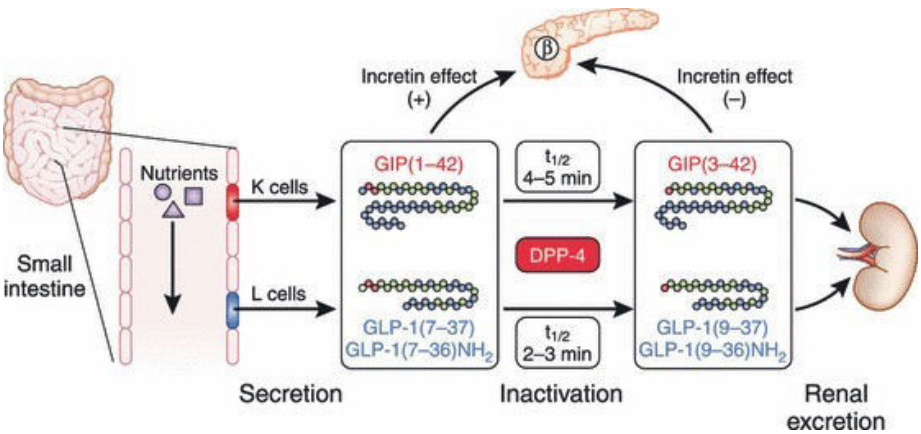


Figure 2. Secretion and metabolism of glucose-dependent insulinotropic polypepide (GIP) and glucagon-like peptide (GLP-1). GIP is secreted from K cells of the upper intestine; GLP-1 is secreted from L cells of the lower intestine. Released GIP and GLP-1 rapidly undergoes proteolytic processing by dipeptidyl peptidase-IV (DPP-IV), and is thereby inactivated and excreted from the kidney. The intact incretins, GIP(1–42), GLP-1(7–37), and GLP-1(7–36)amide, have insulinotropic effects on pancreatic β-cells, whereas the DPP-IV processed incretins, GIP(3-42), GLP-1(9–37), and GLP-1(9–36)amide, have lost their insulinotropic effects. Extracted from [9].

Consequently, the present PhD thesis is focused on using virtual screening (VS) for the prediction of lead compounds which modulate the activity of the DPP-IV target, being this strategy an excellent safe approach to improve glycemic control. VS is a valuable means of focusing experimental efforts on filtered sets of compounds yielding a higher probability of having the desired biological activity. Hit identification using computational screening requires several interactive and iterative steps and requires a careful selection of the methods to be used. VS methods can be roughly organized into two major groups, namely, ligand- and structure-based (*e.g.*, pharmacophore-based methods and protein-ligand docking) [17]. Indeed, in the present thesis both methods have been combined to increase the possibility of identifying active compounds. Novel lead compounds for DPP-IV inhibition

beyond existing drugs (*e.g.*, eleven gliptins are on the market since 2006) might give a remarkably helpful profile regarding safety and viability. A few review articles have revealed the status of the improvements introduced in the development of advanced inhibitors for DPP-IV [18–21]. Recently, computational aided drug discovery approaches have been used as one of the most reliable methods for the development of new DPP-IV inhibitors. Indeed by the use of such computational approaches, many lead compounds were identified and some of these are now under clinical studies [22–32]. Because inhibition of DPP8 and DPP9 (both enzymes belong to the same family as DPP-IV) has been shown to cause severe toxicity in preclinical species [33], high selectivity is therefore an important criterion in selecting DPP-IV inhibitors for antidiabetic clinical development.

We have therefore characterized the binding site of DPP-IV in order to suggest how VS protocols may be improved either to favor the identification of potent and selective DPP-IV inhibitors or to look for new lead molecules ***(Manuscript 1)***. Then, those simple rules were used during the design of a computational strategy suitable for identifying new lead compounds with very low (or no) similarity to existing actives in databases of purchasable compounds ***(Manuscript 2)***. A new study was designed to demonstrate that, at least partly, the described antidiabetic effect of different *Ephedra* species extracts is the result of the DPP-IV inhibitory bioactivity by ephedrine and the ephedrine-derivatives found in these extracts ***(Manuscripts 3)***. Finally, the last section of this thesis is focused on the described cardioprotective effect of the DPP-IV inhibition [34,35]. A new hypothesis has been considered based on the analysis of the physico-chemical features shared by the DPP-IV and β_2 adrenergic receptors (β_2-AR) binding sites and compare them with the aim to evaluate if small molecules with dual bioactivity as DPP-IV inhibitors and β-blockers are possible ***(Manuscripts 4)***.Thus, this relationship awakes great interest in medicinal chemistry to study new alternatives for a a single treatment for hyperglycemia, arrhythmias and hypertension, all of them involved in metabolic syndrome.

[1] International Diabetes Federation, IDF. Diabetes atlas. 7th edition, 2015.

[2] J.A. Sattigeri, S. Sethi, J.A. Davis, S. Ahmed, G. V. Rayasam, B.G. Jadhav, S.M. Chilla, D. Datta, A. Gadhave, V.K. Tulasi, T. Jain, S. Voleti, B. Benjamin, S. Udupa, G. Jain, Y. Singh, K. Srinivas, V.S. Bansal, A. Ray, P.K. Bhatnagar, I.A. Cliffe, Approaches towards the development of chimeric DPP4/ACE inhibitors for treating metabolic syndrome, Bioorg. Med. Chem. Lett. 27 (2017) 2313–2318. doi:10.1016/j.bmcl.2017.04.036.

[3] J. Kaur, A comprehensive review on metabolic syndrome, Cardiol. Res. Pract. 2014 (2014) 943162. doi:10.1155/2014/943162.

[4] R.A. DeFronzo, C.L. Triplitt, M. Abdul-Ghani, E. Cersosimo, Novel agents for the treatment of type 2 diabetes, Diabetes Spectr. 27 (2014) 100–12. doi:10.2337/diaspect.27.2.100.

[5] R.A. DeFronzo, E. Ferrannini, L. Groop, R.R. Henry, W.H. Herman, J.J. Holst, F.B. Hu, C.R. Kahn, I. Raz, G.I. Shulman, D.C. Simonson, M.A. Testa, R. Weiss, Type 2 diabetes mellitus, Nat. Rev. Dis. Prim. 1 (2015). doi:10.1038/nrdp.2015.19.

[6] B. Ahrén, Incretin dysfunction in type 2 diabetes: Clinical impact and future perspectives, Diabetes Metab. 39 (2013) 195–201. doi:10.1016/j.diabet.2013.03.001.

[7] J.J. Meier, M.A. Nauck, Is the diminished incretin effect in type 2 diabetes just an epi-phenomenon of impaired β-cell function?, Diabetes. 59 (2010) 1117–25. doi:10.2337/db09-1899.

[8] J.J. Holst, F.K. Knop, T. Vilsbøll, T. Krarup, S. Madsbad, Loss of incretin effect is a specific, important, and early characteristic of type 2 diabetes, Diabetes Care. 34 Suppl 2 (2011) S251–7. doi:10.2337/dc11-s227.

[9] Y. Seino, M. Fukushima, D. Yabe, GIP and GLP-1, the two incretin hormones: Similarities and differences, J. Diabetes Investig. 1 (2010) 8–23. doi:10.1111/j.2040-1124.2010.00022.x.

[10] M. Monami, I. Dicembrini, E. Mannucci, Dipeptidyl peptidase-4 inhibitors and pancreatitis risk: a meta-analysis of randomized clinical trials, Diabetes. Obes. Metab. 16 (2014) 48–56. doi:10.1111/dom.12176.

[11] H.R. Patil, F.J. Al Badarin, H.A. Al Shami, S.K. Bhatti, C.J. Lavie, D.S.H. Bell, J.H. O'Keefe, Meta-analysis of effect of dipeptidyl peptidase-4 inhibitors on cardiovascular risk in type 2 diabetes mellitus, Am. J. Cardiol. 110 (2012) 826–33. doi:10.1016/j.amjcard.2012.04.061.

[12] P.F. Mora, E.L. Johnson, Cardiovascular outcome trials of the incretin-based therapies: What do we know so far?, Endocr. Pract. 23 (2017) 89–99. doi:10.4158/EP161481.RA.

[13] A. Cahn, S. Cernea, I. Raz, An update on DPP-4 inhibitors in the management of type 2 diabetes, Expert Opin. Emerg. Drugs. 21 (2016) 409–419. doi:10.1080/14728214.2016.1257608.

[14] Y. Zhao, L. Yang, Z. Zhou, Dipeptidyl peptidase-4 inhibitors: multitarget drugs, not only antidiabetes drugs, J. Diabetes. 6 (2014) 21–9. doi:10.1111/1753-0407.12063.

[15] M. Haluzík, J. Frolík, I. Rychlík, Renal effects of DPP-4 inhibitors: A focus on microalbuminuria, Int. J. Endocrinol. 2013 (2013) 895102. doi:10.1155/2013/895102.

[16] J. Kosaraju, V. Murthy, R.B. Khatwal, A. Dubala, S. Chinni, S.K. Muthureddy Nataraj, D. Basavan, Vildagliptin: an anti-diabetes agent ameliorates cognitive deficits and pathology observed in streptozotocin-induced Alzheimer's disease, J. Pharm. Pharmacol. 65 (2013) 1773–84. doi:10.1111/jphp.12148.

[17] D. Schuster, G. Wolber, Identification of bioactive natural products by pharmacophore-based virtual screening, Curr. Pharm. Des. 16 (2010) 1666–81.

[18] L. Juillerat-Jeanneret, Dipeptidyl peptidase IV and its inhibitors: Therapeutics for type 2 diabetes and what else?, J. Med. Chem. 57 (2014) 2197–212. doi:10.1021/jm400658e.

[19] Y. Liu, Y. Hu, T. Liu, Recent advances in non-peptidomimetic dipeptidyl peptidase 4 inhibitors: Medicinal chemistry and preclinical aspects, Curr. Med. Chem. 19 (2012) 3982–99.

[20] A. Smelcerovic, F. Miljkovic, A. Kolarevic, J. Lazarevic, A. Djordjevic, G. Kocic, M. Anderluh, An overview of recent dipeptidyl peptidase-IV inhibitors: linking their structure and physico-chemical properties with SAR, pharmacokinetics and toxicity, Curr. Top. Med. Chem. 15 (2015) 2342–72.

[21] B.D. Patel, M.D. Ghate, Recent approaches to medicinal chemistry and therapeutic potential of dipeptidyl peptidase-4 (DPP-4) inhibitors, Eur. J. Med. Chem. 74 (2014) 574–605. doi:10.1016/j.ejmech.2013.12.038.

[22] A.S. Akarte, B.P. Srinivasan, S. Gandhi, A novel long acting DPP-IV inhibitor PKF-275-055 stimulates β-cell proliferation resulting in improved glucose homeostasis in diabetic rats,

Biochem. Pharmacol. 83 (2012) 241–52. doi:10.1016/j.bcp.2011.10.003.

[23] Y. Miyamoto, Y. Banno, T. Yamashita, T. Fujimoto, S. Oi, Y. Moritoh, T. Asakawa, O. Kataoka, H. Yashiro, K. Takeuchi, N. Suzuki, K. Ikedo, T. Kosaka, S. Tsubotani, A. Tani, M. Sasaki, M. Funami, M. Amano, Y. Yamamoto, K. Aertgeerts, J. Yano, H. Maezaki, Discovery of a 3-pyridylacetic acid derivative (TAK-100) as a potent, selective and orally active dipeptidyl peptidase IV (DPP-4) inhibitor, J. Med. Chem. 54 (2011) 831–50. doi:10.1021/jm101236h.

[24] C. Li, W. Lu, C. Lu, W. Xiao, X. Shen, J. Huang, G. Liu, Y. Tang, Identification of diverse dipeptidyl peptidase IV inhibitors via structure-based virtual screening, J. Mol. Model. 18 (2012) 4033–42. doi:10.1007/s00894-012-1394-3.

[25] R.A. Ward, T.D.J. Perkins, J. Stafford, Structure-based virtual screening for low molecular weight chemical starting points for dipeptidyl peptidase IV inhibitors, J. Med. Chem. 48 (2005) 6991–6. doi:10.1021/jm0505866.

[26] C. Rummey, S. Nordhoff, M. Thiemann, G. Metz, In silico fragment-based discovery of DPP-IV S1 pocket binders, Bioorg. Med. Chem. Lett. 16 (2006) 1405–9. doi:10.1016/j.bmcl.2005.11.038.

[27] I.M. Al-Masri, M.K. Mohammad, M.O. Taha, Discovery of DPP IV inhibitors by pharmacophore modeling and QSAR analysis followed by in silico screening, ChemMedChem. 3 (2008) 1763–79. doi:10.1002/cmdc.200800213.

[28] S. Zhang, W. Lu, X. Liu, Y. Diao, F. Bai, L. Wang, L. Shan, J. Huang, H. Li, W. Zhang, Fast and effective identification of the bioactive compounds and their targets from medicinal plants via computational chemical biology approach, Medchemcomm. 2 (2011) 471. doi:10.1039/c0md00245c.

[29] L. Guasch, M.J. Ojeda, N. González-Abuín, E. Sala, A. Cereto-Massagué, M. Mulero, C. Valls, M. Pinent, A. Ardévol, S. Garcia-Vallvé, G. Pujadas, Identification of novel human dipeptidyl peptidase-IV inhibitors of natural origin (part I): Virtual screening and activity assays, PLoS One. 7 (2012) e44971. doi:10.1371/journal.pone.0044971.

[30] I.M. Almasri, M.O. Taha, M.K. Mohammad, New leads for DPP IV inhibition: Structure-based pharmacophore mapping and virtual screening study, Arch. Pharm. Res. 36 (2013) 1326–37. doi:10.1007/s12272-013-0224-1.

[31] J. Xing, Q. Li, S. Zhang, H. Liu, L. Zhao, H. Cheng, Y. Zhang, J. Zhou, H. Zhang, Identification

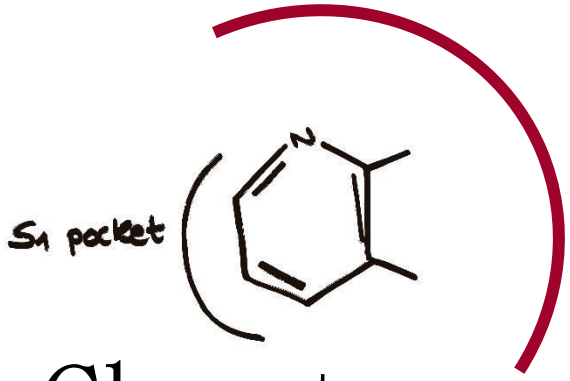
of dipeptidyl peptidase IV inhibitors: Virtual screening, synthesis and biological evaluation, Chem. Biol. Drug Des. 84 (2014) 364–77. doi:10.1111/cbdd.12327.

[32] O. Tanwar, L. Tanwar, M. Shaquiquzzaman, M.M. Alam, M. Akhter, Structure based virtual screening of MDPI database: Discovery of structurally diverse and novel DPP-IV inhibitors, Bioorganic Med. Chem. Lett. 24 (2014) 3447–3451. doi:10.1016/j.bmcl.2014.05.076.

[33] G.R. Lankas, B. Leiting, R.S. Roy, G.J. Eiermann, M.G. Beconi, T. Biftu, C.-C. Chan, S. Edmondson, W.P. Feeney, H. He, D.E. Ippolito, D. Kim, K.A. Lyons, H.O. Ok, R.A. Patel, A.N. Petrov, K.A. Pryor, X. Qian, L. Reigle, A. Woods, J.K. Wu, D. Zaller, X. Zhang, L. Zhu, A.E. Weber, N.A. Thornberry, Dipeptidyl peptidase IV inhibition for the treatment of type 2 diabetes: potential importance of selectivity over dipeptidyl peptidases 8 and 9, Diabetes. 54 (2005) 2988–94.

[34] E. Bonora, M. Cigolini, DPP-4 inhibitors and cardiovascular disease in type 2 diabetes mellitus. Expectations, observations and perspectives, Nutr. Metab. Cardiovasc. Dis. 26 (2016) 273–284. doi:10.1016/j.numecd.2016.03.002.

[35] M.H. Davidson, Potential impact of dipeptidyl peptidase-4 inhibitors on cardiovascular pathophysiology in type 2 diabetes mellitus, Postgrad. Med. 126 (2014) 56–65. doi:10.3810/pgm.2014.05.2756.



Book Chapter

Book Chapter

DPP-IV, an important target for antidiabetic functional food design

María José Ojeda-Montes, Adrià Cereto-Massagué,
Cristina Valls and Gerard Pujadas

Foodinformatics
Applications of chemical information to food chemistry

Karina Martinez-Mayorga and José Luis Medina-Franco
Springer, 2014
DOI: 10.1007/978-3-319-10226-9

Chapter 7
DPP-IV, An Important Target for Antidiabetic Functional Food Design

María José Ojeda, Adrià Cereto-Massagué, Cristina Valls and Gerard Pujadas

7.1 Introduction

7.1.1 Type 2 Diabetes Mellitus

Diabetes is a chronic disease that occurs when the pancreas does not produce sufficient insulin. Diabetes may also arise when the body cannot effectively use the insulin it produces. Hyperglycemia, or increased blood sugar, is a common effect of uncontrolled diabetes. Chronic hyperglycemia leads to serious damage to many body systems, particularly the nerves and blood vessels.

Type 2 diabetes mellitus—formerly referred to as noninsulin-dependent diabetes mellitus (T2DM)—is a chronic metabolic disease that is characterized by hyperglycemia and results from the body's ineffective use of insulin (i.e., a gradual decline in insulin sensitivity and/or insulin secretion). T2DM accounts for 90 % of people with diabetes and has become a worldwide epidemic. Moreover, many countries are now reporting the onset of T2DM at an increasingly young age due to sedentary lifestyles, longer life expectancies, and obesity [1].

G. Pujadas (✉) · M. J. Ojeda · A. Cereto-Massagué · C. Valls
Research Group in Chemoinformatics & Nutrition,
Departament de Bioquímica i Biotecnologia,
Universitat Rovira i Virgili, Campus de Sescelades,
C/ Marcel·lí Domingo s/n, 43007 Tarragona, Catalonia, Spain
e-mail: gerard.pujadas@urv.cat

G. Pujadas
Centre Tecnològic de Nutrició i Salut (CTNS),
TECNIO, CEICS,
Avinguda Universitat 1, 43204 Reus,
Catalonia, Spain

K. Martinez-Mayorga, J. L. Medina-Franco (eds.), *Foodinformatics*,
DOI 10.1007/978-3-319-10226-9_7

The majority of patients with T2DM are obese [2], and many of the current therapeutic options for management of T2DM can cause further weight gain [3, 4]. Concerns about weight gain adversely affect patients' willingness to begin and continue treatment with glucose-lowering medications, such as thiazolidinediones, insulin, and sulfonylureas [5]. In addition to weight gain, a patient's quality of life can be negatively affected by the underlying disease process and its complications, such as polypharmacy, hypoglycemia and micro- and macro-vascular complications [6].

The World Health Organization (WHO) and the International Diabetes Federation (IDF) report that between 347 and 371 million people worldwide currently have diabetes. It is forecasted that the number of diabetes deaths will double between 2005 and 2030, which will make diabetes the seventh leading cause of death in 2030 [7, 8]. According to the WHO and IDF information, this strong correlation between diabetes and death are supported by the following data: (a) between 50 and 80 % of people with diabetes die of cardiovascular disease (primarily heart disease and stroke) [9], (b) diabetes is among the leading causes of kidney failure [10], (c) the overall risk of dying among people with diabetes is at least double the risk of their peers without diabetes [11], and (d) half of all people who die from diabetes are under the age of 60. Moreover, the WHO data also reveal the following: (a) the combination of diabetes with reduced blood flow and neuropathy increases the chance of foot ulcers, infection, and eventual need for limb amputation, and (b) 1 % of global blindness can be attributed to diabetes because it is the result of long-term accumulated damage to the retina's small blood vessels [12].

7.1.2 Current T2DM Incidence in North America and the Caribbean Region

According to the last Diabetes Atlas Update from the IDF [1], approximately 9.6 % of the population between 20 and 79 years old in the North American and Caribbean region (corresponding to 36.8 million people; 24.4 million in the USA) is estimated to be affected by diabetes. By 2035, the number of affected people is expected to increase to 50.4 million. Moreover, 44.2 million people (13.2 % of adults in this region) have impaired glucose tolerance (58.8 million expected by 2035), which increases their risk for developing T2DM. Diabetes-related causes were responsible for 13.5 % (150,000 men and 143,000 women) of all deaths among adults in this region during 2013. In the USA, more than 192,000 people died from diabetes in 2013, which is one of the highest numbers of deaths due to diabetes of any country in the world. The USA is estimated to account for almost half (42 %) of the world's diabetes-related health-care spending.

7.1.3 Pharmacological Treatment of T2DM

There are now ten different drug classes available as adjuncts to diet and exercise for the management of hyperglycemia in T2DM patients in the USA (e.g., sulfonylureas, biguanides, meglitinides, α-glucosidase inhibitors, thiazolidinediones, glucagon-like peptide 1 (GLP-1) agonists, DPP-IV inhibitors, amylin analogs, bile acid sequestrants, and dopamine receptor agonists; Table 7.1) [13]. Despite the many available drugs, there is still a need for new therapies to control glycemia [14]. Many compounds can reduce blood glucose levels. However, clinical use requires an effective antihyperglycemic agent that can meet requirements beyond simply reducing the blood glucose levels [15]. For example, safety profiles (particularly cardiovascular safety) have received significant attention over the past few years.

7.1.4 DPP-IV Inhibition in T2DM Treatment

DPP-IV (also known as adenosine deaminase-binding protein or CD26; EC 3.4.14.5) is a ubiquitous aminodipeptidase that was first described by Hopsu-Havu and Sarimo [16]. It belongs to the α/β-hydrolases (family S9B) and is related to the prolyl oligopeptidase [17]. DPP-IV is expressed on the surface of several cell types including lymphocytes and monocytes and in tissues in the pancreas, kidneys, liver, and the gastrointestinal tract [18]. There are different expression levels in different tissue types. Its expression is particularly high in the kidney cortex, the small intestine brush-border membranes, and the epithelial cells of pancreatic ducts [19]. The widespread expression of DPP-IV means that it can easily access and inactivate a wide variety of biological regulatory peptides. The target peptides include glucose-dependent insulinotropic polypeptide (GIP), GLP-1, growth hormone, peptide YY, and neuropeptide Y [20].

The structure of DPP-IV is a homodimeric transmembrane glycoprotein. Each subunit of the protein is anchored to the plasma membrane by a hydrophobic helix consisting of seven N-terminal amino acids. Each subunit has a large globular extracellular region that contains an active site located in the interface between the β-propeller domain (from residues 39 to 508) and the α/β-hydrolase domain (from residues 509–766; Fig. 7.1) [21–24]. The cleavage of the extracellular portion of DPP-IV from the transmembrane section results in a soluble circulating form of approximately 100 kDa. The soluble form is found in plasma and cerebrospinal fluid [18, 25]. DPP-IV is secreted as a mature monomer, but it requires dimerization to undergo normal proteolytic activity [26].

Recent studies indicated that in addition to the regulation of postprandial glycemia, DPP-IV may have pleiotropic effects (e.g., obesity, tumor growth, and HIV infection), which makes it an attractive target for drug discovery research [27–32]. DPP-IV inhibitors block the degradation of GLP-1 and inhibit the inactivation of several other peptides that may have vasoactive and cardioprotective effects

Table 7.1 The ten different drug classes currently available in the USA that serve as adjuncts to diet and exercise in the management of hyperglycemia in T2DM patients

Antidiabetic agents	Examples	Mode of action	Advantages	Adverse effects
Sulfonylureas	Glipide, glyburide, glimepiride	Induction insulin release from β cells by inhibiting potassium flux through ATP-dependent potassium channels (K_{ATP})	Reduced hepatic uptake, inhibition of glucagon and enhanced insulin sensitivity	Hypoglycemia, body weight gain and possible affection to pancreatic function
Biguanides	Metformin	Suppression of hepatic gluconeogenesis by AMPK phosphorylation	Low rates of hypoglycemia, weight stability/loss, better insulin sensitivity	Gastrointestinal side effects and possible affection to renal or hepatic function
Meglitinides	Repaglinide, nateglinide	Interaction with the voltage-dependent K_{ATP} chanels of pancreatic β cells	Induction of an early insulin response to meals decreasing postprandial blood glucose levels, low rates of hypoglycemia	Weight gain and increased on the insulin deficiency
α-glucosidase inhibitors	Acarbose, miglitol	Competitive inhibition of the α-glucosidase in the intestine	No drug-drug interaction, weight loss, no risk of hypoglycemia, cardioprotective effects, stimulated secretion of GLP-1	Gastrointestinal effects: flatulence, diarrhea, abdominal discomfort
Thiazolidinediones or PPAR-γ agonists	Rosiglitazone, pioglitazone	Binding on the PPAR-γ, it activates the transcription of specific genes of lipid metabolism	Sensitivity to insulin, anti-inflammatory effects and amelioration of hypertension, microalbuminuria and hepatic steatosis	Severe liver failure, death and increased cardiac risk
GLP-1 agonists or mimetics	Exenatide, liraglutide	They are modified GLP-1 molecules that are resistant to DPP-IV induced degradation	Stimulate insulin secretion and inhibit glucagon output in a glucose-dependent manner, slow gastric emptying and decrease appetite	Increased risk of pancreatitis, pre-cancerous cellular changes called pancreatic duct metaplasia and of tumor development at the thyroid gland
DPP-IV inhibitors	Sitagliptin, Saxagliptin	Increase circulating GLP-1 and GIP levels prolonging their action (which lead to decreased levels of blood glucose, HbA1c and glucagon)	Better glucose homeostasis with a lower risk of hypoglycemia and without adversely affecting cardiovascular markers	Headache, nausea, vomiting, loss of appetite

Table 7.1 (continued)

Antidiabetic agents	Examples	Mode of action	Advantages	Adverse effects
Amylin analogues	Pramlintide	Amylin binds to calcitonin receptors in the central nervous system that cooperate with receptor activity modifying proteins	Enhanced satiety, diminished glucagon secretion and delayed gastric emptying	Severe hypoglycemia, nausea, vomiting, anorexia and headache
Bile acid sequestrants	Colesevelam	Binding to the nuclear farnesoid X receptor or the membrane receptor TGR5, where it regulates lipids and glucose levels	No toxicity, no dependency of liver and kidney function	Abdominal and muscle pain, nausea, diarrhea and constipating effects. Associated with dysphagia and esophageal obstruction
Dopamine receptor agonists	Bromocriptine	Activation of hypothalamic-pituitary-adrenal axis	No effects on free fatty acids levels or hepatic glucose production	Nausea, vomiting, diarrhea, stomach cramps and depression

[33–42]. Therefore, the growing body of evidence suggests that DPP-IV inhibitors improve several cardiovascular risk factors, including (a) improvement of endothelium-dependent relaxation, (b) reduction of the vascular inflammation and oxidative stress, (c) reduction of total cholesterol levels, (d) prevention of vascular endothelial dysfunction and atherosclerosis, and (e) reduction of myocardial fibrosis and oxidative stress [42]. Major prospective clinical trials involving various DPP-IV inhibitors with predefined cardiovascular outcomes are currently in progress. These studies are examining T2DM patients who have a high-risk cardiovascular profile to confirm this cardiovascular protective effect [40].

7.1.5 Importance of Selectivity in DPP-IV Inhibition

DPP-IV is in a family of ubiquitous atypical serine proteases with numerous functions, including roles in nutrition, metabolism, the endocrine and immune systems, cancer growth, bone marrow mobilization, and cell adhesion [20]. The DPP-IV family includes four enzymes (DPP-IV, fibroblast activation protein (FAP),DPP8, and DPP9) and two nonenzymes (DPP-IV-like protein-6; DPP6, DPL-1, or DPP-X; and DPP10; DPL-2) [20].

The enzyme FAP, also known as seprase, is the most similar family member to DPP-IV. FAP and DPP-IV share 52% amino acid identity (human enzymes) and similar substrate specificity. Despite these similarities, FAP and DPP-IV differ in their expression patterns because FAP expression is confined predominantly to

Fig. 7.1 A general overview of the 3D fold of the extracellular region for one of the subunits in the human DPP-IV homodimer. The β-propeller domain is shown in *yellow* whereas the α/β-hydrolase domain is shown in *green*. The location of the active site is indicated by the *red* residues from the catalytic triad (Ser630, Asp708 and His740) and the fluoroolefin inhibitor (*in cyan*). This figure has been built with the PDB structure with 3C45 code [92] and with the molecular visualization software RasMol [208]

activated fibroblasts in diseased tissue (e.g., fibrotic and epithelial tumors, invasive cancers [43], and some fetal mesenchymal tissues), but it is absent in the adult human tissues. The other two catalytically active DPP-IV family members, DPP8 and DPP9, share 26 and 21% amino acid identity with the protein sequence of DPP-IV and FAP, respectively (human enzymes). DPP8 and DPP9 are soluble monomeric proteins in the cytoplasm and are very similar proteins because they share 61% amino acid sequence similarity. DPP8 expression is upregulated in activated T cells, and high levels of DPP9 are found in cancer cells, normal skeletal muscle, and the heart and liver [44]. However, their physiological function is not known. Compounds that were previously thought to be specific for DPP-IV could also be inhibitors of other members of the DPP-IV family.

A number of DPP-IV inhibitors have recently been tested for selectivity to DPP-IV, FAP, DPP8, and DPP9 enzymes [45]. In that study, individually selective compounds for DPP-IV, DPP8/9, and FAP were identified, which allowed an evaluation of the potential toxicity and tolerability of each type of inhibition. The DPP8/9

selective inhibitor produced alopecia, thrombocytopenia, reticulocytopenia, multiorgan histopathological changes, enlarged spleen, and mortality in rats. In dogs, the DPP8/9 inhibitor produced gastrointestinal toxicity. However, investigation of the DPP-IV selective inhibitor demonstrated no apparent toxicity [45]. Because inhibition of DPP8 and/or DPP9 has been shown to cause severe toxicity in preclinical species [45], high selectivity is an important criterion in selecting DPP-IV inhibitors for antidiabetic clinical development. Thus, new DPP-IV inhibitors reported on the literature are selective relative to other members of the DPP-IV family [86–105].

7.2 The Incretin System

7.2.1 Overview

Incretin hormones are gut peptides secreted by endocrine cells in the intestinal mucosa in response to nutrient ingestion. These peptides play a key role in the regulation of islet function and blood glucose levels (Fig. 7.2). In humans, the major incretin hormones are GLP-1 and GIP, and, together, they fully account for the incretin effect [46]. The incretin effect is defined as the phenomenon whereby orally ingested glucose elicits a much greater insulin response compared with the response obtained when glucose is infused intravenously to give identical blood glucose levels (the so-called isoglycemic glucose infusion) [47–49]. It has been demonstrated that the incretin effect is responsible for 50–70 % of insulin response in healthy humans [48, 50, 51].

The incretin hormones are released following meal ingestion and are rapidly degraded by DPP-IV [46, 48, 52]. GLP-1 is produced by L cells located in the ileum and in the colon where they are found in high density [49]. In contrast, GIP is secreted by K cells, which are primarily located in the duodenum. Both L cells and K cells are situated in the intestinal mucosa. As a result, these cells can be influenced by direct contact with nutrients from food ingestion [49, 53]. The secretion of GLP-1 and GIP depends not only on the type of macronutrients but also on the rate of gastric emptying and intestinal transit time. Moreover, some evidences show that secretion is modulated by the circadian system, and that higher secretion occurs in the morning than in the afternoon [54, 55]. The incretin metabolites are primarily cleared by the kidneys.

7.2.2 Incretins and Glucose Homeostasis

Both GLP-1 and GIP are able to regulate glucose homeostasis by interacting with G-protein-coupled receptors (GPCR) [56, 57]. The GIP receptor is mainly expressed on islet β cells, but it also occurs in adipose tissue and in the central nervous system. Conversely, the GLP-1 receptor is localized on islet α and β cells and in

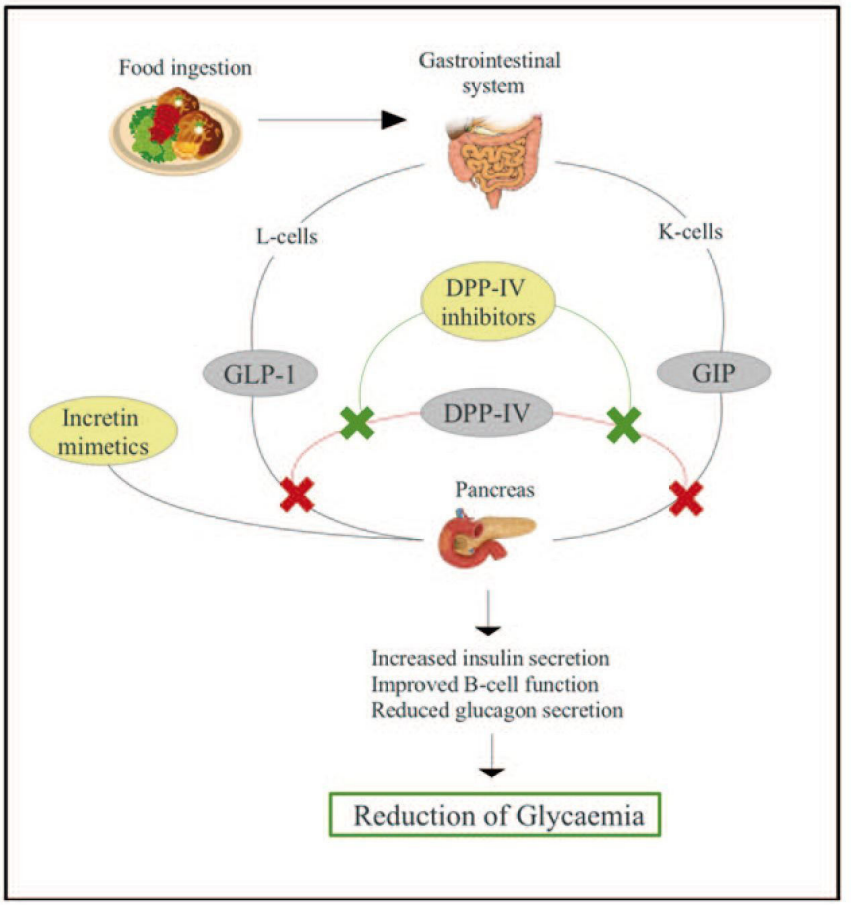


Fig. 7.2 The incretin system. Relationship between the physiological effects of *GLP-1* and *GIP* on insulin secretion and the action of targets implied in T2DM treatment. *GLP-1* and *GIP* are released from enteroendocrine cells after nutrient ingestion to stimulate insulin secretion. However, their activity is reduced because of the cleavage of *DPP-IV* at the second residue of *GLP-1* and *GIP*. Two alternatives to avoid the cleavage are administration of incretin mimetics or *DPP-IV* inhibitors

peripheral tissues, such as the heart, kidneys, lungs, gastrointestinal tract, and peripheral nervous system [57]. As a result of β cell activation, the levels of cAMP and intracellular calcium increase rapidly [57, 58]. This causes insulin secretion in a glucose-dependent manner because of their action after nutrient ingestion [58].

The incretin effect is involved in multiple actions that stimulate all stages of insulin biosynthesis and secretion to reduce the levels of glucose after food ingestion. GLP-1 acts on α cells by suppressing the secretion of glucagon, which has been demonstrated to reduce the risk of hyperglycemia [58]. GLP-1 has a trophic effect on β cells. It not only stimulates their proliferation but also enhances the differentiation of pancreatic cells and reduces apoptosis [49, 59]. Moreover, this gastrointestinal hormone slows gastric emptying and can reduce the postprandial

glucose levels. These effects are similar to inhibiting appetite and food intake [49]. In addition, GLP-1 protects against ischemic and reperfused myocardium injury in rats via mechanisms independent of insulin because of the receptors expressed in this tissue. The hormone may also possess neuroprotective effects. GLP-1 has been proposed as a new therapeutic agent for neurodegenerative diseases such as Alzheimer's disease [49, 58, 59].

Similar to GLP-1, GIP increases insulin biosynthesis and secretion and has a protective activity on β cells. In addition, GIP stimulates the release of glucagon, and it is implicated in lipid metabolism and adiposity [60].

7.2.3 Incretins in T2DM Patients

Although patients with T2DM produce normal levels of GIP, the reduced response to the insulinotropic actions may be related to a reduction in receptor expression or reduced β cell sensitivity to GIP. However, GLP-1 maintains full physiological efficacy, despite being produced in lower concentrations [56, 61, 62]. Although GLP-1 and GIP are responsible for 50–70% of postprandial insulin release in healthy subjects, the incretin effect contributes to only 20–35% of the insulin response to oral glucose in T2DM patients [48]. A reduced insulinotropic effect is also found in healthy subjects with experimental insulin resistance induced by a combination of a high-fat diet, sedentary lifestyle, and steroid therapy [48, 63].

7.3 DPP-IV Inhibition in Detail

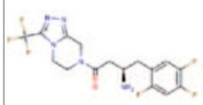
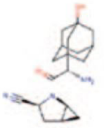
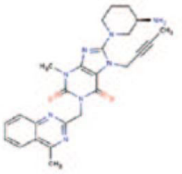
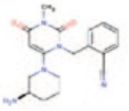
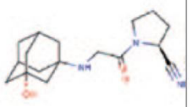
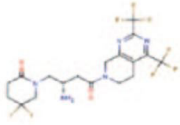
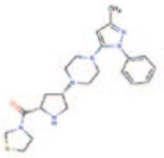
7.3.1 Commercially Available DPP-IV Inhibitors

The inhibition of DPP-IV in humans increases the circulating GLP-1 and GIP levels (and, consequently, prolongs their action), which leads to decreased levels of blood glucose, HbA_{1c}, and glucagon. Therefore, DPP-IV inhibition improves glucose homeostasis with a lower risk of hypoglycemia. As a result, DPP-IV inhibitors are of considerable interest to the pharmaceutical industry [64]. Intensive research activities in this field have resulted in the launch of sitagliptin, saxagliptin, alogliptin, linagliptin, vildagliptin, gemigliptin, and teneligliptin (collectively called as *gliptins*) to the market (Table 7.2) [19, 65].

7.3.2 Side Effects of Commercially Available DPP-IV Inhibitors

A recent post (March 14, 2013) at the sitagliptin [66], saxagliptin [67], and linagliptin [68] pages on MedLinePlus showed that the US Food and Drug Administration (FDA) is evaluating unpublished new findings by a group of academic

Table 7.2 Main features of commercially available DPP-IV inhibitors

Pharmacological name	Commercial name and developer	FDA approval	Advantages	Adverse effects	Selectivity over DPP8/9
 Sitagliptin	**Januvia®** (Merck & Co)	October 17th, 2006	Free from major drug interactions, well-tolerated, moderately efficacious, weight-neutral, low incidence of hypoglycemia, particular role in kidney or liver dysfunction	Abdominal pain, nausea, diarrhea, nasopharyngitis, back pain, osteoarthritis	2600-fold greater affinity
 Saxagliptin	**Onglyza®** (BMS & AstraZeneca)	July 31st, 2009	Well tolerated, safe to use in renal failure, not affect blood pressure, lipid levels, body weight or cardiovascular markers	Headache, upper respiratory infections, arthralgia, nausea, cough	390 and 77-fold greater affinity, respectively
 Linagliptin	**Tradjenta®** (Boehringer Ingelheim International GmbH & Co)	May 2nd, 2011	Once-daily oral dosing, high affinity, no dose restriction in patient with nephropathy, no drug-drug interaction, weight neutrality	Muscle pain, headache, nausea, vomiting, loss of appetite	40000 and >10000-fold greater affinity, respectively
 Alogliptin	**Nesina®** (Furiex pharmaceuticals)	January 25th, 2013	No significant interaction with other drugs, absorption is not affected by food ingestion	Headache, dizziness, constipation	>14000-fold greater affinity
 Vildagliptin	Galvus®, Jalra® or Xiliarx® (Novartis Europharm)	(a)	High specificity, durable response	Upper respiratory infection, dizziness, hypoglycemia, headache	270 and 32-fold greater affinity, respectively
 Gemigliptin	Zemiglo® (LG life Sciences)	(b)	Once-daily oral dosing, well tolerated, low rate of hypoglycemia	headache, dizziness, nausea, epistaxis , and possible increased heart rate	3000-fold greater affinity
 Teneligliptin	TENELIA® Mitsubishi Tanabe Pharma Corporation and Daiichi Sankyo & Co)	(c)	well tolerated, safe, potent and significantly improves glycemic control. inhibited the accumulation of lipids	Risk of hypoglycemia and constipation	700-1500-fold greater affinity

(a) The Europa Union since September 26th, 2007
(b) Korea since June 2012
(c) Japan since September 2012

researchers. The new data suggest an increased risk of pancreatitis and precancerous cellular changes called pancreatic duct metaplasia in patients with T2DM who were treated with these drugs. It is important to note this early communication from the FDA is intended only to inform the public and health-care professionals that the Agency intends to obtain and evaluate the new information before reaching any conclusions about the safety risks of these drugs.

Interestingly, it has been reported that patients with T2DM have a two- to threefold increased risk of suffering from acute pancreatitis [69]. However, other reported studies suggest no increased risk of pancreatitis or malignancy in clinical trials with these drugs [70–75]. For instance, in a pooled analysis of 19 randomized double-blind clinical trials that included data from 10,246 patients, the incidence of acute pancreatitis was 0.10/100 patient–years in the placebo group and 0.08/100 patient–years in the sitagliptin group [71]. A recent analysis has updated the safety and tolerability of sitagliptin by examining pooled data from 25 double-blind clinical studies that lasted up to 2 years. These studies included data from 14,611 patients and concluded that treatment with sitagliptin is not associated with an increased risk of major adverse cardiovascular events, malignancy, or pancreatitis [72]. Therefore, it is likely that sitagliptin does not play a causal role in the reported instances of pancreatitis [72]. Moreover, clinical trials have not demonstrated an increased risk of renal failure with sitagliptin administration [71], and other studies suggest that sitagliptin, saxagliptin, and linagliptin may be used in patients with advanced kidney disease [76, 77].

7.3.3 DPP-IV-Binding Site Description

The DPP-IV binding site is highly druggable in the sense that tight and specific binding to the enzyme can be achieved using small molecules that have drug-like physicochemical properties [56, 78]. It is accessible in two ways: (1) via an opening in the β-propeller domain or (2) via the large side opening, which is formed at the interface of the β-propeller and α/β-hydrolase domain (Fig. 7.1) [18, 19, 23]. The structural features of DPP-IV suggest that substrates and inhibitors enter or leave the binding site via the side opening. Thus, the ligands can directly reach the active site and are correctly oriented for the subsequent cleavage. However, this possibility has not been fully elucidated [18, 79, 80].

In the active site of a protease, there are subsites labeled according to the peptide residue that they bind [81]. The point of peptide cleavage is between the peptide bond that binds residue P_1 with residue P'_1. As a result, the residues that surround this position are labeled relative to the cleavage site as P_2, P_1, P'_1, P'_2, and so on. Therefore, the protein subsites occupied by residues P_2, P_1, P'_1, and P'_2 are labeled as S_2, S_1, S'_1, and S'_2, respectively.

The analysis of the different DPP-IV/inhibitor complexes available at the protein data bank (PDB) has allowed the following different subsites to be identified for DPP-IV (Fig. 7.3 and Table 7.3) [21, 78, 80, 82–86]: (a) the N-terminal recognition is formed by residues Glu205, Glu206, and Tyr662 where the Glu205 (and, in

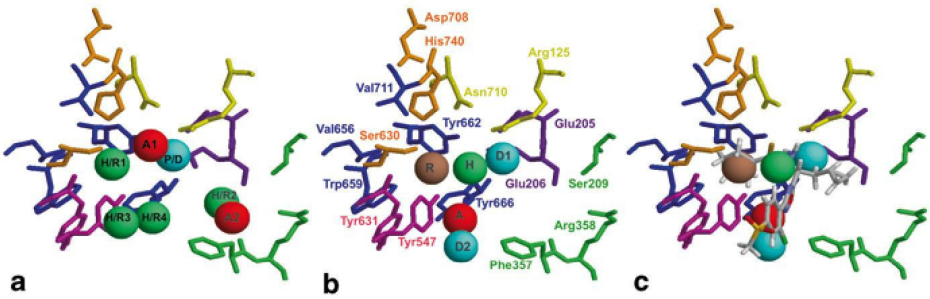


Fig. 7.3 DPP-IV binding site description. Residues belonging to the N-terminal recognition, the S_2 extensive subsite, the S_2 subsite, the S_1 subsite, the catalytic triad, the oxyanion hole and the P_2 amide recognition region are shown in *purple*, *light green*, *green*, *blue*, *orange*, *pink* and *yellow*, respectively. **a** Structure-based energetic pharmacophore built from the PDB structure of 10 complexes of DPP-IV with potent (IC_{50} values ≤ 10 nM) reversible inhibitors of a non-peptide nature [99]; **b** fragment-based energetic pharmacophore built after docking a library of rigid fragments at the DPP-IV binding site and further clustering of the fragments with highest binding energy; **c** the DPP-IV inhibitor from the PDB structure 3C45 in the context of the binding-site and of the fragment-based energetic pharmacophore. Pharmacophore sites are labeled according to their chemical characteristics (*H* hydrophobic, *R* aromatic ring, *P* polar, *D* hydrogen bond donor sites and *A* hydrogen bond acceptor sites; sites labeled as *H/R* and *P/D* accept two different chemical features). All three panels are in the same orientation to facilitate the comparison

some cases, Glu206) forms a salt bridge/hydrogen bond with the peptide's basic amine; (b) the S_2 pocket is formed by the residues Arg125, Ser209, Phe357, Arg358, Tyr547, and Asn710, where Arg125 and Asn710 are essential to coordinate the carbonyl of the P_2 residue and, together with Glu205 and Glu206, align the substrate optimally for the nucleophilic attack by Ser630 [87]; (c) the oxyanion hole is formed by the backbone NH of Tyr631 and the side chain OH of Tyr547 and stabilizes the negatively charged tetrahedral oxyanion intermediate that is generated in the transition state [87]; (d) the S_1 pocket is formed by the residues Tyr631, Val656, Trp659, Tyr662, Tyr666, and Val711; and (e) the catalytic triad is formed by the residues Ser630, Asp708, and His740 (with Ser630 cleaving the peptide bond between P_1 and P'_1 by performing a nucleophilic attack). Although in principle, no subsites are defined further than S_2 in DPP-IV, a recent study has shown that the inhibitors and not the substrates can bind well beyond the S_2 subsite to increase their inhibitory activity [88, 89]. The site beyond S_2 was defined as the S_2 extensive subsite and is formed by Val207, Ser209, Phe357, and Arg358 [23].

Based on the analysis of the DPP-IV crystal structures [90–96] and the interpretation of the structure–activity relationship data, both the lipophilic S_1 pocket and the Glu205/Glu206 dyad can be considered as crucial molecular anchors for DPP-IV inhibition [78]. Moreover, this conclusion is supported by results derived from two different energetic pharmacophores [97, 98] obtained by our group that have quantified the relative contribution of the different pharmacophore sites to the intermolecular interactions with DPP-IV. The first energetic pharmacophore was built from the PDB structure of ten complexes of DPP-IV with potent (IC_{50} values ≤ 10 nM) reversible inhibitors of a nonpeptide nature (Fig. 7.3a) [99]. This study showed that

Table 7.3 Intermolecular interactions between potent (IC_{50} values ≤ 10 nM) and reversible nonpeptide inhibitors in the DPP-IV binding site of available PDB structures

PDB code	Ligand	IC_{50} (nM)	S_2 subsite extensive	N-terminal recognition	S_2 subsite	P_2 amide recognition	Oxyanion hole	S_1 subsite	Enzyme catalitic triad
3C45	317	0.21		SaltB/HBond	Hydroph			Hydroph	
3VJM	W61	0.37	Hydroph	SaltB/HBond	Hydroph	HBond		Hydroph	
3H0C	PS4	0.38		SaltB/HBond	Hydroph	Hydroph		Hydroph	Hydroph
3KWJ	23Q	0.5		SaltB/HBond	Hydroph		Hydroph	Hydroph	
2RGU	356	1		SaltB/HBond			HBond/π-stacking	Hydroph	Hydroph
2QT9	524	2.3	Hydroph	SaltB/HBond	Hydroph	HBond		Hydroph	Hydroph
2IIT	872	2.6		SaltB/HBond	Hydroph	HBond		Hydroph	Hydroph
3HAB	677	4.2		SaltB/HBond	Hydroph			Hydroph	Hydroph
2QTB	474	4.8	Hydroph	SaltB/HBond	HBond/Hydroph	HBond		Hydroph	Hydroph
3G0D	XIH	5		SaltB/HBond		HBond	HBond	Hydroph	
3G0G	RUM	5		SaltB/HBond		HBond	HBond	Hydroph	
3O95	01T	5.3		SaltB/HBond	Hydroph	HBond		Hydroph	HBond
3VJL	W94	5.6		SaltB/HBond	Hydroph	HBond		Hydroph	
2QJR	PZF	6.4	HBond/π-stacking	SaltB/HBond	Hydroph			Hydroph	Hydroph
2IIV	565	6.6		SaltB/HBond	Hydroph			Hydroph	Hydroph
3HAC	361	6.7		SaltB/HBond	Hydroph			Hydroph	
3KWF	B1Q	6.8	Hydroph	SaltB/HBond	Hydroph	Cation-dipole/ HBond		Hydroph	Hydroph
3G0B	T22	7		SaltB/HBond		HBond	HBond	Hydroph	

Rows are sorted according to increasing IC_{50}. The data have been obtained from the literature and from the analysis of the corresponding LigPlot diagrams [207]. *Hydroph*, *SaltB*, and *HBond* refer to hydrophobic contacts, salt bridges, and hydrogen bonds, respectively

two of the six sites of the pharmacophore (**P/D** and **H/R1**): (a) were accomplished by all ten inhibitors, (b) accounted for more than 90 % of the inhibitor/DPP-IV binding energy, and (c) were located in the two previously identified crucial molecular anchors for DPP-IV inhibition (**P/D** and **H/R1** are close to the Glu205/Glu206 dyad and the S_1 pocket, respectively). The second energetic pharmacophore (unpublished results) has been obtained after (1) docking a library of rigid fragments at the DPP-IV binding site and (2) further clustering of the fragments with the highest binding energy. This fragment-based energetic pharmacophore is formed by five relevant sites (i.e., two hydrogen-bond donors**,** one hydrogen-bond acceptor, one hydrophobic site, and one aromatic ring; Fig. 7.3b). According to our results, two of these five sites (**R** and **H**) show a very large contribution to the binding energy (scores of −10.05 and −5.77 kcal/mol, respectively) compared with the remaining three binding energies (scores of −2.71, −2.09 and −1.33 kcal/mol). Interestingly, the comparison of the energetic pharmacophores in Figs. 7.3a and b show that (a) the *P/D* site at Fig. 7.3a matches the *D1* site at Fig. 7.3b; and (b) the *H/R1* site in Fig. 7.3a approximately matches the *R* site at Fig. 7.3b. Therefore, both energetic pharmacophores highlight the importance of the N-terminal recognition performed by the Glu205/Glu206 dyad and the intermolecular interactions at the hydrophobic S_1 site. Moreover, other studies suggest that the binding free energy can be further improved by additional favorable contacts [84] with the following: (a) the catalytic triad, (b) the oxyanion hole, (c) the P_2 amide recognition region (formed by Arg125 and Asn710) where, for instance, Arg125 can stabilize the amide carbonyl moiety of an inhibitor by making a hydrogen bond with it [82], (d) the phenyl rings from Phe357 and Tyr547 (by interacting with different aromatic ligand fragments to give π–π stacking interaction or by making hydrophobic contacts with large aliphatic groups) [80, 84], or (e) Arg358, which uses its positively charged side chain to interact with substituents on the ligand's aromatic rings or to place electronegative groups on the ligands close to its positive-charged side chain [84].

Interestingly, the comparison of Figs. 7.3a and b also shows that there are unexplored ways to inhibit DPP-IV. In the fragment-based pharmacophore sites *A* and *D2* (with scores of −2.71 and −2.09 kcal/mol, respectively) located between the residues Phe357, Tyr547, and Tyr666 (Fig. 7.3b) are not present at the PDB-based energetic pharmacophore (Fig. 7.3a). A similar situation occurs for the *H* site (located at the center of the DPP-IV binding site; Fig. 7.3b) that, as mentioned before, has a very large score (−5.77 kcal/mol). As a result, it is remarkable that only three of the ten experimental poses that were used to derive the structure-based pharmacophore are able to simultaneously fit the *R, H,* and *D1* sites of the fragment-based pharmacophore (unpublished results). Therefore, it can be concluded that the use of the fragment-based pharmacophore in a virtual screening could identify previously undescribed DPP-IV inhibitors in molecular databases by reducing the bias toward the existing covered space of the binding site. Our group is currently using this pharmacophore to identify potent DPP-IV inhibitors in the molecules found in nontoxic mushrooms of the Catalan forests. Our aim is to use extracts rich in these bioactive molecules as food additives for people affected (or potentially affected) by T2DM.

7.3.4 How Differences at the Binding Site Among DPP-IV, DPP8, and DPP9 Explain the Selective Inhibition of DPP-IV

Unlike DPP-IV and FAP, the 3D structures for DPP8 and DPP9 are unknown. However, the structures can be built by homology modeling [100–102]. A comparison of the binding sites in DPP-IV, DPP8, and DPP9 suggests how to look for (or design) potent DPP-IV inhibitors with no (or low) bioactivity on DPP8/9 [103]. This comparison shows the following: (a) the S_1 pocket is significantly smaller in DPP-IV (27.72 Å^3) than it is in DPP8 (99.77 Å^3) and DPP9 (75.89 Å^3) [103, 104–106], which suggests that the excluded volumes obtained for this pocket in DPP-IV can be used to remove DPP8/9 inhibitors during the virtual screening (VS) workflow, (b) the Glu205/Glu206 dyad side chains are oriented towards the ligand site in DPP-IV where they form a salt bridge with ligands whereas in DPP8/9 one of the two equivalent glutamic acids (Glu276 for DPP8 and Glu249 for DPP9) has its side chain oriented away from the active site (consequently, its intermolecular interaction with a ligand hydrophilic group is not as strong as it is in DPP-IV [103, 106], which can result in a lower docking score for the same ligand in DPP8/9 relative to DPP-IV), and (c) whereas the S_2 extensive subsite has not been clearly defined for DPP8/9, it has been shown to be important for the potency and selectivity of DPP-IV inhibitors [23, 27, 78, 88, 100, 105, 107].

7.3.5 How to Predict DPP-IV Selective Inhibition

The relevance of selectivity in the clinical application of DPP-IV inhibitors is an essential step in reducing the toxicity associated with the inhibition of DPP8 and DPP9 [45]. Thus, the importance of computational approaches in designing or looking for selective DPP-IV inhibitors has become indispensable [103]. Various in silico methods have been described, mostly supported by docking studies on DPP8 and DPP9 enzymes [101, 103, 104], which could be subsequently followed either by finding molecules that show a significant higher (i.e., more negative) score for DPP-IV than for DPP8/9 [103], or by a 3D-QSAR study that uses the aligned docked poses to build a predictive model [104]. In contrast, it has been recently used as a conformational-free ligand-based methodology (i.e., holographic QSAR or HQSAR) for predicting DPP-IV selectivity [108] that has the advantage that eliminates the need for generation of the putative binding conformations at the different binding sites and their subsequent 3D-structure alignment. HQSAR involves the investigation of important indications of the molecular fragments that are directly related to biological activity or responsible for the low biological potency of the compounds, and this method is used to propose structural modifications. Therefore, contribution maps indicating the individual contributions to the activity of each atom in a given molecule of the data set can be obtained. Additionally, the most relevant structural fragments can be analyzed.

7.3.6 *Natural Products as DPP-IV Inhibitors*

Dietary intervention is accepted as a key component in the prevention and management of T2DM [109]. Natural products are useful as bioactive components to develop new functional foods for specific population sectors [110–112]. A functional food has been defined as "any modified food or food ingredient that may provide a health benefit beyond the traditional nutrients it contains" [113]. According to the literature, the capacity to inhibit DPP-IV has been identified in natural nonpeptide (Fig. 7.4) [99, 114–125] and peptide products (Table 7.4) [18, 126–142]. Therefore, they could be used as bioactive ingredients in functional foods for T2DM prevention or treatment [99, 126]. These foods may also serve as lead compounds for deriving more potent DPP-IV inhibitors [99, 117, 143].

7.3.6.1 Natural Products of Nonpeptide Nature

There are presently a limited number of DPP-IV inhibitors that have a nonpeptide nature (see Fig. 7.4 for the most relevant examples). Akiyama et al. [144] isolated sulphostin from the culture broth of *Streptomyces sp.* MK251–43F3. This molecule exhibits an antidiabetic activity that is 100-fold stronger than the well-known DPP-IV peptide inhibitor diprotin A [145]. Berberine [115], trigonelline [116], and eight different DPP-IV inhibitors [119] have been isolated from different plants (e.g., *Coptis chinensis, Trigonella foenum-graecum, Bacopa monnieri,* and *Daphne odora*) and are widely used as antihyperglycemic agents in traditional Chinese medicine (TCM). Moreover, curcumin (isolated from the rhizome of the herb *Curcuma longa*), resveratrol, luteolin, apigenin, flavone, and naringin (commonly found in berry wine blends, citrus, berry, grape, and soybean) are plant phenolic compounds that are also DPP-IV inhibitors [117, 121, 122]. Moreover, different natural extracts inhibit DPP-IV, although the specific nonpeptide molecules that are responsible for this bioactivity have not been fully characterized [114, 123–125].

7.3.6.2 Naturally Derived Peptides

Protein–peptide interactions are vital for life because peptides can take part in nearly 40 % of macromolecular interaction-mediating signals [146]. In recent years, studies on peptides derived from food proteins have shown that their bioactivity can significantly improve human health and prevent chronic diseases [126]. These bioactive peptides are short peptide sequences that are typically less than ten amino acids. They are encrypted within the structure of a food protein and can be released by enzyme hydrolysis, microbial fermentation, or physical and chemical processing [18]. The peptides can interact with specific receptors and regulate a variety of physiological functions. Interestingly, peptides offer certain advantages as drugs due to their high biological activity, high specificity, and low toxicity [147].

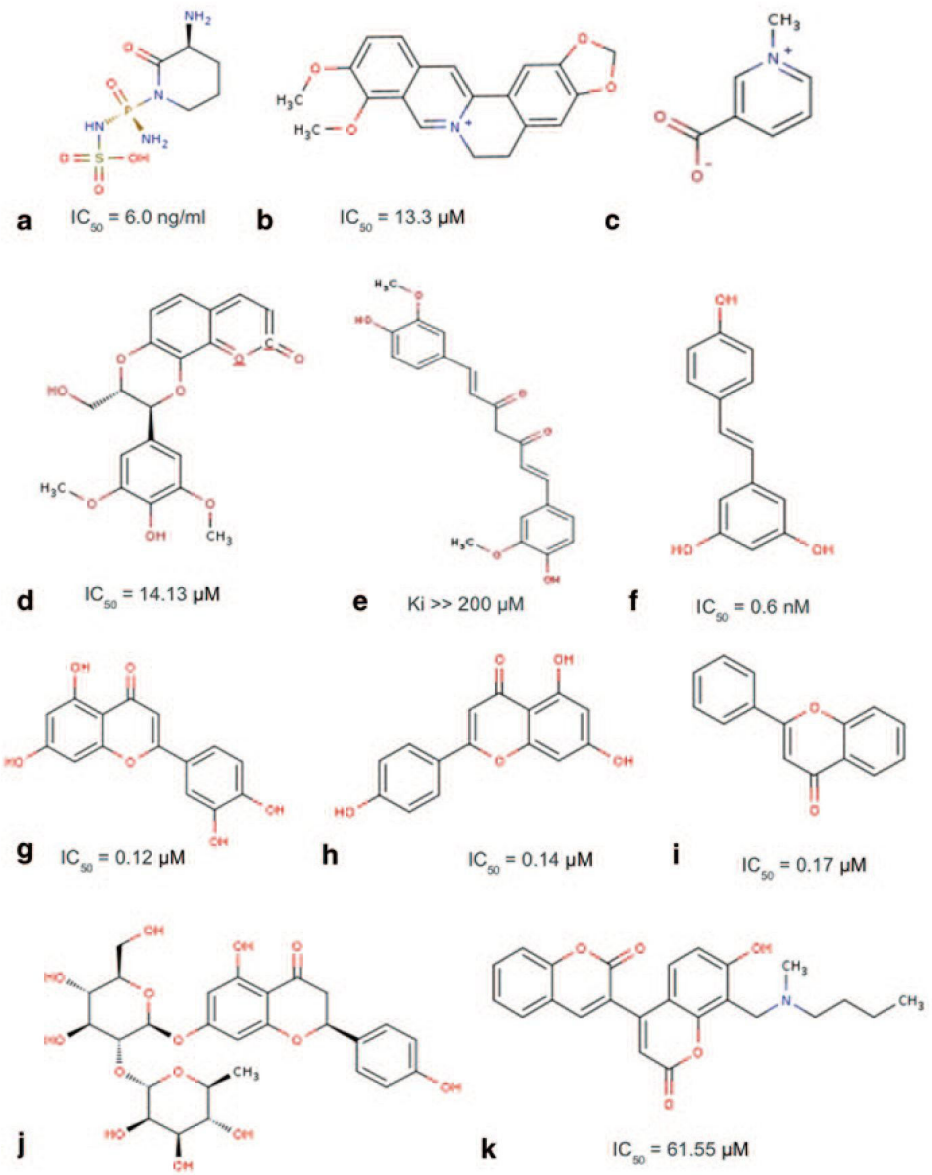


Fig. 7.4 Chemical structures and DPP-IV inhibitory activity for the most relevant natural compounds of non-peptide nature: **a** sulphostin; **b** berberine; **c** trigonelline; **d** compound 4; **e** curcumin; **f** resveratrol; **g** luteolin; **h** apigenin; **i** flavone; **j** naringin; and **k** ZINC02132035

Several recent studies have demonstrated that peptides obtained from proteins from the following sources are able to inhibit DPP-IV: dairy products [126, 127, 129–131, 135, 139–141], defatted rice bran [132], tuna cooking juice [133], dry-cured ham [134], *Amaranthus hypochondriacus* [136], barley [126], canola [126], oat [126], soybean [126], wheat [126], chicken egg [126], bovine meat [126, 142],

Table 7.4 Peptide sequences that inhibit DPP-IV according to the literature

Peptide sequence	IC50 (μM)	Type of inhibition
Ile-*Pro*-Ile (diprotin A)*	3.4–24.7	Competitive
Val-*Pro*-Leu (diprotin B)	5.5	Competitive
Ile-*Pro*-Ile-Gln-Tyr*	35.2	Competitive
Gly-*Pro*-Gly-Ala*	41.9	
Ile-*Pro*-Ala-Val-Phe	44.7	
Leu-Lys-Pro-Thr-Pro-Glu-Gly-Leu-Asp*	45	Un-competitive
Leu-*Pro*-Gln-Asn-Ile-Pro-Pro-Leu	46	
Ile-*Pro*-Ala	49	
Gly-*Pro*-Ala-Glu*	49.6	
Leu-Lys-Pro-Thr-Pro-Glu-Gly-Leu-Asp-Leu-Glu-Ile-Leu*	57	Un-competitive
Trp-Val*	65.69	Non-competitive
Cys-Ala-Tyr-Gln-Trp-Gln-Arg-Pro-Val-Asp-Arg-Ile-Arg*	78	
Leu-*Pro*-Gln	82	
Pro-Ala-Cys-Gly- Gly-Phe-Try-Ile-Ser-Gly-Arg-Pro-Gly*	96.4	
Leu-*Pro*-Tyr-Pro-Tyr *	108.3	Competitive
Val-*Pro*-Ile-Thr-Pro-Thr-Leu	110	
Pro-Gly-Val-Gly-Gly-Pro-Leu-Gly-Pro-Ile-Gly-Pro-Cys-Tyr-Glu*	116.1	
Val-*Pro*-Ile-Thr-Pro-Thr	130	
Trp-Leu-Ala-His-Lys-Ala-Leu-Cys-Ser-Glu-Lys-Leu-Asp-Gln*	141	Un-competitive
Ile-*Pro*-Ala-Val-Phe-Lys	143	
His-Leu*	143.19	Competitive
Ile-*Pro**	149.6	Competitive
Leu-*Pro*-Gln-Asn-Ile-Pro-Pro	160	
Leu-Ala-His-Lys-Ala-Leu-Cys-Ser-Glu-Lys-Leu*	165	Competitive
Thr-Lys-Cys-Glu-Val-Phe-Arg-Glu*	166	Un-competitive
Val-Ala*	168.24	Competitive
Val-Ala-Gly-Thr-Trp-Tyr	174	
Leu-Cys-Ser-Glu-Lys-Leu-Asp-Gln*	186	Non-competitive
Ile-*Pro*-Ala-Val-Phe-Lys-Ile-Asp-Ala*	191	Competitive
Tyr-*Pro*-Tyr-Tyr*	194.4	Competitive
Leu-*Pro*-Leu*	241.4	Competitive
Tyr-*Pro*-Tyr*	243.7	Competitive
Phe-*Pro*-Gly-Pro-Ile-Pro-Asn	260	
Ile-Leu-Asp-Lys-Val-Gly-Ile-Asn-Tyr*	263	Competitive
Trp-Leu-Ala-His-Lys-Ala-Leu*	286	Non-competitive
Thr-*Pro*-Glu-Val-Asp-Asp-Glu-Ala-Leu-Glu-Lys	319.5	
Leu-*Pro*-Leu-Pro-Leu*	325	Competitive
Ile-Val-Gln-Asn-Asn-Asp-Ser-Thr-Glu-Tyr-Gly-Leu-Phe*	337	Non-competitive
Phe-Leu*	399.58	Competitive
Ile-*Pro*	410	Competitive
Val-Leu-Val-Leu-Asp-Thr-Asp-Tyr-Lys	424.4	
Tyr-*Pro**	658.1	Competitive
Tyr-*Pro*-Phe-Pro-Gly-Pro-Ile-Pro-Asn	670	
Leu-*Pro**	712.5	Competitive
Met-*Pro*	870	Competitive
Val-*Pro*	880	Competitive

Table 7.4 (continued)

Peptide sequence	IC50 (μM)	Type of inhibition
Ala-Leu*	882.13	Competitive
Pro-Gly-Pro-Ile-His-Asn-Ser	1000	
Ile-*Pro*-Pro-Leu-Thr-Gln-Thr-Pro-Val	1300	
Pro-Gln-Asn-Ile-Pro-Pro-Leu	1500	
Arg-*Pro*	2240	Competitive
Thr-*Pro*	2370	Competitive
Leu-*Pro*	2370	Competitive
Met*	2381.51	Competitive
Val-*Pro*-Pro-Phe-Ile-Gln-Pro-Glu	2500	
Ser-Leu*	2517.08	Competitive
Lys-*Pro*	2540	Competitive
Gly-Leu*	2615.03	Competitive
His-*Pro*	2820	Competitive
Tyr-*Pro*	3170	Competitive
Glu-Lys*	3216.75	Competitive
Leu*	3419.25	Competitive
Phe-*Pro*	3630	Competitive
Trp*	4280.4	Competitive
Trp-*Pro*	4530	Competitive
Pro-*Pro*	5860	Competitive
Ser-*Pro*	5980	Competitive
Lys-Ala*	6270	
Ala-Ala-Ala-Thr-Pro*	6470	
Ala-*Pro*	7950	Competitive
Ala-Ala-Ala-Ala-Gly*	8130	
Ala-Ala*	9400	
Gly-*Pro**	9690	

Rows are sorted according to increasing IC_{50}. The presence of Pro at the P_1 position of some peptides is highlighted

*The IC_{50} value has been measured with porcine instead of human DPP-IV

and chum and Atlantic salmon (Table 7.4) [126, 138]. They are usually di-, tri-, and oligopeptides that contain proline and/or hydrophobic amino acids within their sequence [126]. Moreover, the sequence of the peptide, not its amino acid composition, influences the DPP-IV inhibitory activity. For instance, the dipeptides Ile–Pro and Trp–Val had DPP-IV inhibitory activity (Table 7.4). However, the reverse peptides Pro–Ile and Val–Trp had no inhibitory activity [130, 132]. Thus, proline is the preferential amino acid residue at the P_1-position. Furthermore, alanine, glycine, and serine are also accepted (Table 7.4). The data in Table 7.4 also show that (a) dipeptides of the general structures Xaa-Pro (except Gly-Pro) are competitive inhibitors of DPP-IV [148], and (b) the residue present at the N-terminus influences inhibitory activity because the dipeptide Leu-Pro has a higher IC_{50} value than Ile-Pro (see Table 7.4) [18].

Longer peptides (larger than 13 residues) have been shown to act as noncompetitive inhibitors by forming interactions at the dimerization interface and blocking the formation of the DPP-IV active dimer [136, 149].

7.4 Using In Silico Tools for Identifying DPP-IV Inhibitors of Natural Origin

The identification of inhibitors with previously undescribed bioactivities in natural extracts exclusively by in vitro or in vivo approaches is a complex and expensive process [114–117, 127, 129–135, 138–140, 144]. The use of in silico approaches can significantly increase this identification of natural extracts. There are successful examples of newly identified DPP-IV inhibitors of natural origin that have been found using either VS workflows [99, 120] or target fishing [119] or sequence similarity tools [126, 141, 142].

7.4.1 Virtual Screening Workflows

7.4.1.1 Defining Virtual Screening Workflows

A VS workflow consists of several sequential *filters* that are used to discern the molecules that share and those that do not share properties that characterize drugs with a specific bioactivity. In a VS workflow, the molecules that survive a filter are then evaluated by the next filter (whereas the rest are rejected). Thus, a VS workflow is described as a funnel shape to indicate the decreasing number of molecules that are evaluated by the successive filters (Fig. 7.5). Some of the most commonly used filters during VS workflows include ADME/Toxicity analysis, protein–ligand docking, pharmacophore matching and similarity/electrostatic comparison (Fig. 7.5) [99, 120].

7.4.1.2 Natural Products Databases

The main goal of using a VS workflow and finding bioactive molecules for functional food design is to find a cheap natural source that can easily provide extracts enriched in the bioactive molecule. Therefore, it is necessary to use databases for naturally occurring molecules that, in addition to showing the molecular structure, include the natural source from which these molecules can be obtained. Examples of such databases are the NuBBE database [150], the TCM database @Taiwan [151], and Reaxys [152].

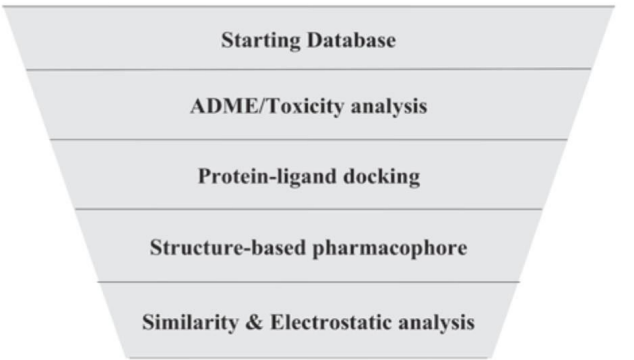


Fig. 7.5 Overview of a typical virtual screening workflow

7.4.1.3 Examples

We have developed a VS workflow to successfully identify molecules that are able to inhibit DPP-IV and molecules that do not inhibit this enzyme [99]. Among other filters, this VS workflow included a structure-based energetic pharmacophore (Fig. 7.3a) that was obtained from the consensus of the different energetic pharmacophores [97] that can be obtained from ten different complexes between human DPP-IV and potent reversible inhibitors (i.e., IC50 values ≤10 nM) of nonpeptide nature available in the PDB [153]. This VS workflow was applied to the *Natural Products* subset of the ZINC database [154]. The results predicted that 446 of the 89,425 molecules present in the database could be potential DPP-IV inhibitors. 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. We found that there were 12 clusters that contained only natural products not previously identified as DPP-IV inhibitors [99]. Nine molecules from 7 of the 12 clusters (from which no antidiabetic activity has been described to date) were selected for in vitro activity testing. The results of the in vitro activity testing showed the following: (a) seven molecules that could be solubilized inhibited DPP-IV, and (b) the most potent compound was ZINC02132035 (with an IC_{50} of 61.55 μM; Fig. 7.4k) [99]. Therefore, we experimentally demonstrated that the VS workflow was able to identify DPP-IV inhibitor molecules that (1) have never been reported to have antidiabetic activity and (2) were not structurally related to any known DPP-IV inhibitor.

We next used a slightly modified version of the VS workflow to evaluate an in-house database of 29,779 natural products annotated with their natural source. We were able to identify 84 molecules (isolated from 95 different natural sources) that were predicted to inhibit DPP-IV [120]. An exhaustive bibliographic search revealed that we predicted 12 potential DPP-IV inhibitors from 12 different plant extracts that are known to have antidiabetic activity (Table 7.5). Six of these 12 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; Table 7.5). Therefore, it is plausible that these 12 molecules could be partially responsible for the antidiabetic activity of these extracts through DPP-IV inhibition [120]. In addition, we identified six potential DPP-IV inhibitor molecules from six

Table 7.5 Natural extracts with reported antidiabetic activity that contain molecules predicted to be DPP-IV inhibitors by our VS protocol [120]

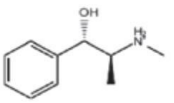
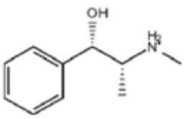
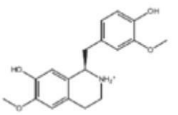
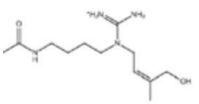
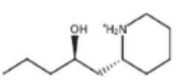
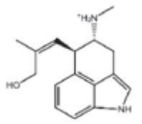
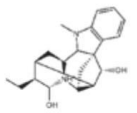
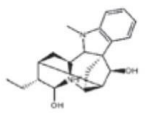
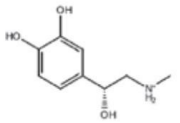
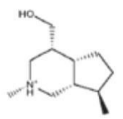
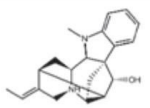
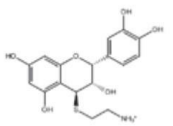
Molecule	Name and CAS number (when available)	Extract	Ref. Isolation molecule from extract	Ref. Antidiabetic extract	Ref. Antidiabetic molecule
	(+)-pseudoephedrine (90-82-4)	*Ephedra alata*	[155]	[156]	[157]
	(-)-ephedrine (299-42-3)	*Ephedra distachya*	[158]	[155]	[155]
	N-nororientalin (29079-44-5)	*Erythrina variegata*	[159]	[160]	[161-163]
	hydroxysmirnovine	*Galega orientalis*	[164]	[165]	
	(-)-halosaline (26648-71-5)	*Haloxylon salicornicum*	[166]	[156]	

Table 7.5 (continued)

Molecule	Name and CAS number (when available)	Extract	Ref. Isolation molecule from extract	Ref. Antidiabetic extract	Ref. Antidiabetic molecule
	isochanoclavin-(I) (1150-43-2)	*Pennisetum typhoideum*	[167]	[168]	
	ajmaline (509-37-5)	*Rauwolfia serpentina*	[169]	[170]	
	isosandwichine (509-37-5)	*Rauwolfia vomitoria*	[171]	[172]	
	epinephrine (51-43-4)	*Scoparia dulcis*	[173]	[174]	[175]
	tecostanine	*Tecoma stans*	[176]	[177]	[178]
	serpinine (509-38-6)	*Vinca major*	[167]	[179]	
	epicatechin derivate	*Vitis vinifera*	[180]	[181]	[182]

The first column shows the 2D structure of each molecule. The second column shows the corresponding common name and the CAS number (when available). 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 studies that describe the purification of the molecule from the corresponding extract, (b) the second column lists studies that describe the antidiabetic activity of the corresponding extract; and (c) the third column lists studies, when available, that describe the antidiabetic activity of the corresponding molecule or one that is very similar to it

different plants with no described antidiabetic activity. These molecules share the same *genus* as plants with known antidiabetic properties (thus suggesting that they could be new sources for antidiabetic extracts; Table 7.6). Moreover, none of the 18 molecules that we predicted as DPP-IV inhibitors exhibits chemical similarity with any previously known DPP-IV inhibitor [120]. Finally, the same study also predicted 77 other sources with no described antidiabetic activity that contain at least one VS hit. Consequently, this work will permit the discovery of new antidiabetic extracts of natural origin that could be of use in the design of functional foods aimed at preventing/treating T2DM [120].

7.4.2 Target Fishing

7.4.2.1 Defining Target Fishing

Target fishing refers to a computer-assisted methodology used to predict the targets of a specific compound (or a limited set of compounds). Therefore, it can be considered the inverse process of a usual VS workflow. Target fishing has applications in drug repositioning [198] and anticipating potential side effects [199]. Other common synonymous for target fishing are chemogenomics [200], drug repurposing [201], polypharmacology [202], virtual target screening [203], and target profiling [204].

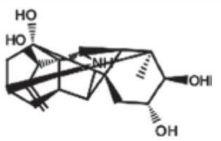
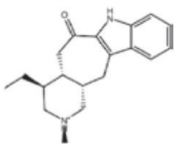
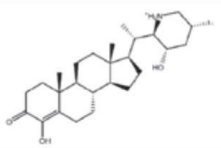
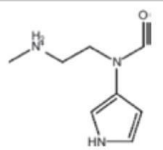
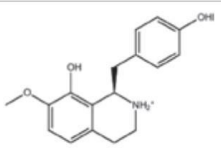
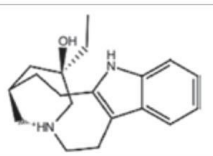
7.4.2.2 Examples

The potential drug target database (PDTD) [205] was searched using the TarFisDock server [206] to identify putative targets for a collection of 19 natural products obtained from *Bacopa monnieri* (L.) Wettst and *Daphne odora* Thunb. var. *marginata* (two plants commonly used by TCM and Ayurvedic medicine in diabetes and inflammation treatment) [119]. This study predicted that from more than 800 drug targets available at PDTD, DPP-IV was one of the most probable for these 19 molecules (consistent with the known therapeutic indications of both plants). Furthermore, an in vitro analysis of the bioactivity of these 19 molecules showed that five have moderate inhibitory activities for DPP-IV (with IC_{50} values ranging from 14.13 to 113.76 μM) [119]. Subsequently, these five molecules were used to identify 27 analogs in the in-house natural products database of the researchers. The in vitro analysis of the bioactivity of these 27 molecules showed that 13 have moderate inhibitory activities for DPP-IV (with IC_{50} values ranging from 22.39 to 87.72 μM) [119].

7.4.3 Sequence Similarity

The aim of these kind of studies consist in performing an in silico evaluation of dietary proteins as potential precursors of biologically active peptides, as well as to

Table 7.6 Natural extracts with no described antidiabetic activity (but from the same *genus* as plants with extracts with described anti-diabetic activity) that contain molecules that are predicted to be DPP-IV inhibitors by our VS protocol [120]

Molecule	CAS number or name	Extract	Ref. Isolation molecule from extract	Extract with antidiabetic activity described	Ref. Antidiabetic extract
	30373-79-6	*Aconitum japonicum*	[183]	*Aconitum carmichaelii*	[184]
				Aconitum moschatum	[185]
				Aconitum violaceum	[185]
	episilicine	*Ervatamia officinalis*	[186]	*Ervatamia microphylla*	[187]
	solanudine	*Solanum nudum*	[188]	*Solanum lycocarpum*	[189]
				Solanum nigrum	[190]
		*Solanum sodomaeum*	[191]	*Solanum xanthocarpum*	[192]
	norjuziphine	*Stephania cepharantha*	[193]	*Stephania hernandifolia*	[194]
				Stephania glabra	[195]
				Stephania tetrandra	[196]
	19637-92-4	*Tabernaemontana eglandulosa*	[197]	*Tabernaemontana divaricata*	[187]

The first column shows the 2D structure of each molecule. The second column shows the corresponding common name and/or the CAS number (when available). The third column lists the source from which the VS hits have been purified (rows in that table are alphabetically sorted based on this column). The fourth column lists the studies that describe the purification of the each molecule from the corresponding extract. The fifth column shows the extracts from the same *genus* where the antidiabetic activity has been described. Finally, the last column lists studies that describe the antidiabetic activity of the corresponding extract

determine whether such peptides can be released by selected proteolytic enzymes [126, 141, 142]. This approach finds biologically active peptides in the protein sequences that remain inactive in precursor protein sequences. However, when released by proteolytic enzymes, these peptides may interact with selected receptors and regulate physiological functions [141]. Thus, the potential of various dietary proteins to serve as DPP-IV inhibitor precursors is predicted by searching for fragments within the protein chains that match the peptide sequences reported in the literature (Table 7.4) to present an inhibitory activity against DPP-IV. This potential is quantified for each protein by calculating A (the occurrence frequency) as $A = a/N$ (where a is the number of peptides with DPP-IV inhibitory activity within the protein chain and N is the number of amino acid residues in the protein chain) [141]. These studies show that β-casein from cow's milk, collagens from bovine meat, and chum salmon have occurrence frequency values of 0.249, 0.380, and 0.305, respectively, and appeared to be the best potential sources of DPP-IV inhibitory peptides among all of the proteins studied [126, 141]. Moreover, it is also shown that DPP-IV inhibitory peptides can be obtained from milk proteins by using serine endopeptidases (e.g., proteinase K, EC.3.4.21.14; pancreatic elastase, EC 3.4.21.36; prolyl oligopeptidase, EC 3.4.21.26; chymotrypsin C, EC 3.4.21.2; and leukocyte elastase, EC 3.4.21.37) or cysteine endopeptidases (papain, EC 3.4.22.2; ficin, EC 3.4.22.3; and bromelain, EC 3.4.22.4) or thermolysin (EC 3.4.24.27). [141] These proteins also hold special interest for the food industry because proteins from the connective tissue (usually with low commercial value) are rich in proline. Therefore, they can be a very important source for DPP-IV inhibitors (Table 7.4) and may represent a new method of generating profit from food industry byproducts.

7.5 Concluding Remarks and Future Perspectives

DPP-IV inhibition appears to be one of the most effective and secure ways of controlling diabetes and related diseases. Three of the seven gliptins that are currently authorized for human use have been released to the market over the last 2 years (Table 7.2). Moreover, DPP-IV inhibitors are orally administered, which makes them compatible with the food additive concept. Therefore, finding naturally available molecules with bioactivity is an area of high interest for the functional food and nutraceutical industry. VS is an essential (and low-cost) tool for predicting new DPP-IV inhibitors from natural molecule databases and recovering them from food-processing byproducts or biomass with low- or no-economic value. Nevertheless, there are some key points that, in our opinion, could improve the performance of VS on DPP-IV and that need to be addressed in future research: (1) including di- and tripeptides in VS studies; (2) improving VS filters to remove molecules that could inhibit FAP, DPP8, or DPP9; and (3) using the dimerization area as the part of the target where ligand binding is predicted during VS. Our lab is making progress in addressing these challenges and has promising results that will be published elsewhere.

References

1. International Diabetes Federation (2013) IDF Diabetes Atlas, 6th edn. Brussels, Belgium: International Diabetes Federation, http://www.idf.org/diabetesatlas
2. Daousi C, Casson IF, Gill GV, MacFarlane IA, Wilding JPH, Pinkney JH (2006) Prevalence of obesity in type 2 diabetes in secondary care: association with cardiovascular risk factors. Postgrad Med J 82:280–284
3. UK Prospective Diabetes Study (UKPDS) Group (1998) Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). Lancet 352:837–853
4. Kahn SE, Haffner SM, Heise MA et al (2006) Glycemic durability of rosiglitazone, metformin, or glyburide monotherapy. N Engl J Med 355:2427–2443
5. Ross SA, Dzida G, Vora J, Khunti K, Kaiser M, Ligthelm RJ (2011) Impact of weight gain on outcomes in type 2 diabetes. Curr Med Res Opin 27:1431–1438
6. Jacobson AM (2004) Impact of improved glycemic control on quality of life in patients with diabetes. Endocr Pract 10:502–508
7. International Diabetes Federation. IDF diabetes atlas. http://www.idf.org/diabetesatlas. Accessed 15 Aug 2013
8. World Health Organization. Diabetes programme. http://www.who.int/diabetes/en/. Accessed 15 Aug 2013
9. Morrish NJ, Wang SL, Stevens LK, Fuller JH, Keen H (2001) Mortality and causes of death in the WHO multinational study of vascular disease in diabetes. Diabetologia 44(Suppl 2):S14–S21
10. World Health Organization (2011). Global status report on noncommunicable diseases 2010. http://www.who.int/nmh/publications/ncd_report2010/en/. Accessed 15 Aug 2013
11. Roglic G, Unwin N, Bennett PH, Mathers C, Tuomilehto J, Nag S, Connolly V, King H (2005) The burden of mortality attributable to diabetes: realistic estimates for the year 2000. Diabetes Care 28:2130–2135
12. World Health Organization (2011). Prevention of blindness and visual impairment. Action plan for the prevention of avoidable blindness. Global data on visual impairment 2010. http://www.who.int/entity/blindness/GLOBALDATAFINALforweb.pdf. Accessed 15 Aug 2013
13. Guthrie RM (2012) Evolving therapeutic options for type 2 diabetes mellitus: an overview. Postgrad Med 124:82–89
14. US Food and Drug Administration (2008). Guidance for industry. Diabetes mellitus—evaluating cardiovascular risk in new anti-diabetic therapies to treat type 2 diabetes. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guid-ances/ucm071627.pdf. Accessed 15 Aug 2013
15. Nathan DM, Buse JB, Davidson MB, Ferrannini E, Holman RR, Sherwin R, Zinman B (2009) Medical management of hyperglycemia in type 2 diabetes: a consensus algorithm for the initiation and adjustment of therapy: a consensus statement of the American Diabetes Association and the European Association for the Study of Diabetes. Diabetes Care 32:193–203
16. Hopsu-Havu VK, Sarimo SR (1967) Purification and characterization of an aminopeptidase hydrolyzing glycyl-proline-naphthylamide. Hoppe Seylers Z Physiol Chem 348:1540–1550
17. Rawlings ND, Tolle DP, Barrett AJ (2004) MEROPS: the peptidase database. Nucleic Acids Res 32:D160–D164
18. Power O, Nongonierma AB, Jakeman P, Fitzgerald RJ (2013) Food protein hydrolysates as a source of dipeptidyl peptidase IV inhibitory peptides for the management of type 2 diabetes. Proc Nutr Soc 73:34–46
19. Mendieta L, Tarrago T, Giralt E (2011) Recent patents of dipeptidyl peptidase IV inhibitors. Expert Opin Ther Pat 21:1693–1741
20. Gorrell MD (2005) Dipeptidyl peptidase IV and related enzymes in cell biology and liver disorders. Clin Sci (Lond) 108:277–292
21. Juillerat-Jeanneret L (2014) Dipeptidyl peptidase IV and its inhibitors: therapeutics for type 2 diabetes and what else? J Med Chem 57:2197–2212

22. Mentlein R (1999) Dipeptidyl-peptidase IV (CD26)–role in the inactivation of regulatory peptides. Regul Pept 85:9–24
23. Nabeno M, Akahoshi F, Kishida H, Miyaguchi I, Tanaka Y, Ishii S, Kadowaki T (2013) A comparative study of the binding modes of recently launched dipeptidyl peptidase IV inhibitors in the active site. Biochem Biophys Res Commun 434:191–196
24. Thoma R, Löffler B, Stihle M, Huber W, Ruf A, Hennig M (2003) Structural basis of proline-specific exopeptidase activity as observed in human dipeptidyl peptidase-IV. Structure 11:947–959
25. Doherty AM, Bock MG, Desai MC, Overington J, Plattner JJ, Stamford A, Wustrow D, Young H, Gwaltney SL, Stafford JA (2005) Inhibitors of dipeptidyl peptidase 4. Annu Rep Med Chem 40:149–165
26. Chien C-H, Huang L-H, Chou C-Y, Chen Y-S, Han Y-S, Chang G-G, Liang P-H, Chen X (2004) One site mutation disrupts dimer formation in human DPP-IV proteins. J Biol Chem 279:52338–52345
27. Engel M, Hoffmann T, Wagner L, Wermann M, Heiser U, Kiefersauer R, Huber R, Bode W, Demuth H-U, Brandstetter H (2003) The crystal structure of dipeptidyl peptidase IV (CD26) reveals its functional regulation and enzymatic mechanism. Proc Natl Acad Sci U S A 100:5063–5068
28. Pederson RA, White HA, Schlenzig D, Pauly RP, McIntosh CH, Demuth HU (1998) Improved glucose tolerance in Zucker fatty rats by oral administration of the dipeptidyl peptidase IV inhibitor isoleucine thiazolidide. Diabetes 47:1253–1258
29. Pospisilik JA, Stafford SG, Demuth H-U, McIntosh CHS, Pederson RA (2002) Long-term treatment with dipeptidyl peptidase IV inhibitor improves hepatic and peripheral insulin sensitivity in the VDF Zucker rat: a euglycemic-hyperinsulinemic clamp study. Diabetes 51:2677–2683
30. Cheng JD, Dunbrack RL, Valianou M, Rogatko A, Alpaugh RK, Weiner LM (2002) Promotion of tumor growth by murine fibroblast activation protein, a serine protease, in an animal model. Cancer Res 62:4767–4772
31. Kajiyama H, Kikkawa F, Suzuki T, Shibata K, Ino K, Mizutani S (2002) Prolonged survival and decreased invasive activity attributable to dipeptidyl peptidase IV overexpression in ovarian carcinoma. Cancer Res 62:2753–2757
32. Ho L, Aytac U, Stephens LC et al (2001) In vitro and in vivo antitumor effect of the anti-CD26 monoclonal antibody 1F7 on human CD30+ anaplastic large cell T-cell lymphoma Karpas 299. Clin Cancer Res 7:2031–2040
33. Ussher JR, Sutendra G, Jaswal JS (2012) The impact of current and novel anti-diabetic therapies on cardiovascular risk. Future Cardiol 8:895–912
34. Zhong J, Rao X, Rajagopalan S (2013) An emerging role of dipeptidyl peptidase 4 (DPP4) beyond glucose control: potential implications in cardiovascular disease. Atherosclerosis 226:305–314
35. Patil HR, Al Badarin FJ, Al Shami HA, Bhatti SK, Lavie CJ, Bell DSH, O'Keefe JH (2012) Meta-analysis of effect of dipeptidyl peptidase-4 inhibitors on cardiovascular risk in type 2 diabetes mellitus. Am J Cardiol 110:826–833
36. Frederich R, Alexander JH, Fiedorek FT, Donovan M, Berglind N, Harris S, Chen R, Wolf R, Mahaffey KW (2010) A systematic assessment of cardiovascular outcomes in the saxagliptin drug development program for type 2 diabetes. Postgrad Med 122:16–27
37. Scheen AJ (2013) Cardiovascular effects of gliptins. Nat Rev Cardiol 10:73–84
38. Simsek S, de Galan BE (2012) Cardiovascular protective properties of incretin-based therapies in type 2 diabetes. Curr Opin Lipidol 23:540–547
39. Dai Y, Dai D, Mercanti F, Ding Z, Wang X, Mehta JL (2013) Dipeptidyl peptidase-4 inhibitors in cardioprotection: a promising therapeutic approach. Acta Diabetol 50:827–835
40. Scheen AJ (2013) Cardiovascular effects of dipeptidyl peptidase-4 inhibitors: from risk factors to clinical outcomes. Postgrad Med 125:7–20
41. Yousefzadeh P, Wang X (2013) The effects of dipeptidyl peptidase-4 inhibitors on cardiovascular disease risks in type 2 diabetes mellitus. J Diabetes Res 2013:459821

42. Balakumar P, Dhanaraj SA (2013) Cardiovascular pleiotropic actions of DPP-4 inhibitors: a step at the cutting edge in understanding their additional therapeutic potentials. Cell Signal 25:1799–1803
43. Wang XM, Yao T-W, Nadvi NA, Osborne B, McCaughan GW, Gorrell MD (2008) Fibroblast activation protein and chronic liver disease. Front Biosci 13:3168–3180
44. Kirby M, Yu DMT, O'Connor S, Gorrell MD (2010) Inhibitor selectivity in the clinical application of dipeptidyl peptidase-4 inhibition. Clin Sci (Lond) 118:31–41
45. Lankas GR, Leiting B, Roy RS et al (2005) Dipeptidyl peptidase IV inhibition for the treatment of type 2 diabetes: potential importance of selectivity over dipeptidyl peptidases 8 and 9. Diabetes 54:2988–2994
46. Deacon CF, Ahrén B (2011) Physiology of incretins in health and disease. Rev Diabet Stud 8:293–306
47. Tortosa F, Dotta F (2013) Incretin hormones and beta-cell mass expansion: what we know and what is missing? Arch Physiol Biochem 119:161–169
48. Ahrén B (2013) Incretin dysfunction in type 2 diabetes: clinical impact and future perspectives. Diabetes Metab 39:195–201
49. Opinto G, Natalicchio A, Marchetti P (2013) Physiology of incretins and loss of incretin effect in type 2 diabetes and obesity. Arch Physiol Biochem 119:170–178
50. Brunton S (2013) Integrating incretin-based therapy into type 2 diabetes management. Vital Signs 62:S1–S8
51. Papamargaritis D, Miras AD, le Roux CW (2013) Influence of diabetes surgery on gut hormones and incretins. Nutr Hosp 28(Suppl 2):95–103
52. Meier JJ, Nauck MA, Schmidt WE, Gallwitz B (2002) Gastric inhibitory polypeptide: the neglected incretin revisited. Regul Pept 107:1–13
53. Green BD, Flatt PR, Bailey CJ (2006) Inhibition of dipeptidylpeptidase IV activity as a therapy of type 2 diabetes. Expert Opin Emerg Drugs 11:525–539
54. Lindgren O, Mari A, Deacon CF, Carr RD, Winzell MS, Vikman J, Ahrén B (2009) Differential islet and incretin hormone responses in morning versus afternoon after standardized meal in healthy men. J Clin Endocrinol Metab 94:2887–2892
55. Ahrén B, Carr RD, Deacon CF (2010) Incretin hormone secretion over the day. Vitam Horm 84:203–220
56. Zettl H, Schubert-Zsilavecz M, Steinhilber D (2010) Medicinal chemistry of incretin mimetics and DPP-4 inhibitors. ChemMedChem 5:179–185
57. Drucker DJ, Nauck MA (2006) The incretin system: glucagon-like peptide-1 receptor agonists and dipeptidyl peptidase-4 inhibitors in type 2 diabetes. Lancet 368:1696–1705
58. Holst JJ, Vilsbøll T, Deacon CF (2009) The incretin system and its role in type 2 diabetes mellitus. Mol Cell Endocrinol 297:127–136
59. Holst JJ, Deacon CF (2004) Glucagon-like peptide 1 and inhibitors of dipeptidyl peptidase IV in the treatment of type 2 diabetes mellitus. Curr Opin Pharmacol 4:589–596
60. Baggio LL, Drucker DJ (2007) Biology of incretins: GLP-1 and GIP. Gastroenterology 132:2131–2157
61. Drucker DJ (2003) Therapeutic potential of dipeptidyl peptidase IV inhibitors for the treatment of type 2 diabetes. Expert Opin Investig Drugs 12:87–100
62. Højberg PV, Vilsbøll T, Rabøl R, Knop FK, Bache M, Krarup T, Holst JJ, Madsbad S (2009) Four weeks of near-normalisation of blood glucose improves the insulin response to glucagon-like peptide-1 and glucose-dependent insulinotropic polypeptide in patients with type 2 diabetes. Diabetologia 52:199–207
63. Hansen KB, Vilsbøll T, Bagger JI, Holst JJ, Knop FK (2012) Impaired incretin-induced amplification of insulin secretion after glucose homeostatic dysregulation in healthy subjects. J Clin Endocrinol Metab 97:1363–1370
64. Demuth H-U, McIntosh CHS, Pederson RA (2005) Type 2 diabetes–therapy with dipeptidyl peptidase IV inhibitors. Biochim Biophys Acta 1751:33–44
65. Kim S-H, Lee S-H, Yim H-J (2013) Gemigliptin, a novel dipeptidyl peptidase 4 inhibitor: first new anti-diabetic drug in the history of Korean pharmaceutical industry. Arch Pharm Res 36:1185–1188

66. US National Library of Medicine. National Institutes of Health. MedlinePlus (2014). Sitagliptin. http://www.nlm.nih.gov/medlineplus/druginfo/meds/a606023.html. Accessed 21 Nov 2013
67. US National Library of Medicine. National Institutes of Health. MedlinePlus (2014). Saxagliptin. http://www.nlm.nih.gov/medlineplus/druginfo/meds/a610003.html. Accessed 21 Nov 2013
68. US National Library of Medicine. National Institutes of Health. MedlinePlus (2014). Linagliptin. http://www.nlm.nih.gov/medlineplus/druginfo/meds/a611036.html. Accessed 21 Nov 2013
69. Noel RA, Braun DK, Patterson RE, Bloomgren GL (2009) Increased risk of acute pancreatitis and biliary disease observed in patients with type 2 diabetes: a retrospective cohort study. Diabetes Care 32:834–838
70. Engel SS, Williams-Herman DE, Golm GT, Clay RJ, Machotka S V, Kaufman KD, Goldstein BJ (2010) Sitagliptin: review of preclinical and clinical data regarding incidence of pancreatitis. Int J Clin Pract 64:984–990
71. Williams-Herman D, Engel SS, Round E, Johnson J, Golm GT, Guo H, Musser BJ, Davies MJ, Kaufman KD, Goldstein BJ (2010) Safety and tolerability of sitagliptin in clinical studies: a pooled analysis of data from 10,246 patients with type 2 diabetes. BMC Endocr Disord 10:7
72. Engel SS, Round E, Golm GT, Kaufman KD, Goldstein BJ (2013) Safety and tolerability of sitagliptin in type 2 diabetes: pooled analysis of 25 clinical studies. Diabetes Ther 4:119–145
73. Monami M, Dicembrini I, Mannucci E (2014) Dipeptidyl peptidase-4 inhibitors and pancreatitis risk: a meta-analysis of randomized clinical trials. Diabetes Obes Metab 16:48–56
74. Scheen A (2013) Gliptins (dipeptidyl peptidase-4 inhibitors) and risk of acute pancreatitis. Expert Opin Drug Saf 12:545–557
75. Deacon CF, Holst JJ (2013) Dipeptidyl peptidase-4 inhibitors for the treatment of type 2 diabetes: comparison, efficacy and safety. Expert Opin Pharmacother 14:2047–2058
76. Zanchi A, Lehmann R, Philippe J (2012) Anti-diabetic drugs and kidney disease–recommendations of the Swiss Society for Endocrinology and Diabetology. Swiss Med Wkly 142:w13629
77. Ramirez G, Morrison AD, Bittle PA (2013) Clinical practice considerations and review of the literature for the use of DPP-4 inhibitors in patients with type 2 diabetes and chronic kidney disease. Endocr Pract 19:1025–1034
78. Kuhn B, Hennig M, Mattei P (2007) Molecular recognition of ligands in dipeptidyl peptidase IV. Curr Top Med Chem 7:609–619
79. Engel M, Hoffmann T, Manhart S, Heiser U, Chambre S, Huber R, Demuth H-U, Bode W (2006) Rigidity and flexibility of dipeptidyl peptidase IV: crystal structures of and docking experiments with DPIV. J Mol Biol 355:768–783
80. Li C, Shen J, Li W, Lu C (2011) Possible ligand release pathway of dipeptidyl peptidase IV investigated by molecular dynamics simulations. Proteins Struct Funct Bioinforma 79:1800–1809
81. Schechter I, Berger A (2012) On the size of the active site in proteases. I. Papain. 1967. Biochem Biophys Res Commun 425:497–502
82. Weber AE (2004) Dipeptidyl peptidase IV inhibitors for the treatment of diabetes. J Med Chem 47:4135–4141
83. Wallace MB, Feng J, Zhang Z, Skene RJ, Shi L, Caster CL, Kassel DB, Xu R, Gwaltney SL (2008) Structure-based design and synthesis of benzimidazole derivatives as dipeptidyl peptidase IV inhibitors. Bioorg Med Chem Lett 18:2362–2367
84. Patel B, Ghate M (2013) Computational studies on structurally diverse dipeptidyl peptidase IV inhibitors: an approach for new anti-diabetic drug development. Med Chem Res 22:4505–4521
85. Al-Masri IM, Mohammad MK, Taha MO (2008) Discovery of DPP IV inhibitors by pharmacophore modeling and QSAR analysis followed by in silico screening. ChemMedChem 3:1763–1779
86. Aertgeerts K, Ye S, Tennant MG, Kraus ML, Rogers J, Sang B-C, Skene RJ, Webb DR, Prasad GS (2004) Crystal structure of human dipeptidyl peptidase IV in complex with a decapeptide reveals details on substrate specificity and tetrahedral intermediate formation. Protein Sci 13:412–421
87. Bjelke JR, Christensen J, Branner S, Wagtmann N, Olsen C, Kanstrup AB, Rasmussen HB (2004) Tyrosine 547 constitutes an essential part of the catalytic mechanism of dipeptidyl peptidase IV. J Biol Chem 279:34691–34697

88. Yoshida T, Akahoshi F, Sakashita H, et al (2012) Discovery and preclinical profile of teneligliptin (3-[(2S,4S)-4-[4-(3-methyl-1-phenyl-1H-pyrazol-5-yl)piperazin-1-yl]pyrrolidin-2-ylcarbonyl]thiazolidine): a highly potent, selective, long-lasting and orally active dipeptidyl peptidase IV inhibitor for t. Bioorg Med Chem 20:5705–5719
89. Yoshida T, Akahoshi F, Sakashita H, Sonda S, Takeuchi M, Tanaka Y, Nabeno M, Kishida H, Miyaguchi I, Hayashi Y (2012) Fused bicyclic heteroarylpiperazine-substituted L-prolylthiazolidines as highly potent DPP-4 inhibitors lacking the electrophilic nitrile group. Bioorg Med Chem 20:5033–5041
90. Edmondson SD, Mastracchio A, Cox JM et al (2009) Aminopiperidine-fused imidazoles as dipeptidyl peptidase-IV inhibitors. Bioorg Med Chem Lett 19:4097–4101
91. Edmondson SD, Mastracchio A, Mathvink RJ et al (2006) (2S,3S)-3-Amino-4-(3,3-difluoropyrrolidin-1-yl)-N, N-dimethyl-4-oxo-2-(4-[1,2,4]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 49:3614–3627
92. Edmondson SD, Wei L, Xu J et al (2008) Fluoroolefins as amide bond mimics in dipeptidyl peptidase IV inhibitors. Bioorg Med Chem Lett 18:2409–2413
93. Biftu T, Scapin G, Singh S et al (2007) Rational design of a novel, potent, and orally bioavailable cyclohexylamine DPP-4 inhibitor by application of molecular modeling and X-ray crystallography of sitagliptin. Bioorg Med Chem Lett 17:3384–3387
94. Eckhardt M, Langkopf E, Mark M et al (2007) 8-(3-®-aminopiperidin-1-yl)-7-but-2-ynyl-3-methyl-1-(4-methyl-quinazolin-2-ylmethyl)-3,7-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 50:6450–6453
95. Kaelin DE, Smenton AL, Eiermann GJ et al (2007) 4-arylcyclohexylalanine analogs as potent, selective, and orally active inhibitors of dipeptidyl peptidase IV. Bioorg Med Chem Lett 17:5806–5811
96. Nordhoff S, Cerezo-Gálvez S, Deppe H, Hill O, López-Canet M, Rummey C, Thiemann M, Matassa VG, Edwards PJ, Feurer A (2009) Discovery of beta-homophenylalanine based pyrrolidin-2-ylmethyl amides and sulfonamides as highly potent and selective inhibitors of dipeptidyl peptidase IV. Bioorg Med Chem Lett 19:4201–4203
97. Salam NK, Nuti R, Sherman W (2009) Novel method for generating structure-based pharmacophores using energetic analysis. J Chem Inf Model 49:2356–2368
98. Loving K, Salam NK, Sherman W (2009) Energetic analysis of fragment docking and application to structure-based pharmacophore hypothesis generation. J Comput Aided Mol Des 23:541–554
99. Guasch L, Ojeda MJ, González-Abuín N et al (2012) Identification of novel human dipeptidyl peptidase-IV inhibitors of natural origin (part I): virtual screening and activity assays. PLoS One 7:e44971
100. Rummey C, Metz G (2007) Homology models of dipeptidyl peptidases 8 and 9 with a focus on loop predictions near the active site. Proteins 66:160–171
101. Janardhan S, Reddy YP (2011) Homology modeling and molecular docking studies of human DPP8 and DPP9. Int J Pharma Res Dev 2:131–146
102. Pitman MR, Menz RI, Abbott CA (2006) Prediction of dipeptidyl peptidase (DP) 8 structure by homology modelling. Adv Exp Med Biol 575:33–42
103. Tanwar O, Deora GS, Tanwar L, Kumar G, Janardhan S, Alam MM, Shaquiquzzaman M, Akhter M (2014) Novel hydrazine derivatives as selective DPP-IV inhibitors: findings from virtual screening and validation through molecular dynamics simulations. J Mol Model 20:2118
104. Kang NS, Ahn JH, Kim SS, Chae CH, Yoo S-E (2007) Docking-based 3D-QSAR study for selectivity of DPP4, DPP8, and DPP9 inhibitors. Bioorg Med Chem Lett 17:3716–3721
105. Patel BD, Ghate MD (2014) Recent approaches to medicinal chemistry and therapeutic potential of dipeptidyl peptidase-4 (DPP-4) inhibitors. Eur J Med Chem 74:574–605
106. Ghate M, Jain SV (2013) Structure based lead optimization approach in discovery of selective DPP4 inhibitors. Mini Rev Med Chem 13:888–914

107. Fukuda-Tsuru S, Anabuki J, Abe Y, Yoshida K, Ishii S (2012) A novel, potent, and long-lasting dipeptidyl peptidase-4 inhibitor, teneligliptin, improves postprandial hyperglycemia and dyslipidemia after single and repeated administrations. Eur J Pharmacol 696:194–202
108. Ghate M, Jain S (2014) Fragment based HQSAR modeling and docking analysis of conformationally rigid 3-azabicyclo hexane derivatives to design selective DPP-4 inhibitors. Lett Drug Des Discov 11:184–198
109. American Diabetes Association (2014) Standards of medical care in diabetes–2014. Diabetes Care 37(Suppl 1):S14–S80
110. Rollinger JM, Stuppner H, Langer T (2008) Virtual screening for the discovery of bioactive natural products. Prog drug Res 65:211, 213–249
111. Schuster D, Wolber G (2010) Identification of bioactive natural products by pharmacophore-based virtual screening. Curr Pharm Des 16:1666–1681
112. Martinez-Mayorga K, Medina-Franco JL (2009) Chemoinformatics-applications in food chemistry. Adv Food Nutr Res 58:33–56
113. Ferguson LLR (2009) Nutrigenomics approaches to functional foods. J Am Diet Assoc 109:452–458
114. Pascual I, Lopéz A, Gómez H, Chappé M, Saroyán A, González Y, Cisneros M, Charli JL, Chávez MDLA (2007) Screening of inhibitors of porcine dipeptidyl peptidase IV activity in aqueous extracts from marine organisms. Enzyme Microb Technol 40:414–419
115. 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 24:1061–1066
116. Hamden K, Bengara A, Amri Z, Elfeki A (2013) Experimental diabetes treated with trigonelline: effect on key enzymes related to diabetes and hypertension, β-cell and liver function. Mol Cell Biochem 381:85–94
117. Antonyan A, De A, Vitali L, Pettinari R, Marchetti F, Gigliobianco MR, Pettinari C, Camaioni E, Lupidi G (2014) Evaluation of (arene)Ru(II) complexes of curcumin as inhibitors of dipeptidyl peptidase IV. Biochimie 99:146–152
118. González-Abuín N, Martínez-Micaelo N, Blay M, Pujadas G, Garcia-Vallvé S, Pinent M, Ardévol A (2012) Grape seed-derived procyanidins decrease dipeptidyl-peptidase 4 activity and expression. J Agric Food Chem 60:9055–9061
119. Zhang S, Lu W, Liu X, Diao Y, Bai F, Wang L, Shan L, Huang J, Li H, Zhang W (2011) Fast and effective identification of the bioactive compounds and their targets from medicinal plants via computational chemical biology approach. MedChemComm 2:471
120. Guasch L, Sala E, Ojeda MJ, Valls C, Bladé C, Mulero M, Blay M, Ardévol A, Garcia-Vallvé S, Pujadas G (2012) Identification of novel human dipeptidyl peptidase-IV inhibitors of natural origin (part II): in silico prediction in anti-diabetic extracts. PLoS One 7:e44972
121. Fan J, Johnson MH, Lila MA, Yousef G, de Mejia EG (2013) Berry and citrus phenolic compounds inhibit dipeptidyl peptidase IV: implications in diabetes management. Evid Based Complement Alternat Med 2013:479505
122. Parmar HS, Jain P, Chauhan DS et al (2012) DPP-IV inhibitory potential of naringin: an in silico, in vitro and in vivo study. Diabetes Res Clin Pract 97:105–111
123. Geng Y, Lu Z-M, Huang W, Xu H-Y, Shi J-S, Xu Z-H (2013) Bioassay-guided isolation of DPP-4 inhibitory fractions from extracts of submerged cultured of Inonotus obliquus. Molecules 18:1150–1161
124. Bharti SK, Krishnan S, Kumar A, Rajak KK, Murari K, Bharti BK, Gupta AK (2012) Antihyperglycemic activity with DPP-IV inhibition of alkaloids from seed extract of *Castanospermum australe*: investigation by experimental validation and molecular docking. Phytomedicine 20:24–31
125. Bellé LP, Bitencourt PER, Abdalla FH, Bona KS de, Peres A, Maders LDK, Moretto MB (2013) Aqueous seed extract of *Syzygium cumini* inhibits the dipeptidyl peptidase IV and adenosine deaminase activities, but it does not change the CD26 expression in lymphocytes in vitro. J Physiol Biochem 69:119–124

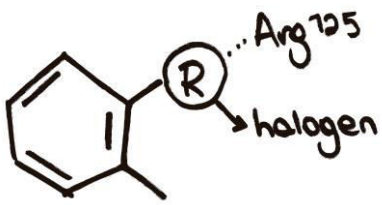
126. Lacroix IME, Li-Chan ECY (2012) Evaluation of the potential of dietary proteins as precursors of dipeptidyl peptidase (DPP)-IV inhibitors by an in silico approach. J Funct Foods 4:403–422
127. Nongonierma AB, Fitzgerald RJ (2014) Susceptibility of milk protein-derived peptides to dipeptidyl peptidase IV (DPP-IV) hydrolysis. Food Chem 145:845–852
128. Rahfeld J, Schierhorn M, Hartrodt B, Neubert K, Heins J (1991) Are diprotin A (Ile-Pro-Ile) and diprotin B (Val-Pro-Leu) inhibitors or substrates of dipeptidyl peptidase IV? Biochim Biophys Acta 1076:314–316
129. Tulipano G, Sibilia V, Caroli AM, Cocchi D (2011) Whey proteins as source of dipeptidyl dipeptidase IV (dipeptidyl peptidase-4) inhibitors. Peptides 32:835–838
130. Nongonierma AB, FitzGerald RJ (2013) Dipeptidyl peptidase IV inhibitory and antioxidative properties of milk protein-derived dipeptides and hydrolysates. Peptides 39:157–163
131. Silveira ST, Martínez-Maqueda D, Recio I, Hernández-Ledesma B (2013) Dipeptidyl peptidase-IV inhibitory peptides generated by tryptic hydrolysis of a whey protein concentrate rich in β-lactoglobulin. Food Chem 141:1072–1077
132. Hatanaka T, Inoue Y, Arima J, Kumagai Y, Usuki H, Kawakami K, Kimura M, Mukaihara T (2012) Production of dipeptidyl peptidase IV inhibitory peptides from defatted rice bran. Food Chem 134:797–802
133. Huang S-L, Jao C-L, Ho K-P, Hsu K-C (2012) Dipeptidyl-peptidase IV inhibitory activity of peptides derived from tuna cooking juice hydrolysates. Peptides 35:114–121
134. Gallego M, Aristoy M-C, Toldrá F (2013) Dipeptidyl peptidase IV inhibitory peptides generated in Spanish dry-cured ham. Meat Sci 96:757–761
135. Lacroix IME, Li-Chan ECY (2012) Dipeptidyl peptidase-IV inhibitory activity of dairy protein hydrolysates. Int Dairy J 25:97–102
136. Velarde-Salcedo AJ, Barrera-Pacheco A, Lara-González S, Montero-Morán GM, Díaz-Gois A, González de Mejia E, Barba de la Rosa AP (2013) In vitro inhibition of dipeptidyl peptidase IV by peptides derived from the hydrolysis of amaranth (*Amaranthus hypochondriacus* L.) proteins. Food Chem 136:758–764
137. Nongonierma AB, Mooney C, Shields DC, Fitzgerald RJ (2013) Inhibition of dipeptidyl peptidase IV and xanthine oxidase by amino acids and dipeptides. Food Chem 141:644–653
138. Li-Chan ECY, Hunag S-L, Jao C-L, Ho K-P, Hsu K-C (2012) Peptides derived from atlantic salmon skin gelatin as dipeptidyl-peptidase IV inhibitors. J Agric Food Chem 60:973–978
139. Uenishi H, Kabuki T, Seto Y, Serizawa A, Nakajima H (2012) Isolation and identification of casein-derived dipeptidyl-peptidase 4 (DPP-4)-inhibitory peptide LPQNIPPL from gouda-type cheese and its effect on plasma glucose in rats. Int Dairy J 22:24–30
140. Uchida M, Ohshiba Y, Mogami O (2011) Novel dipeptidyl peptidase-4-inhibiting peptide derived from β-lactoglobulin. J Pharmacol Sci 117:63–66
141. Dziuba M, Dziuba B, Iwaniak A (2009) Milk proteins as precursors of bioactive peptides. Acta Sci Pol Technol Aliment 8(1):71–90 (http://www.food.actapol.net/volume8/issue1/abstract-7.html)
142. Minkiewicz P, Dziuba J, Michalska J (2011) Bovine meat proteins as potential precursors of biologically active peptides—a computational study based on the BIOPEP database. Food Sci Technol Int 17:39–45
143. Abe M, Akiyama T, Umezawa Y, Yamamoto K, Nagai M, Yamazaki H, Ichikawa Y-I, Muraoka Y (2005) Synthesis and biological activity of sulphostin analogues, novel dipeptidyl peptidase IV inhibitors. Bioorg Med Chem 13:785–797
144. Akiyama T, Abe M, Harada S et al (2001) Sulphostin, a potent inhibitor for dipeptidyl peptidase IV from Streptomyces sp. MK251–43F3. J Antibiot (Tokyo) 54:744–746
145. Umezawa H, Aoyagi T, Ogawa K, Naganawa H, Hamada M, Takeuchi T (1984) Diprotins A and B, inhibitors of dipeptidyl aminopeptidase IV, produced by bacteria. J Antibiot (Tokyo) 37:422–425
146. Trellet M, Melquiond A, Bonvin A (2013) A unified conformational selection and induced fit approach to protein-peptide docking. PLoS One 8:e58769
147. Albericio F, Kruger HG (2012) Therapeutic peptides. Future Med Chem 4:1527–1531

148. Yan TR, Ho SC, Hou CL (1992) Catalytic properties of X-prolyl dipeptidyl aminopeptidase from *Lactococcus lactis* subsp. cremoris nTR. Biosci Biotechnol Biochem 56:704–707
149. Lorey S, Stöckel-Maschek A, Faust J et al (2003) Different modes of dipeptidyl peptidase IV (CD26) inhibition by oligopeptides derived from the N-terminus of HIV-1 Tat indicate at least two inhibitor binding sites. Eur J Biochem 270:2147–2156
150. Valli M, dos Santos RN, Figueira LD, Nakajima CH, Castro-Gamboa I, Andricopulo AD, Bolzani VS (2013) Development of a natural products database from the biodiversity of Brazil. J Nat Prod 76:439–444
151. Chen CY-C (2011) TCM database@Taiwan: the world's largest traditional Chinese medicine database for drug screening in silico. PLoS One 6:e15939
152. Elsevier Reaxys chemistry workflow solution. http://www.reaxys.com. Accessed 20 Jan 2014
153. Parasuraman S (2012) Protein data bank. J Pharmacol Pharmacother 3:351–352
154. Irwin JJ, Sterling T, Mysinger MM, Bolstad ES, Coleman RG (2012) ZINC: a free tool to discover chemistry for biology. J Chem Inf Model 52:1757–1768
155. Black OF, Kelly JW (1927) Pseudo ephedrine from *Ephedra alata*. Am J Pharm 99:748–751
156. 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 44:389–394
157. 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 51:162–163
158. Grue-Sorensen G, Spenser ID (1989) The biosynthesis of ephedrine. Can J Chem 67:998–1009
159. Ito K, Haruna M, Furukawa H (1975) Studies on the erythrina alkaloids. X. Alkaloids of several Erythrina plants from Singapore (author's transl). Yakugaku Zasshi 95:358–362
160. Kumar A, Lingadurai S, Shrivastava TP, Bhattacharya S, Haldar PK (2011) Hypoglycemic activity of *Erythrina variegata* leaf in streptozotocin-induced diabetic rats. Pharm Biol 49:577–582
161. Oh WK, Lee C-H, Seo JH, Chung MY, Cui L, Fomum ZT, Kang JS, Lee HS (2009) Diacylglycerol acyltransferase-inhibitory compounds from *Erythrina senegalensis*. Arch Pharm Res 32:43–47
162. Na M, Jang J, Njamen D, Mbafor JT, Fomum ZT, Kim BY, Oh WK, Ahn JS (2006) Protein tyrosine phosphatase-1B inhibitory activity of isoprenylated flavonoids isolated from *Erythrina mildbraedii*. J Nat Prod 69:1572–1576
163. Bae EY, Na M, Njamen D, Mbafor JT, Fomum ZT, Cui L, Choung DH, Kim BY, Oh WK, Ahn JS (2006) Inhibition of protein tyrosine phosphatase 1B by prenylated isoflavonoids isolated from the stem bark of *Erythrina addisoniae*. Planta Med 72:945–948
164. Benn MH, Shustov G, Shustova L, Majak W, Bai Y, Fairey NA (1996) Isolation and characterization of two guanidines from *Galega orientalis* Lam. Cv. Gale (fodder galega). J Agric Food Chem 44:2779–2781
165. Vuksan V, Sievenpiper JL (2005) Herbal remedies in the management of diabetes: lessons learned from the study of ginseng. Nutr Metab Cardiovasc Dis 15:149–160
166. Michel KH, Sandberg F, Haglid F, Norin T (1967) Alkaloids of *Haloxylon salicornicum* (Moq.-Tand.) Boiss. Acta Pharm Suec 4:97–116
167. Brack A (1962) Verlauf der Alkaloidbildung durch den Clavicepsstamm von Pennisetum typhoideum Rich. in saprophytischer Kultur. 54. Mitteilung über Mutterkornalkaloide. Arch Pharm (Weinheim) 295:510–515
168. Shukla K, Narain JP, Puri P, Gupta A, Bijlani RL, Mahapatra SC, Karmarkar MG (1991) Glycaemic response to maize, bajra and barley. Indian J Physiol Pharmacol 35:249–254
169. Sheludko Y, Gerasimenko I, Kolshorn H, Stöckigt J (2002) New alkaloids of the sarpagine group from *Rauvolfia serpentina* hairy root culture. J Nat Prod 65:1006–1010

170. Benzi G, Villa RF, Dossena M, Vercesi L, Gorini A, Pastoris O (1984) Cerebral and cerebellar metabolic changes induced by drugs during the recovery period after profound hypoglycemia. Farmaco Sci 39:44–56
171. Ronchetti F, Russo G, Bombardelli E, Bonati A (1971) A new alkaloid from *Rauwolfia vomitoria*. Phytochemistry 10:1385–1388
172. Campbell JIA, Mortensen A, Mølgaard P (2006) Tissue lipid lowering-effect of a traditional Nigerian anti-diabetic infusion of *Rauwolfia vomitoria* foilage and *Citrus aurantium* fruit. J Ethnopharmacol 104:379–386
173. 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 (Tokyo) 54:546–549
174. Latha M, Pari L, Sitasawad S, Bhonde R (2004) *Scoparia dulcis*, a traditional anti-diabetic plant, protects against streptozotocin induced oxidative stress and apoptosis in vitro and in vivo. J Biochem Mol Toxicol 18:261–272
175. Ly TT, Hewitt J, Davey RJ, Lim EM, Davis EA, Jones TW (2011) Improving epinephrine responses in hypoglycemia unawareness with real-time continuous glucose monitoring in adolescents with type 1 diabetes. Diabetes Care 34:50–52
176. Andrews KM, Beebe D a, Benbow JW et al (2011) 1-((3S,4S)-4-amino-1-(4-substituted-1,3,5-triazin-2-yl) pyrrolidin-3-yl)-5,5-difluoropiperidin-2-one inhibitors of DPP-4 for the treatment of type 2 diabetes. Bioorg Med Chem Lett 21:1810–1814
177. Aguilar-Santamaría L, Ramírez G, Nicasio P, Alegría-Reyes C, Herrera-Arellano A (2009) Anti-diabetic activities of *Tecoma stans* (L.) Juss. ex Kunth. J Ethnopharmacol 124:284–288
178. Hammouda Y, Rashid A-K, Amer MS (1964) Hypoglycaemic properties of tecomine and tecostanine. J Pharm Pharmacol 16:833–834
179. Van de Venter M, Roux S, Bungu LC et al (2008) Anti-diabetic screening and scoring of 11 plants traditionally used in South Africa. J Ethnopharmacol 119:81–86
180. Torres JL, Bobet R (2001) New flavanol derivatives from grape (*Vitis vinifera*) byproducts. Antioxidant aminoethylthio–flavan-3-ol conjugates from a polymeric waste fraction used as a source of flavanols. J Agric Food Chem 49:4627–4634
181. Pinent M, Blay M, Bladé MC, Salvadó MJ, Arola L, Ardévol A (2004) Grape seed-derived procyanidins have an antihyperglycemic effect in streptozotocin-induced diabetic rats and insulinomimetic activity in insulin-sensitive cell lines. Endocrinology 145:4985–4990
182. Song E-K, Hur H, Han M-K (2003) Epigallocatechin gallate prevents autoimmune diabetes induced by multiple low doses of streptozotocin in mice. Arch Pharm Res 26:559–563
183. 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
184. Konno C, Murayama M, Sugiyama K, Arai M, Murakami M, Takahashi M, Hikino H (1985) Isolation and hypoglycemic activity of aconitans A, B, C and D, glycans of *Aconitum carmichaeli* roots. Planta Med 51:160–161
185. Howes M, Simmonds M (2005) Plants used in the treatment of diabetes. In: Soumyanath A (ed) Traditional medicines for modern times. CRC, Boca Raton.
186. Zhang H, Wang X-N, Lin L-P, Ding J, Yue J-M (2007) Indole alkaloids from three species of the Ervatamia genus: *E. officinalis*, *E. divaricata*, and *E. divaricata* Gouyahua. J Nat Prod 70:54–59
187. Fujii M, Takei I, Umezawa K (2009) Anti-diabetic effect of orally administered conophylline-containing plant extract on streptozotocin-treated and Goto-Kakizaki rats. Biomed Pharmacother 63:710–716
188. Usubillaga A (1988) Solanudine, a steroidal alkaloid from *Solanum nudum*. Phytochemistry 27:3031–3032
189. Yoshikawa M, Nakamura S, Ozaki K, Kumahara A, Morikawa T, Matsuda H (2007) Structures of steroidal alkaloid oligoglycosides, robeneosides A and B, and antidiabetogenic constituents from the Brazilian medicinal plant *Solanum lycocarpum*. J Nat Prod 70:210–214

190. Villaseñor IM, Lamadrid MRA (2006) Comparative anti-hyperglycemic potentials of medicinal plants. J Ethnopharmacol 104:129–131
191. El Sayed KA, Hamann MT, Abd El-Rahman HA, Zaghloul AM (1998) New pyrrole alkaloids from *Solanum sodomaeum*. J Nat Prod 61:848–850
192. Kar DM, Maharana L, Pattnaik S, Dash GK (2006) Studies on hypoglycaemic activity of *Solanum xanthocarpum* Schrad. & Wendl. fruit extract in rats. J Ethnopharmacol 108:251–256
193. 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
194. 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
195. 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 12:99–105
196. 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 26:313–317
197. Beek TAV, Verpoorte R, Svendsen AB (1984) Alkaloids of *Tabernaemontana eglandulosa*. Tetrahedron 40(4):737
198. Ma D-L, Chan DS-H, Leung C-H (2013) Drug repositioning by structure-based virtual screening. Chem Soc Rev 42:2130–2141
199. Meslamani J, Bhajun R, Martz F, Rognan D (2013) Computational profiling of bioactive compounds using a target-dependent composite workflow. J Chem Inf Model 53:2322–2333 doi:10.1021/ci400303n
200. Peng S, Lin X, Guo Z, Huang N (2012) Identifying multiple-target ligands via computational chemogenomics approaches. Curr Top Med Chem 12:1363–1375
201. Swamidass SJ, Lu Z, Agarwal P, Butte AJ (2014) Computational approaches to drug repurposing and pharmacolog- session introduction. Pac Symp Biocomput 19:110–113
202. Peters J-U (2013) Polypharmacology—foe or friend? J Med Chem 56:8955–8971
203. Santiago DN, Pevzner Y, Durand AA, Tran M, Scheerer RR, Daniel K, Sung S-S, Woodcock HL, Guida WC, Brooks WH (2012) Virtual target screening: validation using kinase inhibitors. J Chem Inf Model 52:2192–2203
204. Yue R, Shan L, Yang X, Zhang W (2012) Approaches to target profiling of natural products. Curr Med Chem 19:3841–3855
205. Gao Z, Li H, Zhang H, Liu X, Kang L, Luo X, Zhu W, Chen K, Wang X, Jiang H (2008) PDTD: a web-accessible protein database for drug target identification. BMC Bioinformatics 9:104
206. Li H, Gao Z, Kang L et al (2006) TarFisDock: a web server for identifying drug targets with docking approach. Nucleic Acids Res 34:W219–W224
207. Laskowski RA, Swindells MB (2011) LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. J Chem Inf Model 51:2778–2786
208. Sayle RA, Milner-White EJ (1995) RASMOL: biomolecular graphics for all. Trends Biochem Sci 20:374

Hypothesis and Objectives



Metabolic syndrome refers to a group of metabolic disorders which raises the risk for cardiovascular diseases and insulin resistance. Diabetes has become an important public health problem owing to the increasing prevalence over the past few decades. Despite the many available drugs, there is still a need for new therapies to control glycemia aimed to reduce undesirable side effects, such as hypoglycemia and weight gain. Consequently, one of the main challenges is improving the efficiency of the process of drug discovery and design to find new bioactive compounds. In recent years computational approaches have significantly contributed to the discovery, design and optimization of new bioactive molecules. Thus, *in silico* methodologies can potentially reduce the cost and time, only focusing the experimental efforts on the most promising small molecules. Because of that, virtual screening workflows has become an indispensable tool for the therapeutic development.

The research work of this thesis has been carried out in the Cheminformatics and Nutrition research group (Universitat Rovira i Virgili), which has a wide expertise in the identification of new small molecules for the prevention or reduction of metabolic risk factors associated with the metabolic syndrome (*e.g.*, obesity, inflammation and diabetes). In particular, this thesis focuses on an exhaustive analysis of the binding site of dipeptidyl peptidase IV (DPP-IV) for identifying new lead-compounds as antidiabetic drugs to develop a successful therapeutic approach for the treatment of type II diabetes mellitus.

Accordingly, the hypothesis of this doctoral thesis is as follows:

Potent and selective DPP-IV inhibitors can be identified by computational methodologies.

Virtual screening workflows can be tailored to search for new DPP-IV inhibitors with very low (or no) similarity to existing actives in databases of purchasable compounds

At least part of the beneficial effects on cardiovascular diseases of DPP-IV inhibitors are the result of their interaction with targets different than DPP-IV.

In order to assess the established hypothesis, the following specific objectives were considered:

- ✗ To suggest how virtual screening protocols might be improved to favor the identification of potent and selective DPP-IV inhibitors based on the most important activity and selectivity cliffs for DPP-IV ***(Manuscript 1)***.

- ✗ To design a computational strategy to search for new lead molecules with very low (or no) similarity to existing actives in databases of purchasable compounds and apply it to a target of pharmacological interest, such as DPP-IV ***(Manuscript 2)***.

- ✗ To demonstrate that, at least partly, the described antidiabetic effect of different *Ephedra* species extracts is the result of the DPP-IV inhibitory bioactivity of ephedrine and the ephedrine-derivatives found in these extracts and to suggest how one of these alkaloids can be used as lead compound for developing new DPP-IV inhibitors with better potency and selectivity ***(Manuscript 3)***.

- ✗ To analyze the physico-chemical features shared by the DPP-IV and β_2-adrenergic receptors binding sites and compare them in order to evaluate if small molecules with dual bioactivity as DPP-IV inhibitors and β-blockers are possible ***(Manuscript 4)***.

La síndrome metabòlica fa referència a un grup de desordres metabòlics que incrementen el risc de patir malalties cardiovasculars i resistència a la insulina. La diabetis ha esdevingut durant les últimes dècades un important problema de salut pública degut a l'alta prevalença. Malgrat la disponibilitat de diversos fàrmacs, encara hi ha la necessitat de cercar noves teràpies destinades al control de la glicèmia per tal de reduir els efectes secundaris, com la hipoglucèmia o l'augment de pes. En conseqüència, un dels principals reptes és la millora de l'eficiència del procés de descobriment i disseny de fàrmacs per trobar nous compostos bioactius. En els últims anys, les tècniques computacionals han contribuït significativament al descobriment, disseny i optimització de noves molècules bioactives. D'aquesta forma, les metodologies *in silico* poden reduir potencialment el cost econòmic i temporal, atès que només s'empren els esforços experimentals en aquells lligands més prometedors. Per aquesta raó, els cribratges virtuals han esdevingut una eina indispensable per al desenvolupament farmacològic.

El treball d'investigació d'aquesta tesi s'ha dut a terme en el grup de Quimioinformàtica i Nutrició (Universitat Rovira i Virgili) que té una àmplia experiència en la identificació de molècules per a la prevenció o la reducció dels riscos metabòlics associats amb la síndrome metabòlica (per exemple, inflamació, obesitat i diabetis). En particular, aquesta tesis està centrada en una anàlisi exhaustiva del lloc d'unió de l'enzim dipeptidil peptidasa IV (DPP-IV) per a la identificació de nous compostos de partida que puguin esdevenir nous fàrmacs antidiabètics dirigits al tractament de la diabetis mellitus tipus II.

Per tant, les hipòtesis d'aquesta tesis doctoral són:

Inhibidors potents i selectius de DPP-IV poden ser identificats per tècniques computacionals.

Els cribratge virtuals poden ser adaptats per cercar nous inhibidors de DPP-IV en bases de dades de compostos comercials que presentin baixa (o nul·la) similitud amb els actius existents.

Almenys part dels efectes beneficiosos dels inhibidors de DPP-IV sobre les malalties cardiovasculars són deguts a la seva interacció amb dianes diferents a DPP-IV.

Amb la finalitat d'avaluar les hipòtesis establertes, es van considerar els següents objectius:

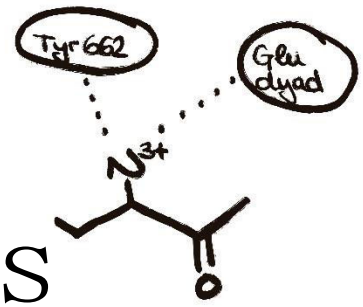
- ✗ Suggerir com els protocols de cribratges virtuals poden ser millorats per afavorir la identificació d'inhibidors en base a les interaccions més rellevants per a potenciar activitat i selectivitat en DPP-IV ***(Manuscrit 1)***.

- ✗ Dissenyar una estratègia computacional per cercar noves molècules de partida en bases de dades de compostos comercials que presentin baixa (o nul·la) similitud amb els actius existents ***(Manuscrit 2)***.

- ✗ Demostrar que, almenys parcialment, l'efecte antidiabètic descrit per als extractes de diferents espècies d'*Ephedra* és el resultat de l'activitat inhibitòria de DPP-IV per part dels compostos d'efedrina i derivats d'aquesta trobats en aquests mateixos extractes i suggerir com un d'aquests alcaloides pot ser utilitzat com a compost de partida per a desenvolupar nous inhibidors de DPP-IV amb millor potència i selectivitat ***(Manuscrit 3)***.

- ✗ Analitzar les característiques fisico-químiques compartides pels llocs d'unió de DPP-IV i del receptor adrenèrgic β_2 i comparar-los amb l'objectiu d'avaluar si és possible que un lligand pugui presentar activitat dual com a inhibidor de DPP-IV i β-bloquejant ***(Manuscrit 4)***.

Results



Manuscript 1

Activity and selectivity cliffs for DPP-IV inhibitors: Lessons we can learn from SAR studies and their application to virtual screening

María José Ojeda-Montes[a,1], Aleix Gimeno[a,1], Sarah Tomás-Hernández[a],, Adrià Cereto-Massagué[a], Raúl Beltrán-Debón[a], Cristina Valls[a], Miquel Mulero[a], Gerard Pujadas[a,b,*] and Santiago Garcia-Vallvé[a,b]

[1]Both authors contributed equally to this work

[a]Research group in Cheminformatics & Nutrition, Departament de Bioquímica i Biotecnologia, Universitat Rovira i Virgili, Tarragona, Catalonia, Spain

[b]EURECAT, TECNIO, CEICS, Reus, Catalonia, Spain

**Correspondence to:*
Gerard Pujadas, Research group in Cheminformatics & Nutrition, phone: +34 977 55 95 65, fax: +34 977 55 82 32. Departament de Bioquímica i Biotecnologia, Facultat de Química, Universitat Rovira i Virgili, C/ Marcel·lí Domingo 1, Edifici N4, 43007 Tarragona, Catalonia, Spain. E-mail: gerard.pujadas@urv.cat

Abstract

The inhibition of dipeptidyl peptidase-IV (DPP-IV) has emerged over the last decade as one of the most effective treatments for type 2 diabetes mellitus, and consequently **(a)** 11 DPP-IV inhibitors have been on the market since 2006 (3 in 2015), and **(b)** 73 non-covalent complexes involving human DPP-IV and drug-like inhibitors are available at the Protein Data Bank. The present review aims to **(a)** explain the most important activity cliffs for DPP-IV non-covalent inhibition according to the binding site structure of DPP-IV, **(b)** explain the most important selectivity cliffs for DPP-IV non-covalent inhibition in comparison with other related enzymes (*i.e.*, DPP8 and DPP9), and **(c)** use the information deriving from this activity/selectivity cliff analysis to suggest how virtual screening protocols might be improved to favor the early identification of potent and selective DPP-IV inhibitors in molecular databases (because they have not succeeded in identifying selective DPP-IV inhibitors with $IC_{50} \leq 100$ nM). All these goals are achieved with the help of available homology models for DPP8 and DPP9 and an analysis of the structure-activity studies used to develop the non-covalent inhibitors that form part of some of the complexes with human DPP-IV available at the Protein Data Bank.

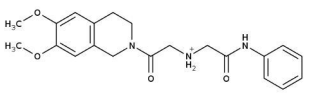
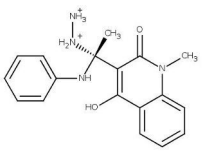
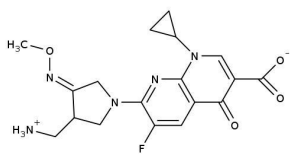
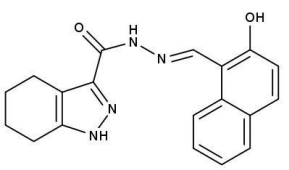
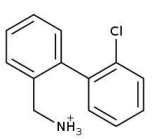
1. Introduction

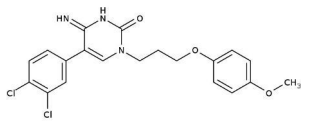
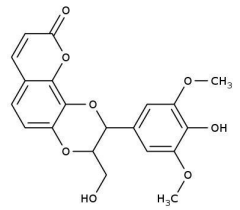
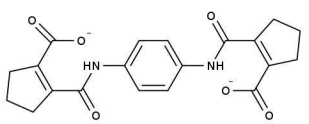
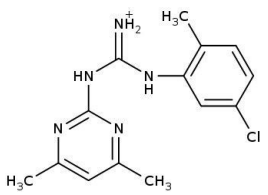
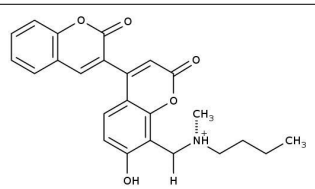
Type 2 diabetes mellitus (T2DM) is a chronic metabolic disease characterized by hyperglycemia and resulting from the body's ineffective use of insulin (*i.e.*, a gradual decline in insulin sensitivity and/or insulin secretion). In 2015, diabetes was responsible for 5 million deaths (46.6% of which were of people under 60 years old) and diabetes deaths are expected to double between 2005 and 2030 and become the 7th leading cause of death by 2030.[1,2] It is estimated that between 415 and 422 million people worldwide currently have diabetes (around 90% of them corresponding to T2DM and 46.5% of them undiagnosed) and that this number will increase to 642 million by 2040.[1,2] This high prevalence means that global expenditure on diabetes treatment was between USD673 and USD1,197 billion in 2015 (12% of worldwide healthcare expenditure) and is expected to rise to between USD802 and USD1,452 billion by 2040.[1] The epidemic dimensions of diabetes therefore make this a priority problem to be solved by healthcare agencies around the world.[1,2]

In the U.S. there are now 12 different drug classes available as an adjunct to diet and exercise to manage hyperglycemia in T2DM patients.[3] DPP-IV inhibitors are one of these and act by increasing circulating levels of GLP-1 and GIP (thereby prolonging their action), which leads to decreased levels of blood glucose, HbA_{1c} and glucagon, thus improving glucose homeostasis with a lower risk of hypoglycemia. Other studies suggest that DPP-IV inhibitors may have cardioprotective effects (although this remains to be confirmed, since studies generally either give contradictory results or show that they are neutral in terms of cardiovascular effects).[4–15] Eleven DPP-IV inhibitors are now commercially available in different countries (*i.e.*, sitagliptin, vildagliptin, saxagliptin, alogliptin, linagliptin, teneligliptin, gemigliptin, anagliptin, trelagliptin, evogliptin and omarigliptin, with the last three being approved during 2015; see Table S1 in the Supplementary Materials)[16] and there are many more in different stages of clinical studies.[17] There is also plenty of information on the 3D structure of human DPP-IV (*i.e.,* apo forms, complexes with oligopeptides, and covalent and non-covalent complexes with drug-like molecules) at the Protein Data Bank (PDB; see Table S2).[18]

Virtual screening (VS) has been used over recent years to discover DPP-IV inhibitors in molecular databases,[19–27] with most of the inhibitors identified having bioactivity in the μM

Table 1. The most active DPP-IV inhibitors identified in molecular databases by means of virtual screening and arranged in order of descending bioactivity.This table shows the protonation states for all the compounds at pH=7 according to Forge.[152]

Structure and activity data for the most active VS hits	Main VS filter[a]							
	Glu205, Glu206, Tyr662	S_2 pocket					S_1 pocket	Others
		Arg125	Ser209	Phe357	Arg358	Tyr547		
 **HWL-892**[26, b] IC_{50} for DPP-IV = 0.148 μM	**HBA**[c]			**HPH**	**HBA**		**HPH**	**HPH**[d]
 **MDPI-12398**[27] IC_{50} for DPP-IV = 0.73 μM	Consecutive dockings using Glide and Gold Ligand interactions: HBD (N-terminal recognition region), HPH (S_1 pocket) and π-π stacking (Phe357)							
 **Gemifloxacin**[21] IC_{50} for DPP-IV = 1.12 μM	**PI**	**HBA**	**HBA**					**HBA** (Gln553)
 Compound **1**[23,114] IC_{50} for DPP-IV = 2.12 μM and 5.77 μM	**HBD**			**HPH**			**HPH**	**HBA**[e] (Arg669, His740)
 Compound **7a**[20] IC_{50} for DPP-IV = 2.3 μM	*In silico* fragment screening within the S_1 pocket of DPP-IV by means of pharmacophore constrained docking. Ligand interactions: PI (N-terminal recognition region), HPH (S_1 pocket)							

 **NCI0211295**[25, f] 40% of DPP-IV inhibition at 10 μM	**PI**					**HPH/ arom.**	**HPH**	**HBA** (Gln553)
 Compound **4**[22] IC_{50} for DPP-IV = 14.13 μM	Reverse docking using the TarFisDock. Ligand interactions: HBA (Arg125), HPH (Tyr547), HBA (Tyr547 and Ser630), HPH (S_1 pocket)							
 **NCI0294730**[21] 34% of DPP-IV inhibition at 10 μM		**NI, HBA**[g]		**HPH**		**HBA, HPH**		**HBA**[g] (Trp629)[h]
 Compound **14**[19, i] 81.9% of DPP-IV inhibition at 30μM	**PI**	**HBA**					**HPH**	
	PI						**HPH**	**HPH**[j]
 **C5**[24] IC_{50} for DPP-IV = 61.55 μM	**PI/HBD**	HBA		HPH/ arom.	HBA	HPH/ arom.	**HPH/ arom.**	HPH/ aromatic (Trp629/ Ser630)

PI: positive ionizable feature
NI: negative ionizable feature
HBA: hydrogen bond acceptor
HBD: hydrogen bond donor
HPH: hydrophobic feature
arom.: aromatic

[a] In those VS where the main filter was a pharmacophore (or where a pharmacophore was used to explain how the ligand binds to DPP-IV), the relative location of the site features in the DPP-IV binding site and the residue that strongly interacts with the pharmacophore site are described. Equivalent sites in different pharmacophores are then found in the same column **(1)** in bold if they were compulsory during the VS, or **(2)** underlined if they were optional but become matched by the most active ligand found by the VS. If the main filter of the VS was a protein-ligand docking, then the intermolecular interactions for the most active ligand found during the corresponding VS are shown.

[b] The assignment of the pharmacophore sites to the different binding site locations was done by visual comparison of Figures 4 and 8C following Xing *et al*.[26]

[c] Although this site is set up as a hydrogen bond acceptor by the pharmacophore authors,[26] the **HWL-892** ligand has a positively ionized amino group forming salt bridges with Glu205 and Glu206 that matches this location. This site should therefore be labeled as a positive ionizable/hydrogen bond donor and not as a hydrogen bond acceptor (indeed, as can be seen for all the other pharmacophores in the table, this site is considered to be a hydrogen bond donor and/or a positive ionizable feature in the other pharmacophores).

[d] The information available in Xing *et al.*[26] did not enable us to assign this site to any feature of the **HWL-892** ligand.

[e] Arg669 for the first pose and His740 for the second.

[f] No data are reported by Al-masri *et al.*[25] relative to the binding mode of this compound.

[g] Corresponds to the same site.

[h] This residue is incorrectly numbered (*i.e.,* Trp637 instead of Trp629) in Figure 7a from Al-masri *et al.*[21]

[i] It is not clear from Ward *et al.*[19] which of the two pharmacophores is matched by compound **14**, but the analysis of Figures 1 and 2 in that paper suggests that it is the first (because the chlorine substituent could interact with Arg125).

[J] After visual inspection of Figure 2 in Ward *et al*.,[19] it is not clear to which DPP-IV pocket this pharmacophore site belongs.

range (see Table 1) but with no measurement of their selectivity over related enzymes like DPP8 and DPP9 (inhibition of either DPP8 or DPP9 has been identified as responsible for alopecia, thrombocytopenia, reticulocytopenia, multiorgan histopathological changes, enlarged spleen and mortality in rats and gastrointestinal toxicity in dogs, while DPP9 inhibition produces neonatal lethality in mice).[28,29] In contrast, several recent structural-activity studies (SAR) in the literature describe the synthesis of novel and selective compounds with nM activity as DPP-IV inhibitors.[30–48]

In the last decade, many reviews focusing on DPP-IV inhibitors have been published.[49–59] Some of these address a wide range of topics such as the incretin system and incretin mimetics,[53] DPP-IV inhibitor selectivity and the implications of DPP-IV inhibition,[52] and even structure optimization in the search for chemical stability, selectivity and favorable pharmacokinetic properties.[58] Most of them devote a major part of their content to classifying DPP-IV inhibitors by structure, listing their activities and their pharmacokinetic

and toxicological properties, describing the binding mode of different inhibitor types, specifying which moieties interact with different DPP-IV subsites and reporting how changes in their substituents affect their bioactivity.[49–51,54–57,59]

In contrast, and far from being a purely descriptive work, the present review aims to gather together published SAR data from studies of different inhibitor series and then extract (with the help of Activity Miner; Cresset BioMolecular Discovery Ltd)[60] relevant information that can be used to **(a)** explain the most important activity cliffs for DPP-IV inhibition according to the binding site structure of DPP-IV, **(b)** explain the most important selectivity cliffs for DPP-IV inhibition in comparison with other related enzymes such as DPP8 and DPP9, and **(c)** on the basis of the information deriving from the activity/selectivity cliff analysis, suggest how VS protocols might be improved to favor the early identification of potent and selective DPP-IV inhibitors in molecular databases. This is of interest, for instance, when it comes to finding potent and selective DPP-IV inhibitors of natural origin that could be used as bioactive compounds in functional food design (in which the chemical modification of bioactive molecules to improve their potency and selectivity is not allowed),[61] and also for finding lead molecules that only need minor changes in their structure before they go to preclinical assays (thereby keeping to a minimum the costs and time needed for new DPP-IV inhibitors to arrive on the market).

2. DPP-IV binding site description

DPP-IV is a homodimeric transmembrane glycoprotein. Each subunit of the protein is anchored to the plasma membrane by a hydrophobic helix consisting of 22 amino acids.[62] Each subunit has a large globular extracellular region that contains an active site located at the interface between the β-propeller domain (from residues 56 to 497) and the α/β hydrolase domain (from residues 509 to 766) (see Figure 1).[56,63–65] The cleavage of the extracellular portion of DPP-IV from the transmembrane section results in a soluble circulating form of approximately 100 kDa. This soluble form is found in plasma and cerebrospinal fluid.[62,66] DPP-IV is secreted as a mature monomer but requires dimerization to undergo normal proteolytic activity.[67]

The DPP-IV binding site is highly druggable in the sense that tight and specific binding to the

enzyme can be achieved using small molecules that have drug-like physicochemical properties.[53,68] It is accessible in two ways: **(1)** via an opening in the β-propeller domain, and **(2)** via the large side opening, which is formed at the interface of the β-propeller and the α/β-hydrolase domain (see Figure 1).[62,65,69] The structural features of DPP-IV suggest that substrates and inhibitors enter or leave the binding site via the side opening. Thus the ligands can directly reach the active site and are correctly oriented for the subsequent cleavage. However, this possibility has not been fully elucidated.[62,70,71]

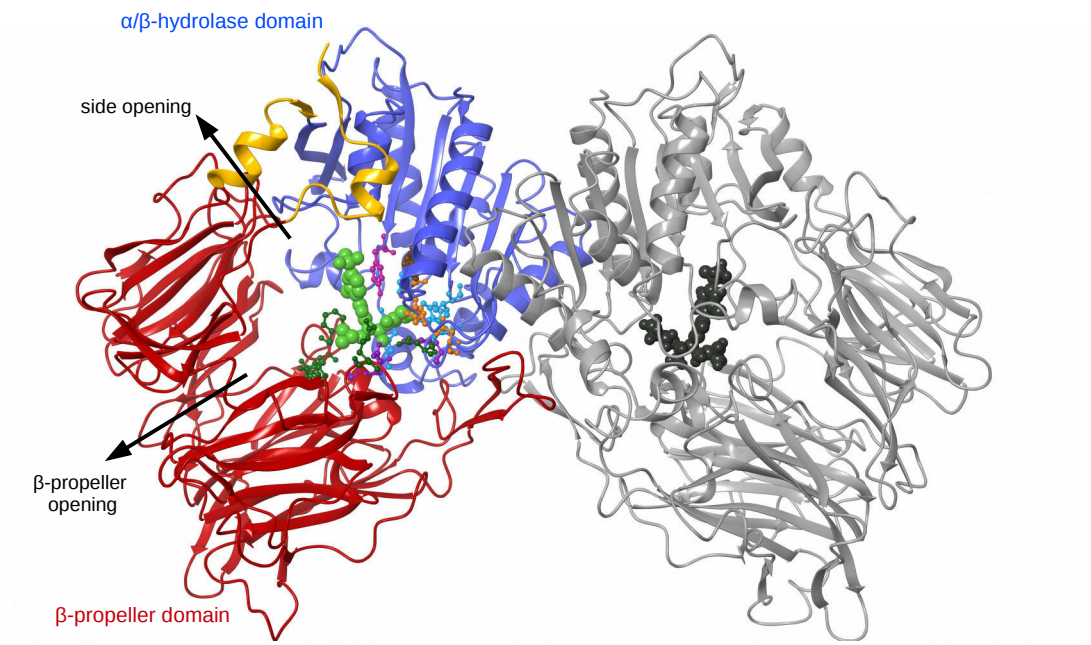


Figure 1. A general overview of the structure of human DPP-IV homodimer. The domain structure for one of the two subunits is also shown (with the β-propeller domain in red, the α/β hydrolase domain in blue and the interdomain region in yellow). The interface between these two domains forms a central cavity which contains the ligand (shown as a spacefill model with atoms in green). Residues that play an important role in the active site (see Figure 2) are shown in the context of the structure of one of the two subunits using a color code to distinguish them (those from the catalytic triad are orange, those from the N-terminal recognition region are lilac, those from the S_1 subsite are light blue, those from the S_2 subsite are dark green, those from the S_1' subsite are pink and those from the S_2' subsite are brown). The DPP-IV binding site is accessible in two ways: **(1)** via an opening in the β-propeller domain, or **(2)** via the large side opening formed at the interface of the β-propeller and the α/β-hydrolase domain (which is the most plausible way for substrates and inhibitors to enter or leave the binding site). PDB entry 3C45[124] was used to obtain this figure with the help of the Maestro program.[165]

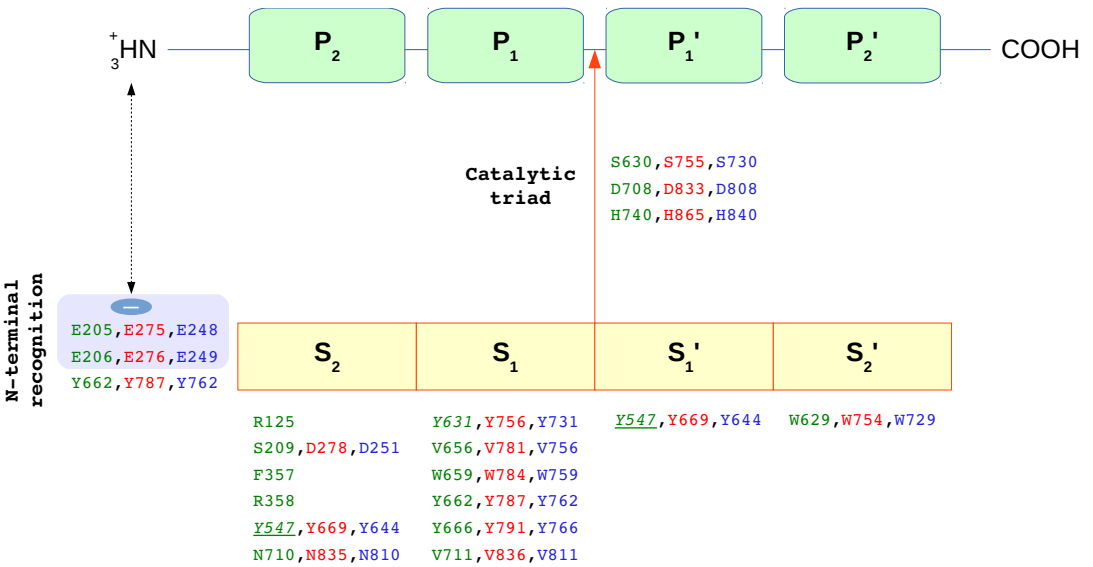


Figure 2. Schematic view of subsite organization in the DPP-IV binding site. The DPP-IV subsites occupied by peptide residues P_2, P_1, P_1', P_2' are labeled S_2, S_1, S_1', S_2' respectively. The point of cleavage of the peptide substrate is between the bond binding residue P_1 with residue P_1'. The DPP-IV residues that form part of the different sites (or other relevant parts of the active site) are shown in green, while residues at equivalent 3D locations for DPP8 and DPP9 in homology models downloaded from ModBase[101,102] are shown in red and blue respectively. The negatively charged Glu205 and Glu206 allow a salt bridge to be formed with the positively charged N-terminal end of the oligopeptide. Tyr547 is underlined because, due to its position, it can be considered part of either the S_2[24,58,73,98,166] or the S_1' pocket.[57,167,168] The Arg125 and Asn710 at the S_2 pocket are essential for coordinating the carbonyl of the P_2 residue and, together with Glu205 and Glu206, align the substrate optimally for the nucleophilic attack by Ser630.[77] Residues forming the oxyanion hole (*i.e.,* Tyr631 and Tyr547) are shown in italics.

On the active site of a protease there are subsites labeled according to the peptide residue that they bind.[72] The point of peptide cleavage is at the peptide bond that binds residue P_1 with residue P_1'. As a result, the residues that surround this position are labeled relative to the cleavage site as P_2, P_1, P_1', P_2' and so on. Therefore the protein subsites occupied by residues P_2, P_1, P_1', P_2' are labeled S_2, S_1, S_1', S_2' respectively. Figure 2 shows the residues that have been identified as part of the different DPP-IV subsites[21,56,68,71,73–76] and those predicted to be their equivalents at DPP8 and DPP9. Apart from these sites, other important groups of residues at the DPP-IV binding site are: **(a)** the N-terminal recognition region formed by residues Glu205, Glu206 and Tyr662 (where Glu205 and in some cases Glu206 form a salt bridge with the peptide's basic amine); **(b)** the oxyanion hole formed by the

backbone amine of Tyr631 and the side chain hydroxyl of Tyr547, which stabilizes the negatively charged tetrahedral oxyanion intermediate generated in the transition state;[77] and **(c)** the catalytic triad formed by residues Ser630, Asp708 and His740 (with Ser630 cleaving the peptide bond between P_1 and P_1' by performing a nucleophilic attack). Additionally, some authors have assigned Val207, Ser209, Phe357 and Arg358 to a site beyond S_2 where the inhibitors but not the substrates can bind well to increase their inhibitory activity[78,79] and which has been termed either the S_2 extensive subsite[17,57,65,78,80] or the S_3 pocket.[81–83] Nevertheless, in line with most authors[23,43,54,56,58,59,84,85] we have assigned Ser209, Phe357 and Arg358 to the S_2 subsite (although throughout this review we also sometimes use the term *S_2 extensive subsite* to refer collectively to these three residues).

The superposition of the large number of experimentally validated structures available nowadays (see Table S2) reveals a slight flexibility of the residues of the active site with the exception of Arg358, Tyr547 and Trp629 (see Figure 3). In this regard (and according to a recent paper)[86], Figures 3E and 3F show how the Arg358 side chain has equivalent orientations regardless of the presence/absence of an inhibitor. It has also been suggested that this flexibility is a consequence of the absence of water molecules around this side chain.[86] In the case of Tyr547, two different orientations are clearly shown where the angle between the aromatic groups is around 70º.[87] According to our analysis, only one Tyr547 conformer is found at the apo form (see Figure 3E), while changes in its conformation are not always related to the formation of a π-π interaction with the ligand (see Figures 3C and 3D and also the following PDB complexes: 3CCC[75], 4N8E[35] and 4N8D[35]). For Trp629, just two conformers are observed (see Figures 3E and 3F), only one of which is present in the two subunits of 1PFQ (apo form) where it shields Ser630 from access.[88]

Two water molecules have recently been identified as being common at the binding site of 92 DPP-IV crystallized structures.[86] It has been suggested that these waters could play two different roles: **(a)** maintaining the proper orientation of the side chains of the Glu205/Glu206 dyad through a network via the water molecules; and **(b)** appropriately arranging the inhibitor at the S_2 subsite.[86]

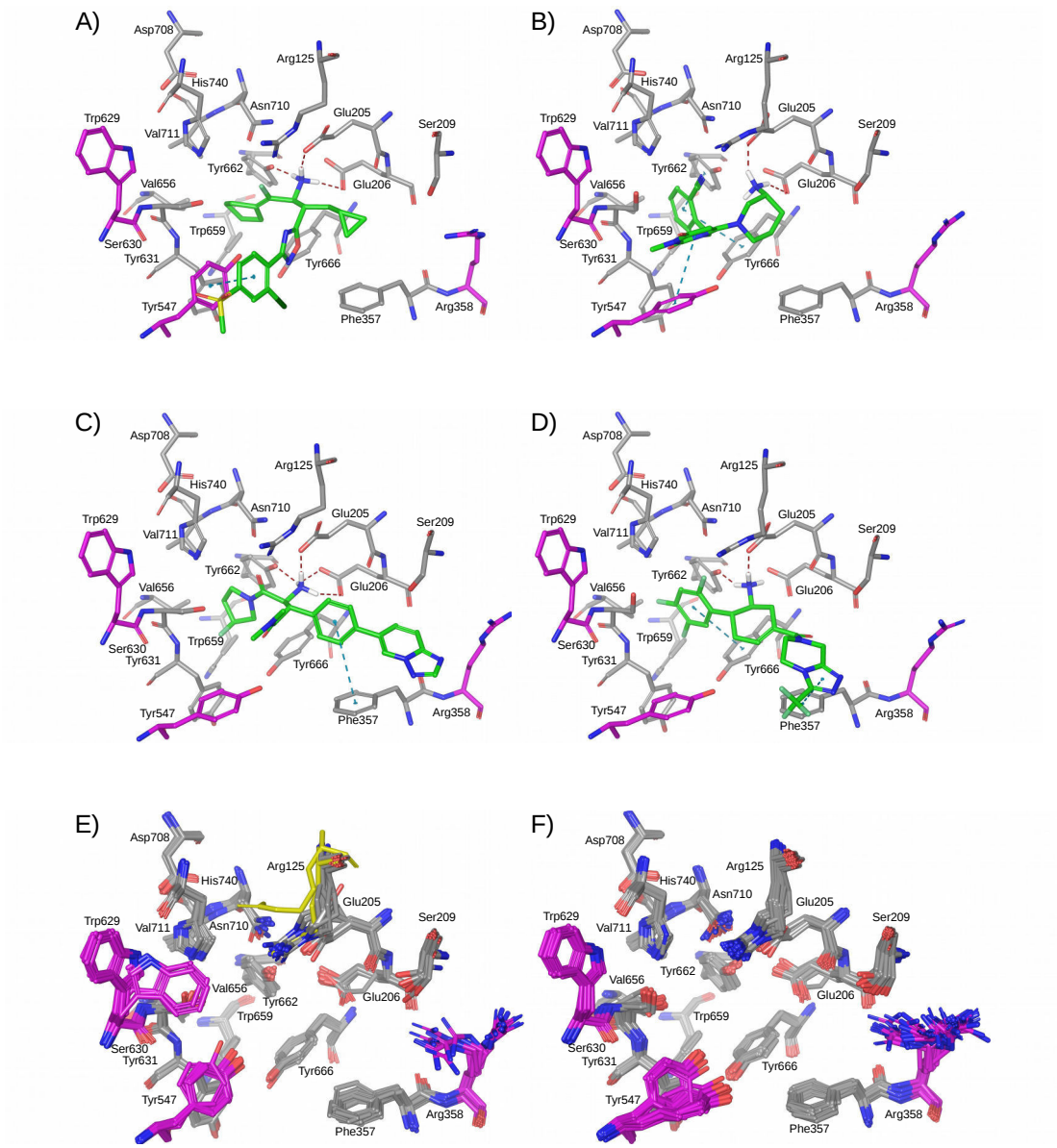


Figure 3. Validated coordinates for the binding site for different DPP-IV chains. Validation was performed either with VHELIBS[169] or by visual comparison with the corresponding electron density map[170]. Only the Arg125 coordinates do not fit well in some electron density map (*i.e.,* 1PFQ,[88] 4KR0[171] and 4L72[172]) and they are therefore shown in yellow to distinguish them from reliable coordinates. The blue dashed lines show the π-π interactions between ligands and aromatic residues in the DPP-IV binding site and were calculated with the help of Maestro [165] using default options. This figure shows that only Arg358, Tyr547 and Trp629 (in magenta) have different conformations depending on crystallization conditions: panel **A** complex with a fluoroolefin inhibitor (PDB code 3C45);[124] panel **B** complex with alogliptin (PDB code 3G0B);[161] panel **C** complex with a β-substituted biarylphenylalanine amide inhibitor (PDB code

2FJP);[173] panel **D** complex with the ABT-341 inhibitor (PDB code 2I78);[174] panel **E** superposition of apo chains; and panel **F** superposition of the 63 non-covalent DPP-IV/inhibitor complexes currently available (one chain per PDB file; see Table S2). The apo forms in panel **E** correspond to all available chains in Table S2 that do not form a complex with any inhibitor (*i.e.,* chains A and B for 1TK3[77] and 1PFQ[88]; chains A, C and D for 2I78[174], 2OAG[175] and 2OQI[140]; chains A, B, C and D for 1W1I;[176] chains A and C for 4QZV[177]; chain A for 4KR0[171] and 4L72;[172] and chain B for 2OQV[140]). This figure was obtained with the help of the Maestro program.[165]

3. Comparing the 3D structures for DPP-IV, DPP8 and DPP9

Like DPP-IV, DPP8 and DPP9 belong to the prolyl oligopeptidase family.[89] Both are to be found as monomers in the cytoplasm of human blood lymphocytes, pulmonary leucocytes and monocytes.[90] Although their physiological role has yet to be verified, DPP8 seems to be involved in T-cell activation, while DPP9 is highly expressed in cancer cells, normal skeletal muscle, and heart and liver tissues.[52,55,74,90] Since two splice variants are described for DPP8 (with a length of 882 and 898 residues respectively),[91] the numbering of residues may not coincide between isoforms. The same occurs with DPP9, which also has two splice variants (with a length of 863 and 892 residues respectively),[91,92] but it has not been demonstrated whether both of these are biologically active.[91,92]

In contrast to DPP-IV (which has been crystallized in several different conditions, see Table S2), the 3D structures for DPP8 and DPP9 proteins have yet to be elucidated. In the future, the availability of their experimental 3D structure will improve the understanding of their catalytic mechanism and their physiological importance. Meanwhile different homology models have been suggested (see Table 2) based on sequence similarity with DPP-IV.[81,93–97] Thus, depending on the sequence alignment used to build the different homology models, the sequence similarity with DPP-IV lies in the 33-55% and 34-43% ranges respectively for DPP8 and DPP9 (see Table 2).[52,89,91,94,97,98] These homology models can be used to hypothesize as to the clues for DPP-IV selectivity by analyzing the structural differences between the three enzymes. The accurate design of the 3D structure around the binding site of DPP8 and DPP9 therefore becomes a powerful tool for identifying activity and selectivity cliffs.

In the present work, in order to explain these structural differences in the active site we have used two different sets of homology models for DPP8 and DPP9 (see Table 2). The first was built using a DPP-IV crystal structure from *Stenotrophomonas maltophilia* as a template

(PDB code 2ECF)[99] and is available from the ModBase database[100] using access numbers 37577089[101] and 123983020[102] for DPP8 and DPP9 respectively. The second set of homology models was reported by Janardhan & Reddy[97] and built using the A chain of a human DPP-IV structure (PDB code 1X70) as a template.[103] Although other homology models for DPP8/DPP9 have been described in the bibliography (see Table 2), their coordinates are not currently available on request and they have therefore not been used in this review (although the information provided in the papers that describe them is nevertheless used when possible).

Table 2. Homology models of DPP8 and DPP9 found in the literature and the best homology models available in ModBase.

Bibliographic reference or ModBase database entry	**DPP8**[a]	**DPP9**[a]
Pitman *et al.* (2006)[93]	1N1M (hDPP-IV; 51%) 1XFD (hDPP6; 52%)	Not modelled
Rummey & Metz, (2007)[94]	2BGR (chain A; hDPP-IV; 37%)	2BGR (chain A; hDPP-IV; 37%)
Kang *et al.* (2007)[81]	Data not available	Data not available
Yazbeck *et al.* (2009)[96]	1R9N (hDPP-IV; 55%) 1Z68 (hFAP) 1XFD (hDPP6)	Not modelled
Park *et al.* (2008)[95]	1N1M (hDPP-IV) 1Z68 (hFAP) 1XFD (hDPP6)	1N1M (hDPP-IV) 1Z68 (hFAP) 1XFD (hDPP6)
Janardhan & Reddy (2011)[97]	1X70 (chain A; hDPP-IV; 43%)	1X70 (chain A; hDPP-IV; 43%)
37577089[100,101]	2ECF (chain A; sDPP-IV; 33%[b])	
123983020[100,102]		2ECF (chain A; sDPP-IV; 34%[b])

[a] Each cell in this column contains the PDB codes (and chain, where necessary) that were used as a template to build the corresponding homology model. The protein to which each PDB file belongs is also identified with its name (*i.e.,* DPP-IV, DPP6 and FAP) and a prefix for its source (*i.e.,* **h** for "human" and **s** for *Stenotrophomonas maltophilia*). When available, the sequence similarity between DPP8/DPP9 and each PDB template reported in the literature is also given.

[b]Corresponds to sequence identity (*i.e.,* the percentage of identical residues in the alignment with the template protein)

The main differences between these two sets of homology models and DPP-IV involve the binding site cavity, in particular around the R-loop[94] (see Figure 4), the P_2-loop[94] (see Figure

5), Cys551 (see Figure 6) and the N-terminal recognition region (see Figure S1). Although some studies have suggested that the S_1 pocket size is smaller in DPP-IV than in DPP8/DPP9,[55,57,81,97,104,105] Figure 7 shows that there are no major differences between them (irrespective of whether the homology models from the ModBase database[101,102] or from Janardhan & Reddy[97] are used).

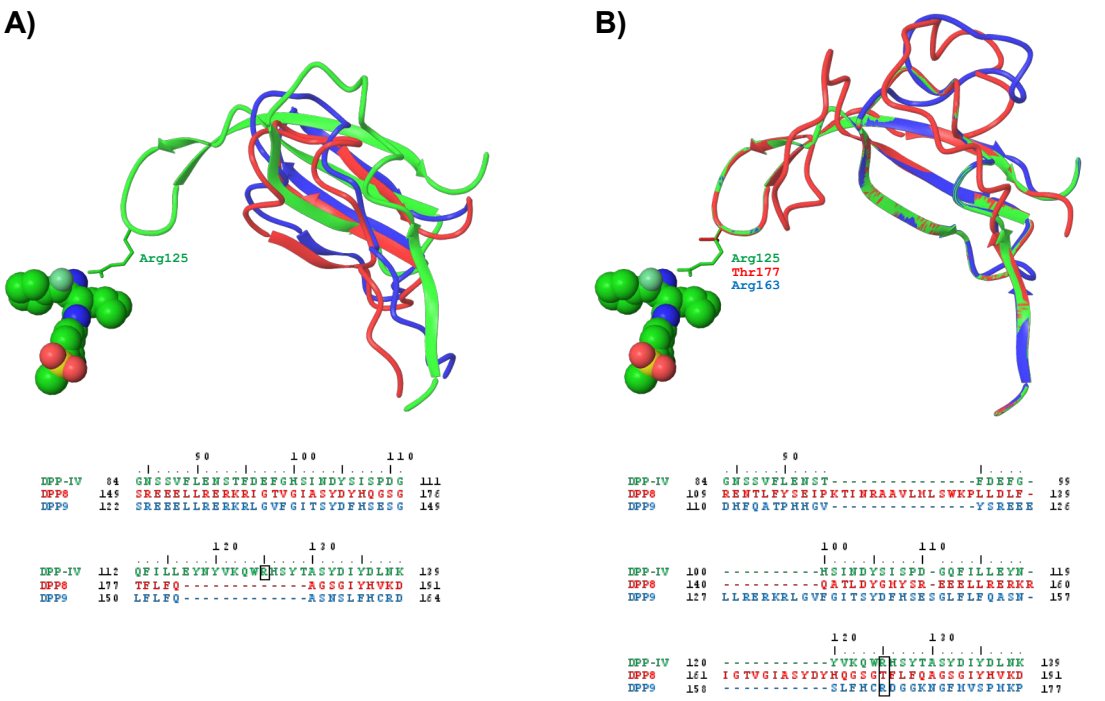


Figure 4. This figure compares the R-loop in DPP-IV with the equivalent one in the ModBase[101,102] and Janardhan & Reddy[97] homology models. Panel **A** corresponds to the superposition of the homology models downloaded from ModBase[101,102] relative to DPP-IV (PDB code: 3C45).[124] Panel **B** corresponds to the superposition of the homology models generated by Janardhan & Reddy[97] relative to DPP-IV (PDB code: 1X70).[103] DPP-IV, DPP8 and DPP9 are shown in ribbons and colored green, red and blue respectively (following the same color schema used in Figure 2 and in the multialignment at the bottom of each panel). The DPP-IV ligand at 3C45 is shown in spacefill in both panels to reference the active site location. The multialignment at the bottom of each panel shows which residues have an equivalent location in the corresponding 3D superposition between DPP-IV, DPP8 and DPP9. Residue Arg125 and its equivalents in DPP8 and DPP9 (if any) are represented in wireframe format in the 3D structures and boxed in the multialignment. This figure was obtained with the help of the Maestro program.[165]

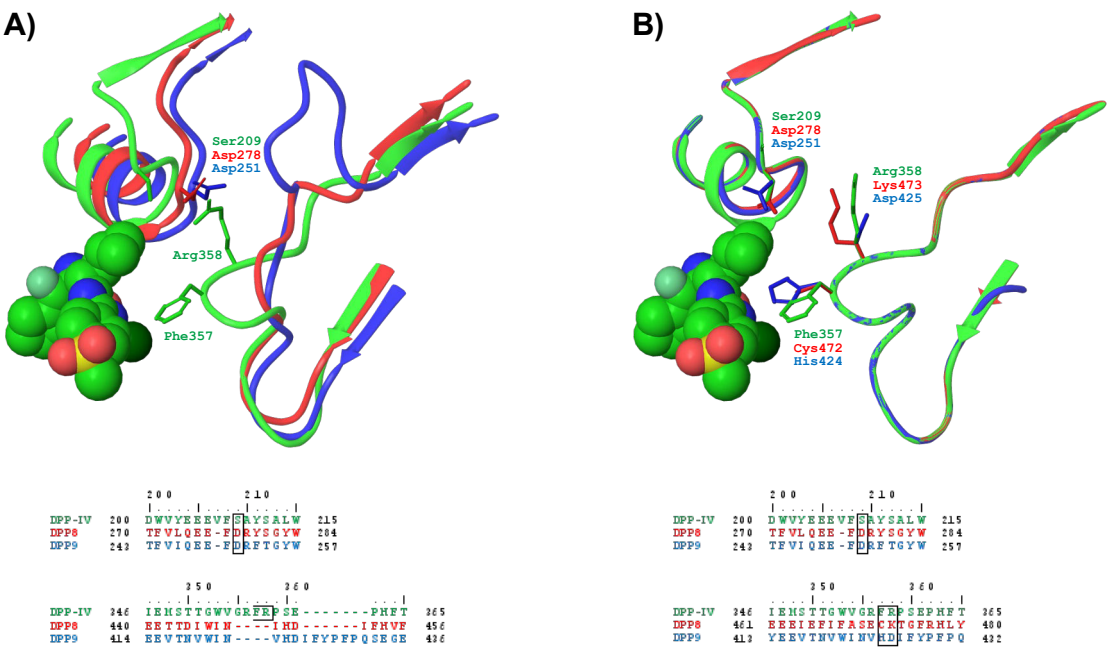


Figure 5. This figure compares the sequence around the P_2-loop in DPP-IV with the equivalent ones in the ModBase[101,102] and Janardhan & Reddy[97] homology models. Panel **A** corresponds to the superposition of the homology models downloaded from ModBase[101,102] relative to DPP-IV (PDB code: 3C45).[124] Panel **B** corresponds to the superposition of the homology models generated by Janardhan & Reddy[97] relative to DPP-IV (PDB code: 1X70).[103] DPP-IV, DPP8 and DPP9 are shown in ribbons and colored green, red and blue respectively (following the same color schema used in Figure 2 and the multialignment at the bottom of each panel). The DPP-IV ligand at 3C45 is shown in spacefill in both panels to reference the active site location. The multialignment at the bottom of each panel shows which residues have an equivalent location in the corresponding 3D superposition between DPP-IV, DPP8 and DPP9. Residues Ser209, Phe357 and Arg358 and their equivalents in DPP8 and DPP9 (if any) are represented in wireframe format in the 3D structures and boxed in the multialignment. This figure was obtained with the help of the Maestro program.[165]

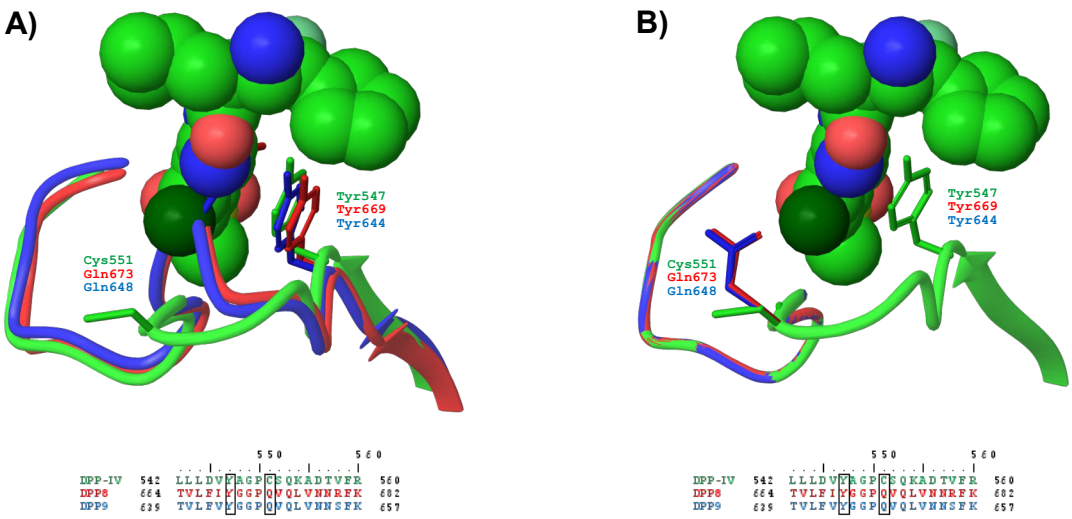


Figure 6. This figure compares the residues surrounding Tyr547 and Cys551 in DPP-IV with the equivalent ones in the ModBase[101,102] and Janardhan & Reddy[97] homology models. Panel **A** corresponds to the superposition of the homology models downloaded from ModBase[101,102] relative to DPP-IV (PDB code: 3C45).[124] Panel **B** corresponds to the superposition of the homology models generated by Janardhan & Reddy[97] relative to DPP-IV (PDB code: 1X70).[103] DPP-IV, DPP8 and DPP9 are shown in ribbons and colored green, red and blue respectively (following the same color schema used in Figure 2 and the multialignment at the bottom of each panel). The DPP-IV ligand at 3C45 is shown in spacefill in both panels to reference the active site location (in panel **A**, covering the equivalent residues for Cys551 in DPP8/DPP9). The multialignment at the bottom of each panel shows which residues have an equivalent location in the corresponding 3D superposition between DPP-IV, DPP8 and DPP9. Residues Tyr547 and Cys551 and their equivalents in DPP8 and DPP9 are represented in wireframe format in the 3D structures and boxed in the multialignment. This figure was obtained with the help of the Maestro program.[165]

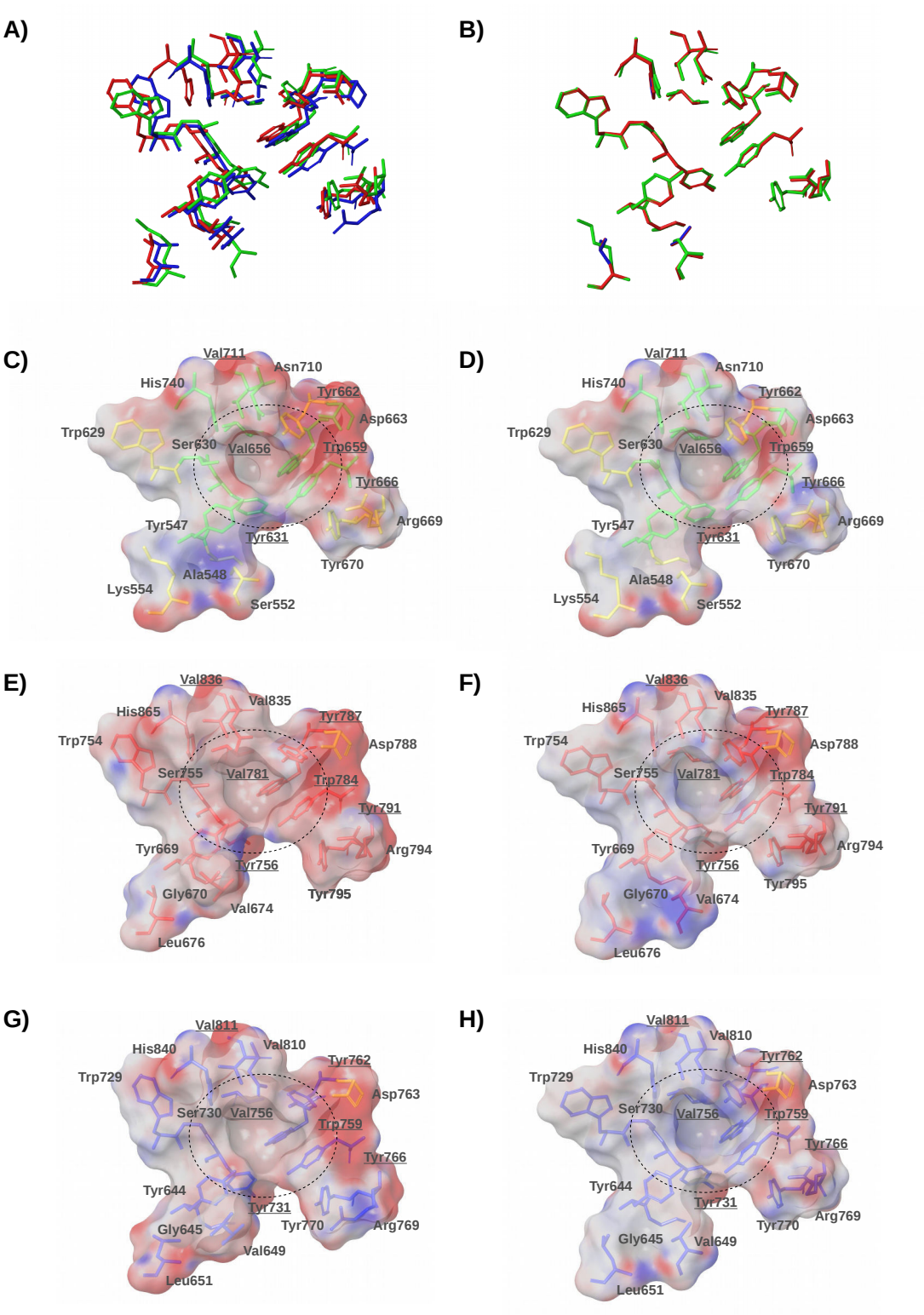

A)
B)
C)
Val711
Asn710
His740
Tyr662
Asp663
Trp629
Ser630
Val656
Trp659
Tyr666
Tyr547
Arg669
Tyr631
Tyr670
Ala548
Lys554
Ser552
D)
Val711
Asn710
His740
Tyr662
Asp663
Trp629
Ser630
Val656
Trp659
Tyr666
Tyr547
Arg669
Tyr631
Tyr670
Ala548
Lys554
Ser552
E)
Val836
Val835
His865
Tyr787
Trp754
Asp788
Ser755
Val781
Trp784
Tyr791
Tyr669
Arg794
Tyr756
Gly670
Tyr795
Val674
Leu676
F)
Val836
Val835
His865
Tyr787
Trp754
Asp788
Ser755
Val781
Trp784
Tyr791
Tyr669
Arg794
Tyr756
Tyr795
Gly670
Val674
Leu676
G)
Val811
Val810
His840
Tyr762
Trp729
Asp763
Ser730
Val756
Trp759
Tyr766
Tyr644
Tyr731
Gly645
Tyr770
Arg769
Val649
Leu651
H)
Val811
Val810
His840
Tyr762
Trp729
Asp763
Ser730
Val756
Trp759
Tyr766
Tyr644
Tyr731
Tyr770
Arg769
Gly645
Val649
Leu651

Figure 7. This figure compares the S_1 pocket in human DPP-IV with the equivalent pockets in the homology models for DPP8 and DPP9 from ModBase[101,102] and Janardhan & Reddy.[97] Panel **A** shows the superposition of all residues in panels **C**, **E** and **G**, while panel **B** shows the superposition of all residues in panels **D**, **F** and **H**. In panels **A** and **B**, the residues are colored according to the protein to which they belong (*i.e.,* DPP-IV, DPP8 and DPP9 are colored green, red and blue respectively, following the same color schema used in Figure 2). Panels **C** to **H** show the protein surfaces for the different DPP8, DPP9 and DPP-IV models or PDB files being compared in this figure. Surfaces are colored from red (negative) to blue (positive) according to their Poisson-Boltzmann electrostatic potentials (where potentials range from -80.0 to 80.0), with the S_1 pocket circled for clarity [the DPP-IV residues that form this pocket according to Kuhn *et al.*[68] (or the equivalent ones in DPP8/DPP9) are underlined]. Panels **C**, **E** and **G** correspond to the 3C45[124] and ModBase models for DPP8 and DPP9, while panels **D**, **F** and **H** correspond to the 1X70[103] and Janardhan & Reddy models for DPP8 and DPP9 respectively. In panels **C** to **H** the residues that according to Janardhan & Reddy form the S_1 pocket in DPP-IV, DPP8 and DPP9 are colored using the same schema as in panels **A** and **B**, while those not mentioned by these authors as forming part of the S_1 pocket (but which are equivalent to other residues that these authors consider part of the S_1 pocket in either DPP-IV, DPP8 or DPP9) are shown in yellow [*i.e.,* Ala548, Ser552, Lys554, Trp629, Tyr662, Arg669 and Tyr670 in panels **C** and **D**; Asp788 in panels **E** and **F** and Asp763 in panels **G** and **H**]. This figure was obtained with the help of the Maestro program.[165]

4. Reviewing virtual screening for DPP-IV inhibitors

In recent years different VS protocols have been proposed for identifying DPP-IV inhibitors in molecular databases[19–27] and, interestingly, **(a)** most of them use pharmacophores as part of their VS workflows, and **(b)** no evaluation of the bioactivity of the VS hits on DPP8 and DPP9 was performed in any of them (see Table 1). In the following paragraphs we describe the main features and achievements of these VS protocols.

- Ward *et al.* developed the first VS protocol for identifying DPP-IV inhibitors in molecular databases.[19] This VS workflow consisted of different sequential filters where the output molecules of one filter were the input molecules for the next, and so on. The initial database containing 800,000 compounds was thus initially prefiltered according to physical and chemical properties. This was followed by the generation of a conformer library for the tautomers and protonation states of the remaining 500,000 molecules. This was screened through two different structure-based pharmacophores (each containing three features; see Table 1) and 20,000 compounds from each of the pharmacophores were selected on the basis of the RMSD overlap of each molecule with the pharmacophore and the overlap with the excluded volumes of the active site. Finally, the 40,000 molecules selected were docked with Glide[106,107] into the DPP-IV binding site and the top 8000 compounds

according to the scoring function were chosen. Further clustering and visual inspection of the 8000 molecules enabled a final subset of 4000 compounds to be selected for bioactivity screening. According to this enzymatic assay, the inhibitory activities of these 4000 molecules ranged from 30% to 82% when tested at a concentration of 30 μM of the corresponding compound. The most active compound identified in this VS can be seen in Table 1 and shows a DPP-IV inhibition of 81.9% at 30μM.

- Rummey *et al.* assembled a fragment database using the Available Chemical Directory and their own in-house collection to identify new molecular anchors for the DPP-IV's S_1 subsite by means of a constrained protein-fragment docking.[20] Thus, in order for a fragment to achieve successful placement, it needed to bind to at least two of four selected acceptor points (located at Glu205, Glu206 and Tyr666) and also fulfill a spatial constraint within S_1. The most active fragment identified by this VS shows an IC_{50} for DPP-IV is 2.3 μM (see Table 1).

- Al-masri *et al.* built two ligand-based 3D pharmacophores by exploring the pharmacophoric space of a large and diverse set of conformers for known DPP-IV inhibitors and integrated it with a predictive QSAR model.[21] The pharmacophores thus allowed them to mine conformer databases, while the QSAR model helped them to prioritize the VS hits for subsequent *in vitro* bioactivity testing. One of the pharmacophores comprised four sites (three hydrogen bond acceptors and one positive ionizable feature) and succeeded in identifying gemifloxacin as a DPP-IV inhibitor (IC_{50} for DPP-IV is 1.12 μM; see Table 1). The second pharmacophore comprised five sites (two hydrogen bond acceptors, two hydrophobic features and one negative ionizable feature) and identified a molecule that causes 34% DPP-IV inhibition at 10 μM (see Table 1).

- Zhang *et al.* used a reverse docking approach to identify putative targets for a collection of 19 natural products (NPs) derived from two medicinal plants [*Bacopa monniera* (L.) Wettst. and *Daphne odora* Thunb. var. *marginata*] used to treat diabetes and inflammation in oriental folk medicine.[22] After screening the *Potential Drug Target Database*[108] with the TarFisDock server,[109] DPP-IV was the most

frequent potential target among the top 5% target candidates. Subsequent *in vitro* measurement of the bioactivity identified that 5 of the 19 NPs were moderate DPP-IV inhibitors (with IC_{50} values ranging from 14.13 μM to 113.76 μM; see Table 1 for the structure of compound **4**, the most active). Subsequently 27 analogs of these five NPs were identified in an in-house NP database and the bioactivity assay showed that 13 of them were moderate DPP-IV inhibitors (with IC_{50} values ranging from 26.92 μM to 87.72 μM).

- Li *et al.*[23] used a VS workflow to predict new DPP-IV inhibitors from the SPECS database.[110] It included the following sequential filters: **(a)** a rigid protein-ligand docking with Glide;[106,107] **(b)** a druglikeness filter inspired by the Lipinski rule of 5;[111] and **(c)** a flexible protein-ligand docking with AutoDock v4.0.[112] The resulting top-ranked 99 compounds were then experimentally tested to measure their bioactivity as DPP-IV inhibitors and 15 were found to have IC_{50} in the 5.77 to 50.32 μM range (the most active compound identified by this VS is shown in Table 1). Subsequent induced-fit docking[113] of these 15 compounds to DPP-IV and further pharmacophore modeling was performed so as to understand how these molecules inhibit DPP-IV. The ability of this pharmacophore to screen a database in search of DPP-IV inhibitors was also confirmed.[23] Afterwards the same research team used the most active compound identified by the VS as a lead compound for obtaining a further 17 derivatives whose activity as DPP-IV was also measured.[114] Only 9 of them were found to inhibit DPP-IV (with IC_{50} values ranging from 3.44 μM to 70.80 μM). In order to explain their SAR, these 9 molecules were docked into the DPP-IV binding site using Glide.[106,107] Furthermore, on the basis of 8 of these 9 compounds, a common pharmacophore hypothesis was developed using the HipHop utility of Catalyst.[115] This common hypothesis consisted of one hydrogen bond donor feature (directed to Glu205), one hydrogen bond acceptor feature (directed to Arg669), and two hydrophobic features (one close to Phe357 and one in the S_1 pocket). The pharmacophore mapping results were in good agreement with the docking results and provided guiding information for further structural optimization.

- Guasch *et al.* used a VS workflow to predict new DPP-IV inhibitors from the NP subset of the ZINC database.[116] This workflow consisted of several sequential steps

in which the output molecules of one step were used as the input molecules for the next step, and so on.[24] First, the 89,165 molecules that were part of this ZINC subset were submitted to an ADME/Tox filter[117] in order to discard molecules that were either potentially toxic or exhibited poor ADME properties. Conformers were then obtained with the help of OMEGA[118,119] for the remaining molecules and filtered through a structure-based common pharmacophore. This pharmacophore was designed by **(a)** selecting PDB complexes from DPP-IV and drug-like reversible inhibitors with IC_{50} ≤ 10 nM, **(b)** using the protein structure to superimpose the corresponding PDB files, **(c)** predicting the contribution of each ligand's functional group to the binding affinity,[120] **(d)** finding which functional group features were spatially equivalent in the different ligands, **(e)** identifying common functional group features that strongly contribute to the binding affinity and setting them as mandatory pharmacophore sites, and **(f)** identifying less common functional group features that contribute less to the binding affinity and setting them as optional pharmacophore sites. The resulting pharmacophore had two compulsory sites (one positive/donor and one hydrophobic/aromatic ring), while the remaining two hydrogen-bond acceptors and three hydrophobic/aromatic ring sites were optional (see Table 1). Phase[121] was then used to filter conformers with the pharmacophore and only those molecules with at least one conformer matching the two compulsory and at least one of the optional sites were considered for the subsequent protein-ligand docking performed with eHiTS.[122] The resulting ligand poses were then filtered again with the pharmacophore but without pose reorientation (*i.e.,* the *score in place* option was set to *on*). Finally, using EON[123] those poses that matched the pharmacophore were submitted to a shape and electrostatic-potential comparison with the experimental pose of the DPP-IV inhibitor at PDB file 3C45[124] (which had the smallest IC_{50} of all the drug-like reversible inhibitors found in DPP-IV inhibitor complexes at the PDB). The reliability of the VS was then demonstrated using an *in vitro* test to determine the inhibitory activity of representative hits (*i.e.,* hits that were chemically different not only from one another but also relative to any known DPP-IV inhibitor). Lastly, in order to predict more potent derivatives, a lead-optimization of the most active compound (IC_{50} = 61.55 μM; see Table 1) was carried out with the help of CombiGlide[125]. The combinatorial screening suggested that the activity of this VS hit

could be increased by **(1)** replacing the original butyl group by a substituent containing a ring with a positive formal charge that could improve the interaction with the S_1 pocket by forming a π-cation interaction with Tyr662 and Tyr666 and also by enclosing the two sides of the ring in the lipophilic protein environment in the pocket, and **(2)** making hydrogen bonds with the S_2 pocket (through either Ser209 or Arg358) or with Arg669. Interestingly, the same VS workflow was also used in another manuscript to predict DPP-IV inhibitors in natural extracts with known antidiabetic activity.[126]

- Al-masri *et al*. used a VS with 2D and 3D filters implemented in a hierarchical cascade to identify new DPP-IV inhibitors.[25] Structure-based pharmacophore models were generated from co-crystallized ligands with potent DPP-IV inhibitory activities using Discovery Studio Visualizer software.[127] The optimum pharmacophore model was then selected by using an in-house database containing active and inactive DPP-IV inhibitors and employed to screen two 3D conformer databases (the NCI[128] and an in-house built database) with the help of Catalyst.[115] This pharmacophore was made up of four different features (one positive ionizable, one hydrogen bond acceptor, one hydrophobic and one hydrophobic/aromatic; see Table 1) and enabled compound conformations with the desired features to be identified. After the pharmacophore screening, 2D virtual filters based on molecular weight, ΔlogP and the number of heavy atoms, rotatable bonds and hydrogen bond acceptors and donors were applied with the help of FILTER (which was also used to remove those molecules with unstable, toxic or reactive functional groups).[129] Next, conformations were built for the remaining compounds with the help of OMEGA[118,119] and the resulting conformer library was used by FRED[130,131] to predict their binding mode at the DPP-IV binding site. Finally, *in vitro* bioassays were performed that confirmed the finding of five novel DPP-IV inhibitors (with inhibition at 10 μM ranging from 17% to 40%; see Table 1 for the most potent DPP-IV inhibitor found) together with another 11 DPP-IV inhibitors already described in a previous VS by the same authors.[21]

- Xing *et al.* performed a hierarchical VS via a multistage workflow.[26] A pharmacophore was built using the HypoGen module from Discovery Studio v2.5[132] and prioritizing

hydrogen-bond acceptor and hydrophobic features relative to the hydrogen-bond donor, positive ionizable and ring/aromatic features. In addition, a maximum of five excluded volumes were automatically added to the pharmacophore in order to improve specificity. The resulting pharmacophore models were validated by the following four different approaches: **(a)** an external test set, **(b)** a systematic cost analysis, **(c)** a Fisher's randomization test, and **(d)** a receiver operating characteristic analysis. Afterwards an in-house database containing 5034 drug-like compounds was screened using the validated pharmacophore (containing two hydrogen-bond acceptors, three hydrophobic features and five excluded volumes; see Table 1). The top 500 pharmacophore hits then underwent a parallel and independent docking study with two different docking program/scoring function combinations (LigandFit/DockScore[133] and Glide/GScore[106,107,134]) and the top 100 docked poses from each docking strategy were compared in order to identify which were common to both (*i.e.,* RMSD < 3.0 Å). This analysis identified 51 poses common to both docking protocols that were visually checked for the presence of either hydrogen bonds or salt bridges with Glu205/Glu206 (considered a prerequisite for potential DPP-IV). Finally, all common poses fulfilling this prerequisite were re-ranked according to their binding energy (calculated by using molecular mechanics generalized Born surface area). After the VS workflow, the hit compounds **HWL-405** and **HWL-892** (with IC_{50} 271 nM and 148 nM respectively) showed the highest inhibitory activity *in vitro*. Several analogs of these hits were synthesized for *in vitro* evaluation (with IC_{50} values ranging from 78 µM to 494 µM) and *in vivo* analysis.

- Tanwar *et al.* used a structure-based VS strategy to look for DPP-IV inhibitors in the MDPI database.[27] Initially the database was filtered so as to remove those molecules that are either reactive or show poor ADME properties. The remaining molecules (together with some approved DPP-IV drugs used to validate the virtual screening protocol) were then docked with Glide[106,107] to 1RWQ[135] by using a grid containing two docking constraints (one hydrophobic constraint at the S_1 pocket and one hydrophilic constraint close to the Glu dyad). During this docking, three consecutive steps were performed (the first with Glide-HTVS, the second with GlideSP and the

third with GlideXP) and the sample for each step was the top 10% according to the results of the previous scoring function (*i.e.,* around 18000 molecules for Glide-HTVS, 1800 for GlideSP and 180 for GlideXP). To further validate the reliability of the scores supplied by GlideXP, the same 180 VS hits were docked to DPP-IV with GOLD[136] in the area around 10 Å of the co-crystalized ligand by using its genetic algorithm with default parameters. Finally, the ligands that were identified as being among the top scorers simultaneously by GlideXP and GOLD were visually checked for proper interactions with the S_1 and the Glu dyad constraints. Six ligands were selected on the basis of docking scores and the availability of sufficient quantities of compounds to perform the biological assays. Interestingly, all the approved DPP-IV drugs that were included in the VS and that bind non-covalently to DPP-IV (*i.e.,* alogliptin, gemigliptin, linagliptin and sitagliptin) were retrieved among the top 100 scored ligands by both programs. After the VS workflow, the hit compound **MDPI-12398** (IC_{50} = 730 nM; see Table 1) showed the highest inhibitory activity *in vitro,* while in the oral glucose tolerance test it also showed the most significant reduction in blood glucose excursion in fed female Wistar rats.

5. Selecting the DPP-IV inhibitor series that give clues on how to favor potency and selectivity and predicting their binding modes

All the series of DPP-IV inhibitors analyzed in the present review[103,124,137–150] fulfill the following criteria: **(a)** they contain compounds with bioactivity in humans in the nM range, **(b)** they also contain compounds that are at least 10-fold less potent in humans than the most active compounds in their corresponding series, and **(c)** at least one compound of the series (or a very similar one from elsewhere) has been crystallized in a complex with human DPP-IV. This last point is crucial because correctly predicting the binding mode of all the compounds in a series is necessary in order to offer valid explanations for the activity and selectivity cliffs relative to the protein environment. Once all the series that fulfill all these requirements were identified, all their compounds were downloaded from Reaxys Medicinal Chemistry[151] and superposed to crystallized ligands with Cresset's Forge v10.4[152] using the *Maximum Common Substructure* and either the default *Accurate But Slow* or the *Very Accurate But Slow* set-up for the conformational search.

Once the alignments were performed and their correctness visually checked, the Activity Miner module in Forge[60,152] was used to calculate the similarity and disparity values between all pairs of compounds within each series. The disparity between a pair of molecules is calculated as the difference in their activity divided by the distance between them (where the distance between a pair of molecules is found from their similarity expressed in either 3D or 2D), as in the following formula:[152]

$$Disparity = \Delta activity/(1-Similarity)$$

where:

(1) if the absolute value of *Δactivity* is smaller than the error associated with activity, *Δactivity* is considered to be zero.

(2) if *Similarity* is greater than 0.95, it is clipped to 0.95 to avoid *Disparity* assuming very large values.

In our case, the distance between a pair of molecules is found from their 3D similarity (where 3D molecules are compared using field and shape similarity terms, with a final 50% contribution from each to the similarity). Therefore high disparity values indicate that a small change in the molecule (*i.e.,* high similarity) has made a big change in the activity and refer to important areas of the SAR landscape.

When comparing compound pairs in this review, two premises were followed: **(1)** in order to correlate the differences in activity with the particular residue or subsite of the protein responsible for them, we focused exclusively on those comparisons between compounds that differ in only one substituent; and **(2)** in pairs of compounds in which one compound has an acute decrease in activity and also clashes with the protein used as a reference in Forge[152], it was assumed that this protein-ligand steric clash was responsible for the decrease in activity and, in consequence, these pairs of compounds were excluded from the analysis.

After the compound pairs with high disparity values were identified, their protein environment was carefully inspected. In each case an explanation for the change in activity was proposed based on the differences between the compounds and their intermolecular interaction with

the protein environment. This allowed us to identify which residue and subsite interactions were responsible for the changes in activity and selectivity. Moreover, the robustness of our conclusions was verified, when possible, with supporting evidence from several independent studies (using results deriving from different series) and from bibliographic data.

6. How to favor potent and selective DPP-IV inhibitors according to the analysis of SAR studies

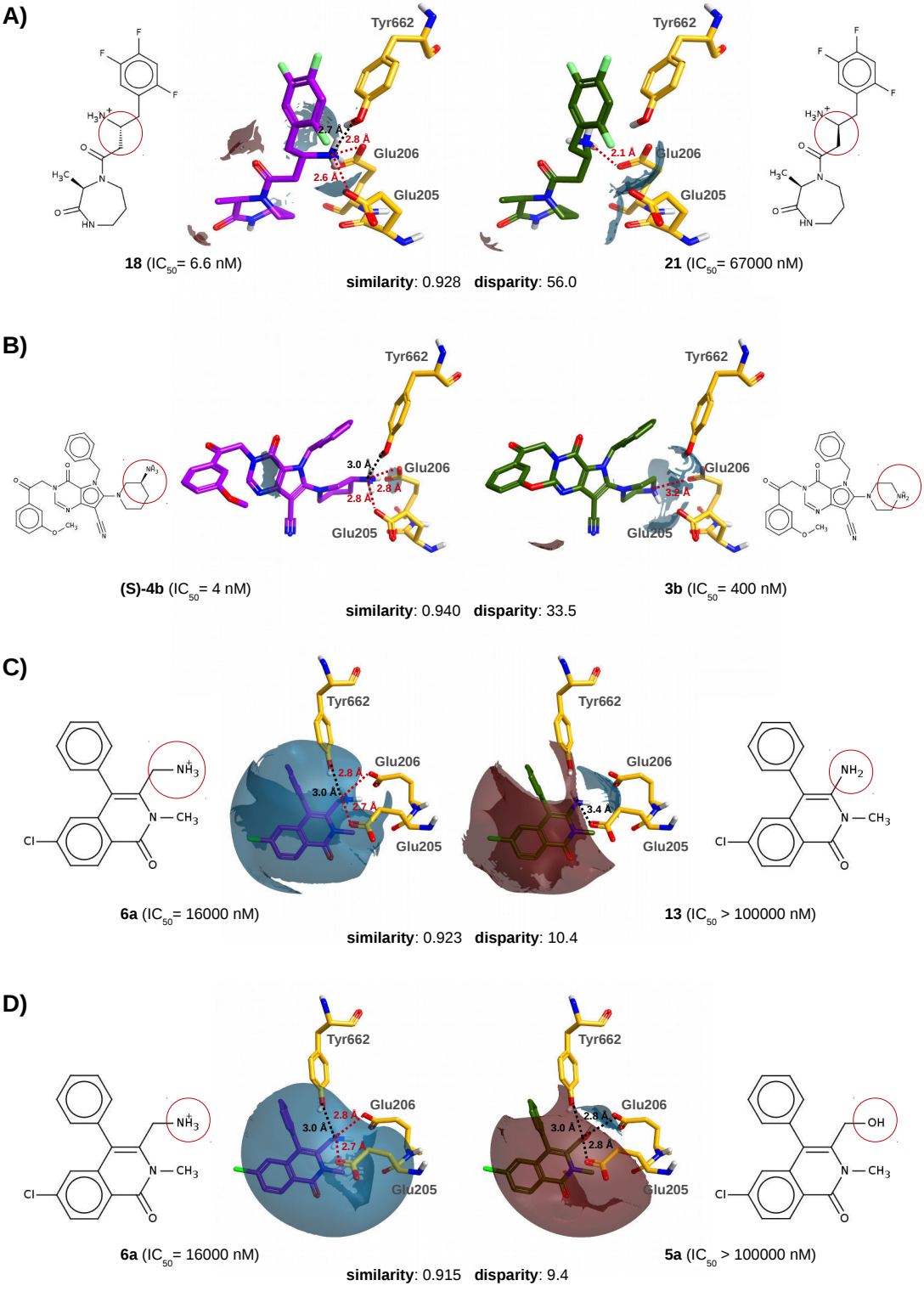
Glu205, Glu206 and Tyr662 (N-terminal recognition region)

The Glu205, Glu206 and Tyr662 residues form the N-terminal recognition region (see Figure 2) and, together with the S_1 pocket, are considered to be the most important anchor points for inhibitor recognition by DPP-IV.[19,67,68,89,153] Indeed it has already been shown that optimized interactions with these two key recognition motifs result in large gains in binding free energy, which can be further improved by additional favorable contacts to side chains that flank the active site.[68] The salt bridge is the strongest non-covalent interaction in nature and it is formed between two ionized sites that simultaneously make a hydrogen bond and an electrostatic interaction.[154] The Glu205, Glu206 and Tyr662 residues create a negative environment in the binding site of DPP-IV that favors the presence of a positively charged group (*e.g.* the NH_3^+ group) facing them and forming salt bridges with the Glu205/Glu206 dyad (see Figures 8 and S2). In this regard, the NH_2 group present in a large number of DPP-IV inhibitors (NH_3^+ in solution at physiological pH) simulates the N-terminus of the peptide that would normally bind to the binding site of DPP-IV (see Figure 2). Site-directed mutagenesis experiments[153,155] along with the fact that inhibitors of a different chemical nature establish salt bridges with Glu205/Glu206 and hydrogen bonds with Tyr662[52] have demonstrated that this interaction is essential for DPP-IV activity.[53,68,156] For instance, in the case of sitagliptin, the lack of interaction with these three residues results in a 25-fold loss of activity.[53] Therefore, in order to observe the importance for bioactivity of these intermolecular interactions with Glu205, Glu206 and Tyr662, we have compared different pairs of DPP-IV inhibitors whose structure differs only in the region facing these residues, focusing on how their bioactivities are affected by **(a)** their different capacity to form salt bridges/hydrogen

bonds with the N-terminal recognition region, and **(b)** the electrostatic surfaces they create in this area. We have observed activity cliffs subjected to changes of this NH_3^+ group in different situations: **(a)** when the configuration of the carbon containing the NH_3^+ group is switched from **R** to **S** (see Figures 8A,[139] S2B[139] and S2E[103]); **(b)** when the NH_3^+ group is replaced by a hydrogen atom and a charged secondary amine is introduced in the adjacent carbon (see Figures 8B, S2A, S2C, S2D and S2F);[142,146] **(c)** when the relative location of the amino substituent in the compound makes its protonation more difficult (see Figure 8C);[145] and **(d)** when the positive NH_3^+ group is replaced by an alcohol group that is also capable of making hydrogen bonds with Glu205, Glu206 and Tyr662 but not of producing the electrostatic interaction with the Glu205/Glu206 dyad (see Figure 8D).[145]

In short, for the N-terminal recognition region we have observed activity cliffs deriving from the loss of activity when this NH_3^+ group is either away from the Glu205/206 dyad [situation **(a)** in the previous paragraph] or not present [situations **(b)**, **(c)** and **(d)** in the previous paragraph]. This loss of activity is associated with either the presence of a more negative electrostatic environment facing the Glu205, Glu206 and Tyr662 residues (see Figures 8C and 8D) or the loss (or partial loss) of hydrogen bonds with Tyr662 and salt bridges with the Glu205/206 dyad (see Figures 8 and S2). Generating a positive electrostatic surface toward the Glu205/Glu206 residues and/or allowing the formation of salt bridges with Glu205/Glu206 and a hydrogen bond with Tyr662 therefore significantly increases the bioactivity of DPP-IV inhibitors (with associated disparity values that reach a maximum value of 56.0, see Figure 8A).

Table 1 shows that, in general, most of the pharmacophores used for the identification of DPP-IV inhibitors have a mandatory positive ionizable or a hydrogen bond donor site in the proper location for interacting with the N-terminal recognition region. Interestingly, the pharmacophore developed by Xing et al.[26] does not show any positive ionizable or hydrogen bond donor sites (see Figure 4 in that paper). Instead, the authors claim that in the two most active compounds "*the ionized amino group forms two hydrogen bonds with Glu205 and Glu206, which are the common features for most of DPP-4 inhibitors*".[26] This suggests that their pharmacophore should also contain the positive ionizable/hydrogen bond donor site that allows interaction with the N-terminal recognition region.



18 (IC_{50}= 6.6 nM) **21** (IC_{50}= 67000 nM)

similarity: 0.928 **disparity**: 56.0

(S)-4b (IC_{50}= 4 nM) **3b** (IC_{50}= 400 nM)

similarity: 0.940 **disparity**: 33.5

6a (IC_{50}= 16000 nM) **13** (IC_{50} > 100000 nM)

similarity: 0.923 **disparity**: 10.4

6a (IC_{50}= 16000 nM) **5a** (IC_{50} > 100000 nM)

similarity: 0.915 **disparity**: 9.4

Figure 8. Comparison of the distribution of electrostatic surfaces between pairs of compounds that differ in their interactions with residues Glu205, Glu206 and Tyr662. For each panel the compound with the highest activity is shown in purple on the left and the compound with the lowest activity is shown in green on the right. Molecules are labeled with the same names that identify them in the corresponding paper.[139,145,146] The negative and positive electrostatic surface differences are shown in garnet and blue respectively (where the default value – *i.e.,* 2.0 – was used as the threshold for the surface difference between each pair). Dotted lines represent either donor and acceptor atoms with the potential to form hydrogen bonds (in black) or atom pairs with the potential to form salt bridges (in red). In the 2D representation of each ligand, the structural differences between the compounds compared are highlighted. The different panels are arranged in order of decreasing disparity and correspond to different situations: in panel **A**, the configuration of the carbon containing the NH_3^+ group is switched from (R) in **18** to (S) in **21**;[139] in panel **B**, the NH_3^+ group is replaced by H and a positively charged secondary amine is introduced into the adjacent carbon;[146] in panel **C**, the substituent containing the amine group is shortened and the resulting amine is more difficult to protonate at pH = 7;[145] and in panel **D**, the positive NH_3^+ group is replaced by an alcohol group.[145] The ligand orientations are the result of their superposition with co-crystallized ligands from the same or very similar chemical series (*i.e.,* 2IIV[139] for panel **A**, 4A5S[146] for panel **B** and 3OPM[145] for panels **C** and **D**; residue locations in each panel are also taken from the corresponding PDB file). This figure was obtained with the help of the Forge[152] and MarvinSketch programs.[178]

There are also equivalent residues at DPP8 and DPP9 for the Glu205/Glu206 dyad (Glu275/Glu276 for DPP8 and Glu248/Glu249 for DPP9; see Figure 2) and also for Tyr662 (Tyr787 for DPP8 and Tyr762 for DPP9; see Figure 2) and, depending on the homology models used, the Glu dyad has been modeled either facing[101,102] or not facing[93,97] the DPP8/DPP9 binding site, whereas there are fewer differences for the Tyr residue (see Figure S1). Nevertheless, to our knowledge there is no experimental evidence that supports that the small differences in the N-terminal recognition region suggested by some homology models[93,97] have been successfully used to design selective DPP-IV inhibitors.

Arg125 (S_2 pocket)

It has been described that Arg125 is essential to coordinate the carbonyl group of the P_2 residue and, together with Glu205 and Glu206, align the substrate optimally for the nucleophilic attack by Ser630.[77] Moreover, SAR studies suggest that an electrostatic intermolecular interaction with the positively charged Arg125 results in an affinity gain for DPP-IV.[103,138,139,144,147] In this regard, an optimal protein-ligand fit might be reached by an inhibitor moiety that introduces a negative (or less positive) environment close to Arg125. Various strategies aimed at achieving this goal have been reported in the literature, including: **(a)** the introduction of a halogen in the ortho position of a phenyl ring that faces Arg125 and acts as an electron withdrawing substituent (see Figures 9B, S3A, S3B, S3C,

S3E, S3F, S3I and S3J) that lowers the positive environment which, when absent, is located close to Arg125;[103,147] **(b)** bringing an oxygen atom closer to the Arg125 sidechain [either by modulating the size of the ring bearing the oxygen (see Figures 9A and S3D),[144] introducing a methoxy substituent (see Figures S3H and S3K),[147] or changing the configuration of a chiral center (see Figure 9C)[138]]; and **(c)** the use of a pyrazol-1-ylmethyl substituent (see Figure S3G).[139] Thus, according to the SAR data available for these three strategies,[103,138,139,144,147] the main increase in bioactivity comes from adding the pyrazol-1-ylmethyl substituent (which involves a 483-fold increase in IC_{50} for compound **26** relative to compound **22**; see Figure S3G),[139] then introducing a halogen in the ortho position of a phenyl ring (which involves a 54-fold increase in IC_{50} for compound **3n** relative to compound **3a**; see Figure S3A),[147] and finally bringing an oxygen atom closer to the Arg125 sidechain (which involves a 33-fold increase in IC_{50} for compound **18** relative to compound **19**; see Figure 9C).[138]

Figure 4 shows that the equivalent area around Arg125 in DPP8/DPP9 (the so-called R-loop) can be modeled in very different ways depending on how the sequences are aligned relative to the sequence of the corresponding DPP-IV template. For instance, the Modbase models[101,102] have been built bearing in mind that the 117-129 segment in DPP-IV lacks equivalents in DPP8/DPP9 and consequently there is no counterpart for Arg125 in DPP8/DPP9 (see Figure 4A). In contrast, as Figure 4B shows, the homology models in Janardhan & Reddy[97] were built from a multiple alignment where **(1)** the R-loop of DPP-IV is well aligned with other segments of DPP8/DPP9, and **(2)** the counterpart for Arg125 is another Arg residue in DPP9 (Arg163) but Thr177 in DPP8. As a result, according to Janardhan & Reddy's models,[97] the R-loop has a similar 3D structure for the three enzymes and the basic environment caused by Arg125 is only lost at DPP8 (see Figure 4B). Finally, in the model described by Rummey & Metz,[94] the R-loop has not only equivalent sequence segments in the multialignments for DPP-IV, DPP8 and DPP9 but also basic counterparts for Arg125 (Lys174 for DPP8 and Arg163 for DPP9), and thus the R-loop was predicted to closely follow the DPP-IV conformation. An analysis of the available SAR data[103,138,139,144,147] suggests that, although the optimization of the intermolecular interaction with Arg125 increases inhibitor bioactivity for DPP-IV, this is not essential for the selectivity relative to DPP8/DPP9 because even the compounds that make non-optimized interactions with Arg125 (*e.g.*, **22**[103]) are selective relative to DPP8/DPP9 (data not shown). Therefore the

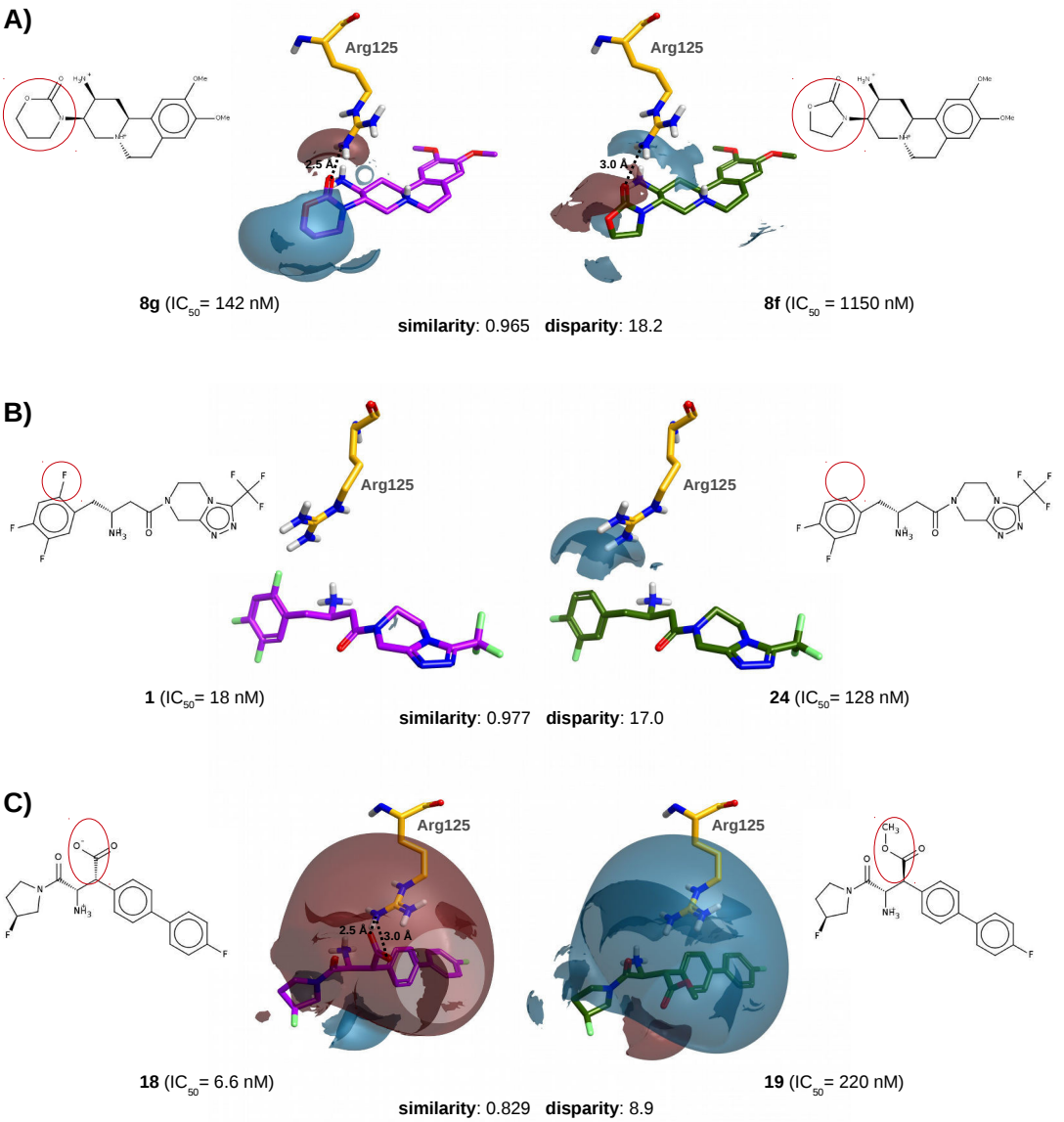


Figure 9. Comparison of the distribution of electrostatic surfaces between pairs of compounds that differ in their interactions with Arg125. For each panel the compound with the highest activity is shown in purple on the left and the compound with the lowest activity is shown in green on the right. Molecules are labeled with the same names that identify them in the corresponding paper.[103,138,144] The negative and positive electrostatic surface differences are shown in garnet and blue respectively (where the default value – *i.e.,* 2.0 – was used as the threshold for the surface difference between each pair). Dotted lines represent donor and acceptor atoms with the potential to form intermolecular hydrogen bonds. In the 2D representation of each ligand, the structural differences between the compounds compared are highlighted. The different panels are arranged in order of decreasing disparity and

correspond to different situations: in panel **A**, the ring size is increased;[144] in panel **B**, a halogen is added in the ortho position of the phenyl ring;[103] and in panel **C**, a carboxylic acid is placed near Arg125.[138] The ligand orientations are the result of their superposition with co-crystallized ligands from the same or very similar chemical series (*i.e.,* 3KWF[144] for panel **A**, 1X70[103] for panel **B** and 2FJP[173] for panel **C**; residue locations in each panel are also taken from the corresponding PDB file). This figure was obtained with the help of the Forge[152] and MarvinSketch programs.[178]

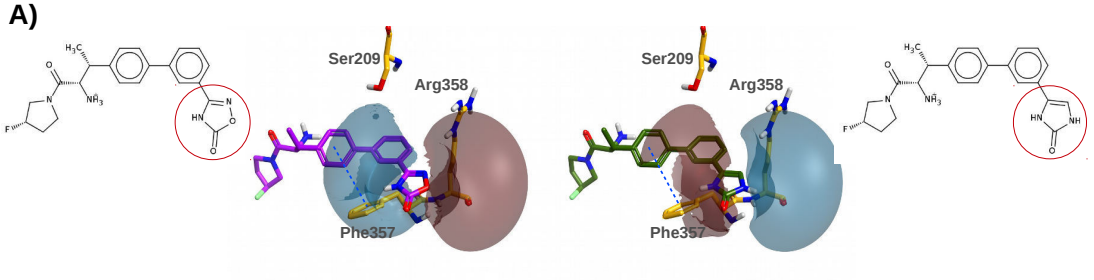
lack of relevance as regards the selectivity of the interactions with Arg125 suggests that, of the three sets of homology models described in this paragraph[94,97,101,102], the one that best describes the structure of the R-loop at DPP8/DPP9 according to the bioactivity data is Rummey & Metz's[94] because **(1)** the structure of the R-loop is very similar for DPP-IV and DPP8/DPP9, and **(2)** the residues equivalent to Arg125 at DPP8/DPP9 are also basic residues.

Phe357 and Arg358 (S_2 extensive subsite)

Figures 10 and S4 show how interactions with Phe357 and Arg358 can increase the activity of DPP-IV inhibitors. At this point it is worth remembering that the side chain of Arg358 is highly flexible (see Figure 3) and consequently, when doing the SAR analysis with the help of Activity Miner,[60] the position of this residue has been taken from the corresponding reference in the protein-ligand complex (see captions for Figures 10 and S4 for more details). There is, for instance, some variability in the examples in Figures 10 and S4 that result in either a larger (*e.g.* Figure S4G) or smaller (*e.g.* Figure S4C) S_2 extensive subsite.

Favoring a π-π interaction with Phe357 significantly increases ligand bioactivity. Some examples are: **(a)** the addition to compound **10** of different substituents (*i.e.,* 5-methoxy-2,3-dihydro-1H-indol-1-yl for **22e**, 2,3-dihydro-1H-indol-1-yl for **22c**, (5-cyanopyridin-2-yl)aminyl for **8** and [(4-cyanophenyl)methyl]aminyl for **15b**) that are associated with increases in bioactivity that range from 16- to 116-fold (see Figures S4G, S4H, S4I and S4J);[148] **(b)** the substitution of a morpholin-4-yl ring in **23** by the aromatic 4-pyridyl group to obtain ligand **24** (associated with a 25-fold increase in bioactivity; Figure S4F);[140] and **(c)** the replacement of a urea linker in **38** by a pyrimidine in **41** (associated with a 23-fold increase in bioactivity; Figure 10D).[140] More modest improvements of ligand bioactivity can be achieved by using alkyl groups to interact with Phe357. This can be seen

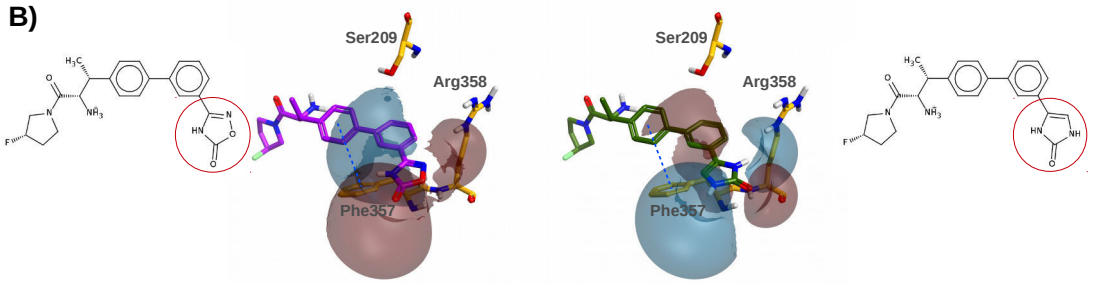
A)



44 (IC_{50}= 3 nM) **40** (IC_{50}= 110 nM)

similarity: 0.928 **disparity**: 21.6

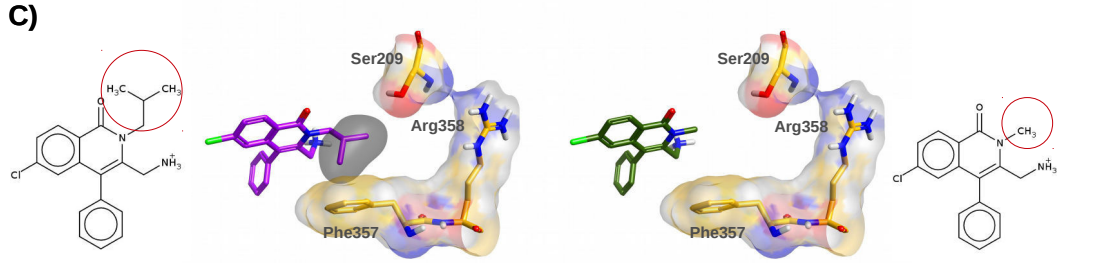
B)



44 (IC_{50}= 3 nM) **40** (IC_{50}= 110 nM)

similarity: 0.898 **disparity**: 15.3

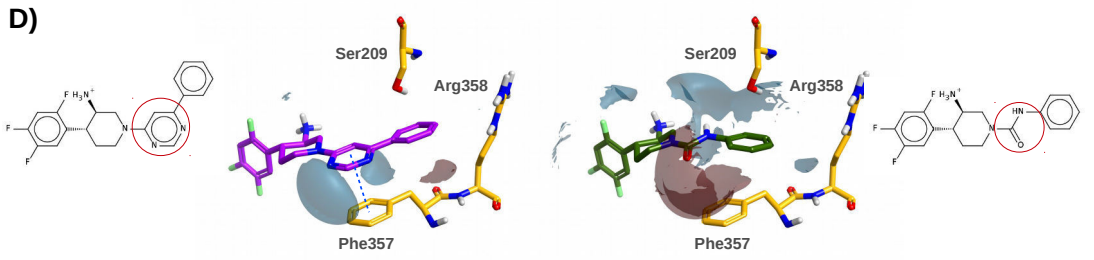
C)



6e (IC_{50}= 1600 nM) **6a** (IC_{50} = 16000 nM)

similarity: 0.927 **disparity**: 13.7

D)



41 (K_i = 6.1 nM) **38** (K_i = 140 nM)

similarity: 0.853 **disparity**: 9.2

Figure 10. Comparison of the distribution of electrostatic and hydrophobic surfaces between pairs of compounds that differ in their interactions with residues Ser209, Phe357 and Arg358. For each panel the compound with the highest activity is shown in purple on the left and the compound with the lowest activity is shown in green on the right. Molecules are labeled with the same names that identify them in the corresponding paper.[137,140,145] In panels **A**, **B** and **D**, the negative and positive electrostatic surface differences are shown in garnet and blue respectively. In panel **C**, the hydrophobic surface differences are shown in gray, while the protein surface has been colored according to atom color. The field surface difference is established by default at 2.0. In the 2D representation of each ligand, the structural differences between the compared compounds are highlighted. The different panels are arranged in order of decreasing disparity and correspond to different situations: in panels **A** and **B** a negative environment is placed around Arg358 irrespective of the orientation of the alignment between compounds **44** and **40**;[137] in panel **C**, a hydrophobic interaction is established with Phe357;[145] and in panel **D**, a π-π interaction is established with Phe357.[140] The blue dashed lines show the π-π interactions between the corresponding ligand and Phe357 and were calculated with the help of Maestro[165] using default options. The ligand orientations are the result of their superposition with co-crystallized ligands from the same or very similar chemical series (*i.e.,* 2FJP[173] for panels **A** and **B**, 3OPM[145] for panel **C** and 2OQV[140] for panel **D**; residue locations in each panel are also taken from the corresponding PDB file). This figure was obtained with the help of the Forge[152] and MarvinSketch programs.[178]

in compounds **6e** (see Figure 10C)[145] and **6g** (see Figure S4C),[145] where the inclusion of an isobutyl and a neopentyl group resulted in a 10- and 8-fold increase of DPP-IV potency relative to **6a**.

Placing a negative environment close to Arg358 also significantly improves ligand bioactivity. Some examples are: **(a)** replacing the trifluoromethoxy group in **14r** by a carboxylic acid in **14t** (associated with a 260-fold increase in bioactivity; see Figure S4A);[141] and **(b)** replacing the 2-oxo-1,3-dihydroimidazol-4-yl group at **40** by the 5-oxo-4H-1,2,4-oxadiazol-3-yl in **44** (associated with a 37-fold increase in bioactivity; see Figures 10A and 10B).[137] Interestingly, although the superposition of **40** and **44** can be done in two different ways, in both of them **44** places a negative environment around Arg358 (see Figures 10A and 10B).[137] It has been also described that placing electronegative groups on ligand aromatic rings near the positive charge of Arg358 have led to a 4-fold increase in affinity in sitagliptin.[57,68]

Another way of increasing ligand bioactivity can be achieved by improving the occupancy of the small cavity located between Arg358 and Ser209 with a hydrophobic substituent. For example, the replacement of a piperidin-1-yl group in **12s** by an aromatic ring in **12q** is associated with a 66-fold increase in bioactivity (see Figure S4B).[137] Other examples of this are compounds **1**, **26**, **22** and **24** (see Figures S4D and S4E),[103] which in all cases have a 5H,6H,7H,8H-[1,2,4]triazolo[4,3-a]pyrazin-7-yl substituent with the 1,2,4-triazole moiety making a π-π interaction with Phe357, but the additional presence of either a

difluoro(iodo)methyl or a trifluoromethyl group bound to the 1,2,4-triazole moiety brings about a 4-fold improvement in the inhibitory activity of **1** and **24** relative to **26** and **22**.

Regarding how interactions with Phe357 and Arg358 can influence the selectivity of DPP-IV inhibitors, it should be noted that, similarly to what happens with the R-loop, the sequence of the segment that goes from Ser349 to Glu362 in DPP-IV (which includes the so-called P_2-loop)[94] is very different in DPP8 and DPP9 and can therefore be modeled in very different ways in DPP8/DPP9 (depending on how their sequences are aligned relative to the P_2-loop sequence of the corresponding DPP-IV template; see Figure 5).[94,97,101,102] For instance, depending on the homology model used, there may either be no residue equivalent to Phe357[101,102] or there may be a His (*i.e.,* His434 for DPP8[94] and His424 for DPP9[94,97]) or a Cys (*i.e.,* Cys472 for DPP8).[97] Similarly, there may either be no residue equivalent to Arg358[101,102] or there may be an Asp (*i.e.,* Asp435[94] for DPP8 and Asp425 for DPP9[94,97]) or a Lys (*i.e.,* Lys473 for DPP8).[97] As a result, the interaction with the S_2 extensive subsite becomes the most important site governing selectivity as well as contributing to the achievement of nanomolar affinity,[52,57,68,79,82,97,105] because, regardless of what the correct alignment for the P_2-loop may be, the differences in this subsite are important enough to be exploited to achieve selectivity due to the fact that either **(a)** Phe357 and Arg358 could orient toward the DPP-IV binding site, thus favoring additional interactions with the ligands not possible in DPP8 or DPP9 (see Figure 5A), or **(b)** a substantial difference in the electrostatic environment of the S_2-pocket in the three proteins could be found (*i.e.,* replacing Phe357 by polar residues in DPP8/DPP9[94,97] and Arg358 by Asp in DPP8[94] and DPP9[94,97]; see Figure 5B). Moreover, a nearby residue (*i.e.,* Ser209; see Figure 5) is replaced in both models by either Asp278 (for DPP8)[97,101] or Asp251 (for DPP9),[97,102] creating a different charge environment in DPP8 and DPP9 relative to DPP-IV (whereas in Rummey & Metz models for DPP8/DPP9 there are no residues equivalent to Ser209)[94]. Thus all these differences between DPP-IV and DPP8/DPP9 show the potential of that region to be targeted to increase selectivity for DPP-IV.

Tyr547 (S_2/S_1' pockets and oxyanion hole)

The hydroxyl group of Tyr547 plays an oxyanion-stabilizing role in the catalytic mechanism of DPP-IV,[77,89] and therefore it is essential for the catalytic activity of the enzyme.[52,77,157] Figure

3F shows that Tyr547 can adopt two different conformations but, in contrast with previous studies,[68,86] the conformational change seems to be independent to the formation of π-π interactions with the ligand (see Figures 3A, 3B, 3C and 3D). Together with the phenyl ring of Phe357, interaction with the phenyl ring of Tyr547 is often sought to achieve nanomolar affinity, either by π-π interactions or by hydrophobic contacts with large aliphatic groups.[57,68] Indeed steered molecular dynamics simulations have shown that interactions with Tyr547 are important in preventing the inhibitor from leaving the active site[71] (which can contribute to the nM activity of the DPP-IV inhibitors that interact with it). It has also been shown that when using the nicotinic acid derivative that is co-crystallized with DPP-IV at 3O9V[158] as lead compound, it is possible to achieve a 10-fold improvement in DPP-IV bioactivity by introducing an aromatic ring into compound **13b** that is thought to occupy the hydrophobic pocket between Tyr547 and Trp629.[159] Unfortunately there are no other examples from SAR studies to enable us to analyze how changes in the ligand moieties close to Tyr547 affect ligand bioactivity.

Although Tyr547 is conserved in the three proteins (*i.e.,* Tyr669[97,101] or Tyr653[94] for DPP8 and Tyr644[94,97,102] for DPP9), a nearby residue in DPP-IV is mutated from Cys (*i.e.,* Cys551) to Gln (Gln673[97,101] or Gln657[94] in DPP8 and Gln648[94,97,102] in DPP9; see Figure 6). However, while in Janardhan & Reddy's models[97] these residues have been modeled following the same conformation as DPP-IV (Figure 6B), in ModBase models[101,102] the Gln residues occupy part of the binding site and thus could block the proper binding to DPP8/DPP9 of the DPP-IV inhibitors that make π-π interactions with Tyr547 (see Figure 6A). Other authors[160] have also suggested that Tyr547 is involved in inhibitor selectivity because the mobility of this residue is not the same in DPP8/DPP9 due to the replacement of Ser552 by a bulkier Val (*i.e.,* Val674[97,101] or Val658[94] for DPP8 and Val649[94,97,102] for DPP9). This would therefore open the possibility of exploiting these differences in order to find/design selective DPP-IV inhibitors. For instance, alogliptin, which shows excellent selectivity (see Table S1), interacts with Tyr547, the Glu dyad, the S_1 pocket and Arg125 (see Figure 3B).[161] As seen above, SAR data suggest that neither the Glu dyad nor Arg125 are involved in selectivity.[103,138,139,144,147] In the case of the S_1 pocket, the data given below show no significant differences in its size for DPP8/DPP9 relative to DPP-IV, and therefore the S_1 pocket seems not to be involved in selectivity either. Thus it could be concluded that the selectivity of alogliptin would be due to its interaction with Tyr547.

S_1 pocket

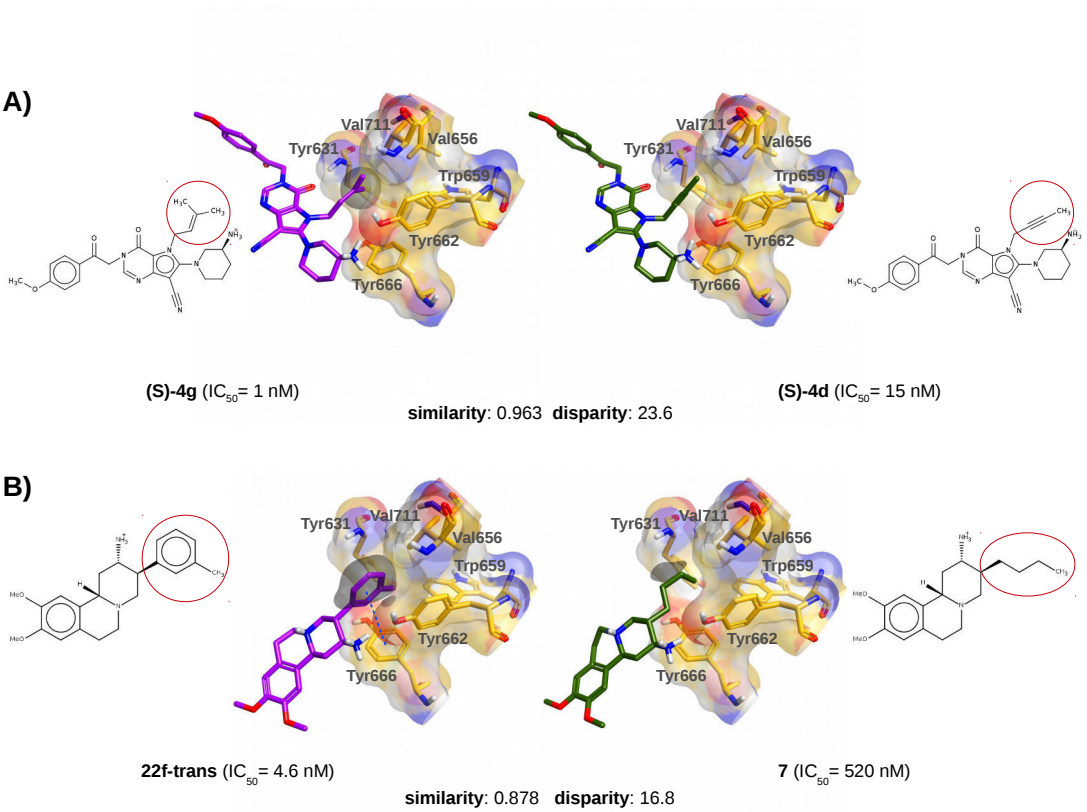


Figure 11. Comparison of the distribution of hydrophobic surfaces between pairs of compounds that differ in their interactions with the S_1 subsite. The compound with the highest activity is shown in purple on the left and the compound with the lowest activity is shown in green on the right. Molecules are labeled with the same names that identify them in the corresponding paper.[146,149] The hydrophobic surface difference is shown in gray, while the protein surface has been colored according to atom color. The field surface difference is established by default at 2.0. In the 2D representation of each ligand, the structural differences between the compared compounds are highlighted. The different panels are arranged in order of decreasing disparity. The blue dashed line shows the π-π interaction between the **22f-trans** ligand and Tyr666[149] (this π-π interaction was calculated with the help of Maestro[165] using default options). In panel **A** a but-2-yn-1-yl substituent in the S_1 subsite is replaced by a prenyl substituent,[146] while in panel **B** a monobutyl group is replaced by an m-tolyl group.[149] The ligand orientations are the result of their superposition with co-crystallized ligands from the same or very similar chemical series (*i.e.,* 4A5S[146] for panel **A** and 3KWF[144] for panel **B**; residue locationsa in each panel are also taken from this PDB file). This figure was obtained with the help of the Forge[152] and MarvinSketch programs.[178]

The lipophilic S_1 pocket is considered a crucial molecular anchor point for DPP-IV

inhibitors,[68] and the residues that constitute this pocket are conserved among the peptidases DPP-IV, DPP8 and DPP9.[81,97] Figures 11 and S5 describe two different ways for increasing the bioactivity of DPP-IV inhibitors through hydrophobic interactions with the S_1 pocket: **(a)** replacing a but-2-yn-1-yl substituent by a prenyl group (with associated improvements in bioactivity in a 1.7 to 125-fold range; see Figures 11A, S5A, S5B, S5C, S5D, S5E, S5F, S5G and S5H);[142,146] and **(b)** replacing a monobutyl substituent by either an m-tolyl or a phenyl group (with respectively 113- and 2.6-fold associated improvements in bioactivity; see Figures 11B and S5I).[149] In all the comparisons in Figures 11 and S5 we observe a tendency in which those compounds presenting a better occupancy of the S_1 pocket achieved higher bioactivities. Interestingly, a comparison of compounds **22f-trans**, **13** and **7** (which only differ in the substituent of the piperidin-4-aminium moiety; see Figures 11B and S5I) shows how introducing a π-π interaction with Tyr666 contributes modestly to improving bioactivity (2.6-fold when comparing compounds **13** and **7**; see Figure S5I) in comparison to filling the S_1 pocket better using a methyl substituent added to the phenyl ring (43.5-fold when comparing compounds **22f-trans** and **13**; see Figures 11B and S5I). This demonstrates that full occupation of the hydrophobic S_1 pocket plays a role in the determination of DPP-IV activity.[68,150] In fact all the crystallized ligands in the PDB occupy the S_1 pocket, with most of them showing very few changes in the size and shape of the ligand moiety in this place (see Table S3).

Some authors have suggested that the S_1 pocket is significantly smaller in DPP-IV (27.72 $Å^3$) than it is in DPP8 (99.77 $Å^3$) and DPP9 (75.89 $Å^3$).[97] Nevertheless, an analysis of the homology models available for DPP8/DPP9[97,101,102] show that these differences are caused by considering certain residues as part of the S_1 pocket in DPP8 and DPP9 and but not considering the equivalent ones in DPP-IV (see Figure 7).[97] In this regard, whereas Gly670/Gly645, Val674/Val649, Leu676/Leu651, Trp754/Trp729, Tyr787/Tyr762, Arg794/Arg769 and Tyr795/Tyr770 are included as part of the S_1 in DPP8/DPP9, their spatial counterparts in DPP-IV (*i.e.,* Ala548, Ser552, Lys554, Trp629, Tyr662, Arg669 and Tyr670) are not.[97] All this would suggest that, irrespective of the homology models used for DPP8/DPP9, there are no significant differences in either the size or the electrostatic potential of S_1 between DPP-IV, DPP8 and DPP9 (see Figure 7). Similarly, other papers that describe homology models for DPP8 and DPP9 find no significant differences between S_1 sizes for the three enzymes.[81,93–95] To our knowledge, few investigations have carried out

SAR studies to discern whether or not the S_1 pocket is important for selectivity.[28,104,137,162] Some of these studies replaced the moiety of the lead compound that is thought to bind to the S_1 with a larger substituent and found a marked decrease in DPP-IV bioactivity and an improvement in the relationship of the DPP8/DPP-IV and/or DPP9/DPP-IV bioactivities (see Figure S6).[28,104] These results may suggest that the decrease in DPP-IV bioactivity is related to the smaller size of the S_1 pocket in DPP-IV relative to DPP8/DPP9. Nevertheless, Figure S6 shows that the 2,3-dihydro-1H-isoindol-2-yl moiety of **4** and the 1,2,3,4-tetrahydroisoquinolin-2-yl moiety of **7** can be also accommodated in the rigid S_1 pocket of DPP-IV (see Figure 3), and therefore S_1 size seems not to be involved in the bioactivity differences between **1** and **4** and between **3** and **7**. Considering all this information together, we can conclude that there is no clear evidence to suggest that the S_1 pocket plays a role in the selectivity of DPP-IV inhibitors.

Lys554

More recently, a new mechanism to enhance the bioactivities of DPP-IV inhibitors has been discovered, consisting of establishing interactions with residue Lys554.[145,163] Figures 12 and S7 show three different situations described in the literature that lead to increased DPP-IV activity in this manner: **(a)** the introduction of a carboxylic acid (see Figure 12A),[150] **(b)** the introduction of a methanesulfonyl group (see Figures 12B and S7A),[124] and **(c)** the introduction of a substituent that ends in a carbamoyl group (see Figures 12C and S7B).[145] In the case of compound **8n**, adding a carboxylic acid in the right location to make a salt bridge interaction with Lys554 results in a 213-fold increase in bioactivity relative to compound **8f** (which has a methyl instead of a carboxylic acid; see Figure 12A). In the case of compounds with a methanesulfonyl group, when these are compared to a compound with another substituent (see Figures 12B and S7A, in which the methanesulfonyl group is replaced by either a trifluoromethyl or a trifluoromethylsulfonyl group respectively), the outcome is the generation of a negative electrostatic surface toward Lys554 created by the methanesulfonyl group. For instance, a 70-fold increase in bioactivity is obtained in this way by compound **26** relative to **21** (see Figure 12B).[124] In the case of compounds with a substituent that ended with a carbamoyl group, these not only created a negative electrostatic surface toward Lys554 but also used the carbamoyl oxygen as an acceptor in a hydrogen bond with Lys554

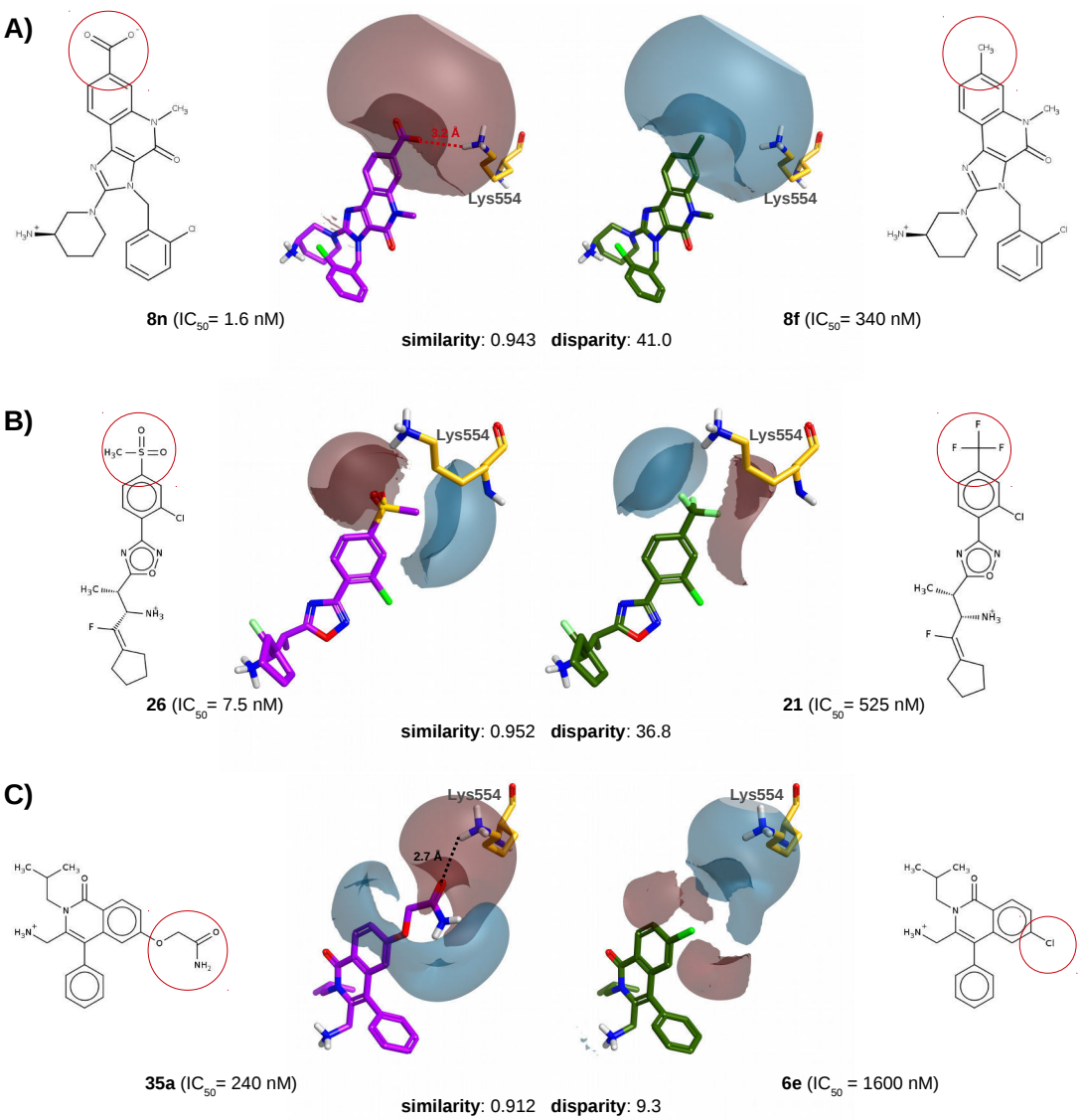


Figure 12. Comparison of the distribution of electrostatic surfaces between pairs of compounds that differ in their interactions with residue Lys554. For each panel, the compound with the highest activity is shown in purple on the left and the compound with the lowest activity is shown in green on the right. Molecules are labeled with the same names that identify them in the corresponding paper.[124,145,150] The negative and positive electrostatic surface differences are shown in garnet and blue respectively (where the default value – *i.e.,* 2.0 – was used as the threshold for the surface difference between each pair). In the 2D representation of each ligand, the structural differences between the compared compounds are highlighted. The different panels are arranged in order of decreasing disparity and correspond to different situations: in panel **A**, a carboxylic acid replaces a methyl group;[150] in panel **B**, a methanesulfonyl replaces a trifluoromethyl group;[124] and in panel **C**, a

(carbamoylmethyl)oxidanyl group replaces a chlorine group.[145] The salt bridge between Lys554 and compound **8n** in panel **A** is shown as a red dotted line, while the hydrogen bond with compound **35a** in panel **C** is shown as a black dotted line. The ligand orientations are the result of their superposition with co-crystallized ligands from the same or very similar chemical series (*i.e.,* 3G0D[161] for panel **A**, 3C45[124] for panel **B** and 3OPM[145] for panel **C**; residue locations in each panel are also taken from the corresponding PDB file). This figure was obtained with the help of the Forge[152] and MarvinSketch programs.[178]

(features that are absent in those compounds that do not have a substituent able to reach Lys554; see Figures 12C and S7B). In that case, however, the comparison between **35a** and **6e** (see Figure 12C)[145] or between **23a** and **6e** (see Figure S7B)[145] shows that the enhancement of bioactivity is more modest (6-7 fold) than when introducing either a carboxylic acid or methanesulfonyl group. These three strategies therefore result in an increase in the negative electrostatic surface oriented toward the positively charged moiety of the Lys554 sidechain, thereby favoring intermolecular interactions between the corresponding ligand and DPP-IV.

Interestingly, different homology models[94,97,101,102] suggest that the equivalent residue in DPP8/DPP9 is a Leu residue (*i.e.,* Leu676[97,101] or Leu660[94] in DPP8 and Leu651[94,97,101] in DPP9). Therefore the charged environment provided by Lys at the DPP-IV binding site is absent from them, and consequently this difference can be exploited to design potent and selective DPP-IV inhibitors. Unfortunately, none of the SAR studies that show how the interaction with Lys554 improves DPP-IV potency provides data on DPP8/DPP9 selectivity to enable us to confirm this.[124,145,150,158,163]

Trp629 (S_2' pocket)

Another feature that is also explored in various inhibitor series is the ability of some compounds to extend to the S_2' subsite of DPP-IV. This strategy is, for instance, used by the drug linagliptin, which forms a π-π interaction with Trp629, achieving a very high bioactivity value (see Table S1).[142] With the aim of determining the importance of exploring the S_2' pocket, we have compared the compounds that extend to this subsite with compounds of the same congeneric series that do not have a substituent able to reach it. In order to observe the differences between these compounds, we have represented the differences between their respective hydrophobic surfaces.

Although it is true that higher bioactivities can be accomplished by reaching the S_2' subsite (see Figures 13A and S8),[142,146] in some cases this may result in a huge decrease in activity (see Figure 13B)[142] that could be related to a possible conformational change for Trp629 (see Figure 3E), which would produce a steric hindrance with the ligand. Moreover, even though compounds can be optimized by their extension to the S_2' subsite, this is not a premise for obtaining compounds with bioactivities in the nM range, as exemplified by compounds **(S)-4l** (IC_{50} = 9 nM; see Figure 13A) and **6ac** (IC_{50} = 35 nM; see Figure 13B).

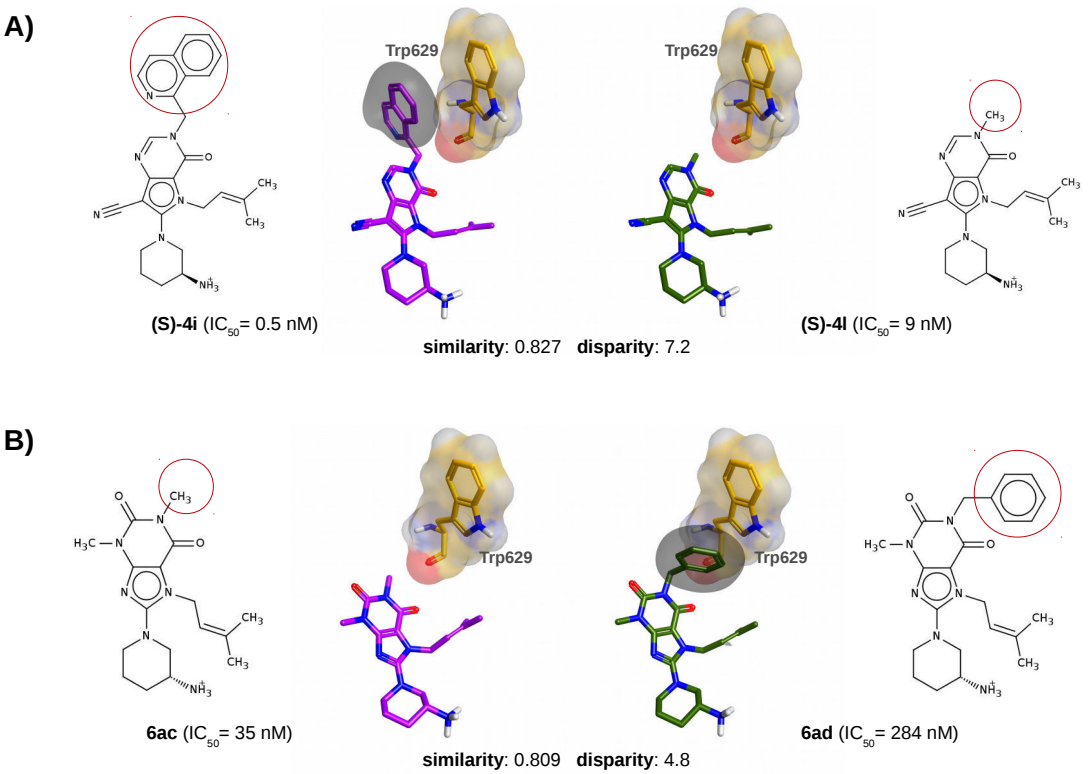


Figure 13. Comparison of the distribution of hydrophobic surfaces between pairs of compounds that differ in their interactions with the S_2' subsite. For each panel, the compound with the highest activity is shown in purple on the left and the compound with the lowest activity is shown in green on the right. Molecules are labeled with the same names that identify them in the corresponding paper.[142,146] The hydrophobic surface differences are shown in gray, while the protein surface has been colored according to atom color. The field surface difference is established by default at 2.0. In the 2D representation of each ligand, the structural differences between the compared compounds are highlighted. The different panels are arranged in order of decreasing disparity and correspond to different situations: in panel **A**, the extension to the S_2' pocket correlates with an improvement in DPP-IV activity;[146] in panel

B, the extension to the S_2' pocket correlates with a worsening in DPP-IV activity.[142] The ligand orientations are the result of their superposition with co-crystallized ligands from the same or very similar chemical series (*i.e.,* 4A5S[146] for panel **A** and 2RGU[142] for panel **B**; residue locations in each panel are also taken from the corresponding PDB file). This figure was obtained with the help of the Forge[152] and MarvinSketch programs.[178]

Therefore, taking into account that the dynamics that govern conformational changes on Trp629 are not well understood (see Figure 3), the fact that nM can be achieved without interacting with the S_2' pocket and the risk involved in requiring a compound to reach the S_2' subsite, we would suggest that this feature is more promising from a drug design perspective (to obtain higher bioactivities based on a lead compound or even for fragment-screening purposes) than for VS purposes (where the requirement of an aromatic ring pharmacophore site at the S_2' pocket would produce false positives).

7. How to obtain potent and selective DPP-IV inhibitors in a virtual screening following the analysis of SAR studies

From the analysis of previous SAR studies, we have reached several conclusions regarding how to obtain DPP-IV inhibitors with high bioactivity values (see Table 3). In order to evaluate whether these rules have been considered (at least implicitly) in previous VS searching for DPP-IV inhibitors, we have looked at the most potent DPP-IV inhibitors identified by VS methods (see Table 1), analyzed how they interact with the binding site, and proposed how to optimize some of them in accordance with the rules set out above. Since the selectivity against DPP8 and DPP9 has not been evaluated by biological assays in any of these studies, we have also tried to hypothesize whether or not the hit compounds analyzed are selective for DPP-IV.

- **HWL-892**[26] is the most potent DPP-IV ever identified by means of a VS (IC_{50} = 0.148 μM; see Table 1). We hypothesize that compound **HWL-892** is selective since it targets the S_2 extensive subsite by a π-π interaction between the phenyl moiety of its 1,2,3,4-tetrahydroisoquinolin group and Phe357 as well as by the formation of a hydrogen bond with Arg358 by one of its methoxy groups. Despite the fact that its bioactivity is already high, we suggest that the placement of a halogen in the ortho position of the phenyl group may establish an additional electrostatic interaction with Arg125 and improve its inhibitory potency even more.

Table 3. Eight simple rules when searching for potent and selective DPP-IV inhibitors through virtual screening.

N-terminal recognition region

Place a positively charged donor group (preferably a primary amine) that can establish a salt bridge interaction with the Glu205 and the Glu206 dyad and a hydrogen bond with the Tyr662 hydroxyl group.

S_1 pocket

Place a group (preferably an aromatic ring) that can establish the maximum number of hydrophobic contacts with the S_1 pocket.

S_2 pocket

Place a negatively charged group (with a partial or net charge), an acceptor group or a phenyl ring with a halogen substituent in the ortho position that can establish an electrostatic interaction with Arg125.

Place an aromatic ring that can establish π-π interactions with Phe357 so as to increase both DPP-IV activity and selectivity against DPP8 and DPP9.

Place a negatively charged group (with a partial or net charge) that can establish an electrostatic interaction with Arg358 so as to increase both DPP-IV activity and selectivity against DPP8 and DPP9.

Oxyanion-hole (Tyr547)

Place an aromatic ring near Tyr547 to form additional π-π interactions and further increase activity and selectivity.

Lys554

Place a negatively charged group (with a partial or net charge) that can establish an electrostatic interaction with Lys554.

S_2' pocket (Trp629)

Do not reach the S_2' pocket since this may result in a decrease in activity.

- Compound **MDPI-12398**[27] uses a phenyl substituent to occupy the S_1 pocket and one hydroxyl and two positively charged amino groups to form three salt bridges and one hydrogen bond with Glu205 and Glu206. Moreover, its primary amine forms an additional hydrogen bond with Tyr662 and its 4-hydroxy-1-methyl-2-oxoquinolin-3-yl moiety forms a π-π interaction with Phe357. Collectively, this large number of

interactions with the Glu dyad, the S_1 hydrophobic pocket and the S_2 extensive subsite may explain its significant bioactivity (IC_{50} = 0.73 μM; see Table 1). According to the predicted π-π interaction with Phe357, this compound is expected to be selective for DPP-IV. In order to obtain **MDPI-12398** derivatives with increased bioactivity, we would suggest placing **(a)** an electron withdrawing substituent in the ortho position of the phenyl substituent to favor interaction with Arg125, and **(b)** a negative or hydrophobic group bound to the phenyl ring of the 4-hydroxy-1-methyl-2-oxoquinolin-3-yl moiety to improve either the electrostatic interaction with Arg358 or the occupancy of the small cavity located between Arg358 and Ser209.

- **Gemifloxacin**[21] contains a primary amine that is able to establish a salt bridge with the Glu206 and a hydrogen bond with the Tyr662 hydroxyl group. In addition, a highly electronegative aromatic fluoro substituent is oriented toward Arg125, which creates a favorable electrostatic environment and could act as a hydrogen bond acceptor. As well as this, the carboxylic acid moiety makes a hydrogen bond with Gln553 and is possibly involved in a charge-charge interaction with Lys554. The achievement of these interactions supports its activity value (*i.e.,* IC_{50} = 1.12 μM; see Table 1), but we believe that this could be further improved by optimally filling the S_1 hydrophobic pocket for a substantial gain in activity and incorporating a substituent able to interact with Phe357 and Arg358 at the S_2 extensive subsite (which, apart from improving its potency, would provide it with selectivity).

- Compound **1**[23,114] belongs to the SPECS database and has been identified by different VS protocols[23,114] resulting in different hypothetical binding modes obtained from docking. The same biological assay was applied in both studies to determine the *in vitro* activity of the compound, and two IC_{50} values were reported (IC_{50} = 2.12 μM and IC_{50} = 5.77 μM; see Table 1). Two different docked poses were suggested that accomplish the same pharmacophore. In the first docked pose, the Glu dyad is targeted by the amine at the amide group,[114] whereas in the second it is targeted by the hydroxyl group.[23] In contrast, the S_1 and S_2 sites in both poses are filled by the same groups (the 4,5,6,7-tetrahydro-1H-indazol-3-yl ring fills the S_1, while the naphthalene ring fills the S_2). Moreover, in the second pose the hydroxyl group of the

naphthalene ring is hydrogen-bonded to Arg125. Bearing in mind that the two suggested binding modes involve a hydrophobic interaction of the naphthalene ring with Phe357, we would suggest that this compound could be selective for DPP-IV. In order to use compound **1** as a lead molecule to obtain more potent DPP-IV inhibitors, we suggest **(1)** favoring the first binding mode by replacing the amine at the amide group by a carbon atom bound to a primary amine (which would improve interactions with the N-terminal recognition region through salt bridges with the Glu dyad and a hydrogen bond with Tyr662), and **(2)** adding a substituent to the naphthalene ring to favor ligand interaction with Arg358 by either placing a negative environment close to Arg358 or improving the occupancy of the small cavity located between Arg358 and Ser209 with a hydrophobic substituent.

- Compound **7a**[20] is a biaryl fragment that has been docked in the hydrophobic S_1 pocket. Its positively charged $-NH_3^+$ group is able to make two salt bridges with the Glu dyad. These interactions explain the bioactivity value (IC_{50} = 2.3 μM; see Table 1). Because of its small size, this fragment is not able to reach other relevant protein residues such as Phe357 and Arg358 and is therefore not expected to have DPP-IV selectivity. In order to increase its bioactivity (and selectivity), we would suggest adding a substituent that could reach either the S_2 extensive subsite or Tyr547.

- **NCI0211295** was not the most active DPP-IV inhibitor found in the VS, but it was the most potent of the five new DPP-IV inhibitors discovered by Al-masri *et al.*[25] Although their paper reports that all the identified DPP-IV inhibitors accomplish the pharmacophore, no information is provided to allow us to infer how this ligand binds to DPP-IV (*e.g.* **NCI0211295** has no positively charged group at neutral pH), and therefore we cannot suggest how its bioactivity might be improved and infer whether or not it is selective.

- Compound **4**[22] has a 2H-1-benzopyran-2-one ring that makes hydrophobic contacts in the S_1 pocket but lacks interactions with the Glu dyad. However, there are other intermolecular interactions that may explain the preservation of its inhibitory activity, such as **(a)** a second hydrophobic interaction between the 2H-1-benzopyran-2-one ring and Tyr547, **(b)** the hydrogen bonds between the 2H-1-benzopyran-2-one ring

oxygen and the Tyr547 and Ser630 hydroxyls, and **(c)** the hydrogen bond between the hydroxymethyl group and Arg125. This compound is expected to be selective for DPP-IV because in the pose initially suggested it interacts with Tyr547 (see Table 1),[22] whereas in a new pose that has just been suggested for a very close derivative it additionally interacts with the S_2 extensive pocket.[48] The same compound **4** derivative has also been used as a lead compound to obtain more potent DPP-IV inhibitors.[48] For instance, the replacement of the three substituents of the phenyl moiety by fluoro groups and of the hydroxymethyl by an amino group able to make salt bridges with the Glu dyad results in a 91-fold improvement in potency (from 14.13 μM to 155 nM for compound **8a**).[48] Interestingly, these changes result in a radically new orientation of the derivative that locates the 2H-1-benzopyran-2-one ring in the S_2 and the phenyl moiety in the S_1.[48]

- The **NCI0294730** optimal pose is shown in Figure 7 in the original paper.[21] According to this proposed binding mode, the compound does not interact with either the S_1 pocket or the Glu dyad. Instead, Table 1 shows how this is compensated by interactions with Trp629 and the S_2 pocket (via a salt bridge from one of the two carboxyl groups with Arg125, a hydrophobic interaction between one of the two cyclopentene moieties and Phe357, another hydrophobic interaction between the second cyclopentene moiety and the Tyr547 aromatic ring, and a hydrogen bond between the Tyr547 hydroxyl and the carbonyl oxygen at one of the two amide bonds). As a result, 34% DPP-IV inhibition was observed at 10μM,[21] which shows that the interactions with the S_2 pocket were able to achieve similar bioactivity relative to other compounds that interact with the Glu dyad and the S_1 subsite (see compound **14** in Table 1).[19] Moreover, due to the interactions with Phe357 and Tyr547, **NCI0294730** is expected to be selective relative to DPP8 and DPP9. Therefore, in order to obtain derivatives of **NCI0294730** with increased bioactivity we would suggest introducing a primary amine able to make salt bridges with the Glu dyad and an aromatic group (*e.g.* a phenyl) that fits well in the S_1 pocket.

- Compound **14**[19] is relatively small and only requires a hydrophobic interaction with the S_1 pocket and a salt bridge with the Glu dyad to achieve a basal inhibitory bioactivity (*i.e.,* 81.9% DPP-IV inhibition at 30μM; see Table 1). Unfortunately, the

paper in which compound **14** was identified does not show which of the two possible binding modes is the one adopted by this DPP-IV inhibitor (which makes it difficult to predict how to use compound **14** as a lead).[19] Nevertheless, neither of the two binding modes shows interactions with either the S_2 extensive subsite or Tyr547, and we can therefore conclude that compound **14** is not selective against DPP8 and DPP9.

- Compound **C5**[24] forms a salt bridge with Glu206 via its tertiary amine and uses its monobutyl chain to interact with the S_1 pocket through hydrophobic interactions with Tyr662, Tyr666 and Val711. Therefore, in both cases its intermolecular interaction with the Glu dyad and the S_1 pocket is not optimal and requires an additional π-π interaction between its chromene ring and Phe357 to achieve an IC_{50} of 61.55 μM (see Table 1). Moreover, this compound is predicted to be selective against DPP8 and DPP9 because of its π-π interaction with Phe357. In order to optimize the compound, it has been suggested that the original monobutyl chain should be replaced by a group that improves interactions with the S_1 pocket and a substituent added that could interact with Arg358.[24]

To summarize, 7 out of the 10 hit compounds (**HWL-892**, **MDPI-12398**, compound **1**, compound **7a**, **NCI0211295**, compound **14** and compound **C5**) present DPP-IV activity achieved by interactions with the hydrophobic S_1 pocket and salt bridges or hydrogen bonds with the Glu dyad.[24,58,59] Despite the importance of the S_2 extensive subsite for selectivity, this cavity was only exploited by 5 out of the 10 hit compounds (**HWL-892**, **MDPI-12398**, compound **1**, **NCI0294730** and compound **C5**).

8. Conclusions

Using our approach we have identified activity cliffs by focusing exclusively on comparisons between pairs of compounds **(1)** with big differences in their DPP-IV bioactivity that are not the product of steric hindrances with the protein, and **(2)** that differ in only one substituent. The protein environment of each pair was then carefully inspected and an explanation for the change in activity proposed based on the differences between the compounds and their intermolecular interaction with the protein environment. The robustness of our conclusions

has been verified with supporting evidence from several independent studies (using results deriving from different series) and bibliographical data. To our knowledge, only certain studies have so far looked at DPP-IV inhibition from the receptor point of view, instead either classifying DPP-IV inhibitors on the basis of which subsites they occupy[65] or describing specific molecular recognition interactions from crystal structure data[68] or quantifying electrostatic and hydrophobic interactions with binding site residues[164] (but nevertheless, the approach they use is different to the one used by us). However (and not considered by the three studies cited),[65,68,164] the relative importance for selectivity of the intermolecular interactions between DPP-IV and DPP-IV inhibitors was evaluated here by correlating data from SAR studies with available homology models for DPP8 and DPP9.[94,97,101,102] All this has allowed us to draw conclusions about which interactions are important for improving DPP-IV activity and selectivity and to favor early recognition of potent and selective DPP-IV during VS (see Table 3). As far as we know, only one review offers a summary of design clues for enhancing DPP-IV potency and selectivity but limits itself to the Glu dyad, the S_1 pocket and Phe357/Arg358 (S_2 extensive subsite) without considering the role of Arg125 (S_2 pocket), Tyr547 (S_2/S_1' pockets and oxyanion hole), Lys554 and Trp629 (S_2' pocket).[90]

Finally, we would note that focusing on the protein environment and finding out about receptor-ligand interaction from a binding site perspective is crucial in those situations where a compound has to be used *as is* (like in the case of natural ingredients to be used as bioactive ingredients in functional foods or dietary supplements).[61]

Acknowledgements

This study was supported by research grants 2014PFR-URV-B2-67 and 2015PFR-URV-B2-67 from our University. AG's contract is supported by grant 2015FI_B00655 from the Catalonia Government. We thank Cresset BioMolecular Discovery Ltd. and ChemAxon Ltd. for generously providing us with a software bursary for using their programs. We would also like to thank Prof. Sridhara Janardhan for kindly supplying us with his homology models for DPP8 and DPP9 and Prof. Maribel Matheu for her advice on organic nomenclature. This manuscript has been edited following the English language usage of our University.

References

1. International Diabetes Federation. *IDF Diabetes Atlas Seventh Edition 2015*. *International Diabetes Federation* (2015).

2. World Health Organization. *Global report on diabetes*. (2016).

3. Mittermayer, F. *et al.* Addressing unmet medical needs in type 2 diabetes: a narrative review of drugs under development. *Curr Diabetes Rev* **11,** 17–31 (2015).

4. Scheen, A. J. Cardiovascular effects of gliptins. *Nat. Rev. Cardiol.* **10,** 73–84 (2013).

5. Avogaro, A., de Kreutzenberg, S. & Fadini, G. P. Dipeptidyl-peptidase 4 inhibition: Linking metabolic control to cardiovascular protection. *Curr. Pharm. Des.* **20,** 2387-94 (2013).

6. Kwok, A. J., Mashar, M., Khavandi, K. & Sabir, I. DPP-IV inhibitors: Beyond glycaemic control? *Trends Cardiovasc. Med.* **24,** 157–164 (2014).

7. Chinda, K., Sanit, J., Chattipakorn, S. & Chattipakorn, N. Dipeptidyl peptidase-4 inhibitor reduces infarct size and preserves cardiac function via mitochondrial protection in ischaemia-reperfusion rat heart. *Diab. Vasc. Dis. Res.* **11,** 75–83 (2014).

8. Avogaro, A., Vigili de Kreutzenberg, S. & Fadini, G. P. Cardiovascular actions of GLP-1 and incretin-based pharmacotherapy. *Curr. Diab. Rep.* **114,** 1788-803 (2014).

9. Wu, S., Hopper, I., Skiba, M. & Krum, H. Dipeptidyl peptidase-4 inhibitors and cardiovascular outcomes: Meta-analysis of randomized clinical trials with 55,141 participants. *Cardiovasc. Ther.* **32,** 147–158 (2014).

10. Ussher, J. R. & Drucker, D. J. Cardiovascular actions of incretin-based therapies. *Circ. Res.* **114,** 1788–1803 (2014).

11. Salles, T., Santos, L., Barauna, V. & Girardi, A. Potential role of dipeptidyl peptidase IV in the pathophysiology of heart failure. *Int. J. Mol. Sci.* **16,** 4226–4249 (2015).

12. Barkas, F., Elisaf, M., Tsimihodimos, V. & Milionis, H. Dipeptidyl peptidase-4 inhibitors and protection against stroke: A systematic review and meta-analysis. *Diabetes Metab.* **43,** 1-8(2016).

13. Dokken, B. Mechanisms of cardiovascular injury in type 2 diabetes and potential effects of dipeptidyl peptidase-4 inhibition. *J. Cardiovasc. Nurs.* **31,** 274–283 (2016).

14. Kubota, A. *et al.* DPP-4 inhibition has beneficial effects on the heart after myocardial infarction. *J. Mol. Cell. Cardiol.* **91,** 72–80 (2016).

15. Kuramitsu, S. *et al.* Effect of sitagliptin on plaque changes in coronary artery following acute

coronary syndrome in diabetic patients: The ESPECIAL-ACS study. *J. Cardiol.* **69,** 369–376 (2017).

16. Deacon, C. F. & Lebovitz, H. E. A comparative review of DPP-4 inhibitors and sulphonylureas. *Diabetes. Obes. Metab.* **18,** 333-47(2015).

17. Kushwaha, R. N., Haq, W. & Katti, S. B. Discovery of 17 gliptins in 17-years of research for the treatment of type 2 diabetes: A synthetic overview. *Chem. Biol. Interface* **4,** 137–162 (2014).

18. Rose, P. W. *et al.* The RCSB Protein Data Bank: New resources for research and education. *Nucleic Acids Res.* **41,** D475-82 (2013).

19. Ward, R. A., Perkins, T. D. J. & Stafford, J. Structure-based virtual screening for low molecular weight chemical starting points for dipeptidyl peptidase IV inhibitors. *J. Med. Chem.* **48,** 6991–6996 (2005).

20. Rummey, C., Nordhoff, S., Thiemann, M. & Metz, G. In silico fragment-based discovery of DPP-IV S1 pocket binders. *Bioorg. Med. Chem. Lett.* **16,** 1405–1409 (2006).

21. Al-masri, I. M., Mohammad, M. K. & Taha, M. O. Discovery of DPP IV inhibitors by pharmacophore modeling and QSAR analysis followed by in silico screening. *ChemMedChem* **3,** 1763–79 (2008).

22. Zhang, S. *et al.* Fast and effective identification of the bioactive compounds and their targets from medicinal plants via computational chemical biology approach. *MedChemComm* **2,** 471–477 (2011).

23. Li, C. *et al.* Identification of diverse dipeptidyl peptidase IV inhibitors via structure-based virtual screening. *J. Mol. Model.* **18,** 4033–4042 (2012).

24. Guasch, L. *et al.* Identification of novel human dipeptidyl peptidase-IV inhibitors of natural origin (Part I): Virtual screening and activity assays. *PLoS One* **7,** e44971 (2012).

25. Al-masri, I. M., Taha, M. O. & Mohammad, M. K. New leads for DPP IV inhibition: Structure-based pharmacophore mapping and virtual screening study. *Arch. Pharm. Res.* **36,** 1326–1337 (2013).

26. Xing, J. *et al.* Identification of dipeptidyl peptidase IV inhibitors: Virtual screening, synthesis and biological evaluation. *Chem. Biol. Drug Des.* **84,** 364–77 (2014).

27. Tanwar, O., Tanwar, L., Shaquiquzzaman, M., Alam, M. M. & Akhter, M. Structure based virtual screening of MDPI database: Discovery of structurally diverse and novel DPP-IV inhibitors. *Bioorg. Med. Chem. Lett.* **24,** 3447–3451 (2014).

28. Lankas, G. R. *et al.* Dipeptidyl peptidase IV inhibition for the treatment of type 2 diabetes: Potential importance of selectivity over dipeptidyl peptidases 8 and 9. *Diabetes* **54,** 2988–2994 (2005).

29. Gall, M. G. *et al.* Targeted inactivation of dipeptidyl peptidase 9 enzymatic activity causes mouse neonate lethality. *PLoS One* **8,** e78378 (2013).

30. Zeng, S. *et al.* Discovery of potent dipeptidyl peptidase IV inhibitors through pharmacophore hybridization and hit-to-lead optimization. *Bioorg. Med. Chem.* **21,** 1749–1755 (2013).

31. Liu, Y. *et al.* Synthesis and biological evaluation of novel benzyl-substituted (S)-phenylalanine derivatives as potent dipeptidyl peptidase 4 inhibitors. *Bioorg. Med. Chem.* **21,** 5679–5687 (2013).

32. Biftu, T. *et al.* Novel tetrahydropyran analogs as dipeptidyl peptidase IV inhibitors: Profile of clinical candidate (2R,3S,5R)-2-(2,5-difluorophenyl)-5-[2-(methylsulfonyl)-2,6-dihydropyrrolo[3,4-c]pyrazol-5(4H)-yl]tetrahydro-2H-pyran- 3-amine (23). *Bioorg. Med. Chem. Lett.* **23,** 5361–5366 (2013).

33. Xie, H. *et al.* Highly potent dipeptidyl peptidase IV inhibitors derived from alogliptin through pharmacophore hybridization and lead optimization. *Eur. J. Med. Chem.* **68,** 312–320 (2013).

34. Wang, J. *et al.* Synthesis and biological evaluation of pyrrolidine-2-carbonitrile and 4-fluoropyrrolidine-2-carbonitrile derivatives as dipeptidyl peptidase-4 inhibitors for the treatment of type 2 diabetes. *Bioorg. Med. Chem.* **21,** 7418–29 (2013).

35. Namoto, K. *et al.* Discovery of C-(1-aryl-cyclohexyl)-methylamines as selective, orally available inhibitors of dipeptidyl peptidase IV. *Bioorg. Med. Chem. Lett.* **24,** 731–736 (2014).

36. Ji, X. *et al.* Design, synthesis and biological evaluation of hetero-aromatic moieties substituted pyrrole-2-carbonitrile derivatives as dipeptidyl peptidase IV inhibitors. *Eur. J. Med. Chem.* **75,** 111–122 (2014).

37. Lai, Z.-W. *et al.* Discovery of highly potent DPP-4 inhibitors by hybrid compound design based on linagliptin and alogliptin. *Eur. J. Med. Chem.* **83,** 547–560 (2014).

38. Shu, C., Ge, H., Song, M. & Chen, J. Discovery of imigliptin, a novel selective DPP-4 inhibitor for the treatment of type 2 diabetes. *ACS Med. Chem. Lett.* **5,** 921–926 (2014).

39. Ji, X. *et al.* Design, synthesis and biological evaluation of 4-fluoropyrrolidine-2-carbonitrile and octahydrocyclopenta[b]pyrrole-2-carbonitrile derivatives as dipeptidyl peptidase IV inhibitors. *Eur. J. Med. Chem.* **86,** 242–56 (2014).

40. Jiang, T. *et al.* Design, synthesis, and pharmacological evaluation of highly potent and selective dipeptidyl peptidase-4 inhibitors. *Arch. Pharm. (Weinheim).* **348,** 399–407 (2015).

41. Jiang, T. *et al.* Design, synthesis, and pharmacological evaluation of fused β-homophenylalanine derivatives as potent DPP-4 inhibitors. *ACS Med. Chem. Lett.* **6,** 602–6 (2015).

42. Ran, Y., Pei, H., Shao, M. & Chen, L. Synthesis, biological evaluation and molecular docking of

(R)-2-((8-(3-aminopiperidin-1-yl)-3-methyl-7-(3-methylbut-2-en-1-yl)-2,6-dioxo-2,3,6,7-tetrahydro-1H-purin-1-yl)methyl)benzonitrile as dipeptidyl peptidase IV inhibitors. *Chem. Biol. Drug Des.* **87,** 290-5 (2015).

43. Wang, S. *et al.* (*R*)-3-amino-1-((3aS,7aS)-octahydro-1*H*-indol-1-yl)-4-(2,4,5-trifluorophenyl)butan-1-one derivatives as potent inhibitors of dipeptidyl peptidase-4: design, synthesis, biological evaluation, and molecular modeling. *Bioorg. Med. Chem.* **22,** 6684–93 (2014).

44. Chen, P. *et al.* Structure-activity-relationship of amide and sulfonamide analogs of omarigliptin. *Bioorg. Med. Chem. Lett.* **25,** 5767-71 (2015).

45. Wu, W.-L. *et al.* Discovery of novel tricyclic heterocycles as potent and selective DPP-4 inhibitors for the treatment of type 2 diabetes. *ACS Med. Chem. Lett.* **7,** 498–501 (2016).

46. Schwehm, C. *et al.* Synthesis of new DPP-4 inhibitors based on a novel tricyclic scaffold. *ACS Med. Chem. Lett.* **6,** 324–8 (2015).

47. Gomha, S. M., Eldebss, T. M. A., Badrey, M. G., Abdulla, M. M. & Mayhoub, A. S. Novel 4-heteroaryl-antipyrines as DPP-IV inhibitors. *Chem. Biol. Drug Des.* **86,** 1292–303 (2015).

48. Li, S. *et al.* Discovery and rational design of natural-product-derived 2-phenyl-3,4-dihydro-2H-benzo[f]chromen-3-amine analogs as novel and potent dipeptidyl peptidase 4 (DPP-4) inhibitors for the treatment of type 2 diabetes. *J. Med. Chem.* **59,** 6772–90 (2016).

49. Gwaltney, S. L. Medicinal chemistry approaches to the inhibition of dipeptidyl peptidase IV. *Curr. Top. Med. Chem.* **8,** 1545–1552 (2008).

50. Havale, S. H. & Pal, M. Medicinal chemistry approaches to the inhibition of dipeptidyl peptidase-4 for the treatment of type 2 diabetes. *Bioorg. Med. Chem.* **17,** 1783–1802 (2009).

51. Gaba, M., Singh, S. & Gaba, P. Dipeptidyl peptidase-4 inhibitors: a new approach in diabetes treatment. *Int. J. Drug. Dev. Res.* **1,** 146–156 (2009).

52. Kirby, M., Yu, D. M. T., O'Connor, S. & Gorrell, M. D. Inhibitor selectivity in the clinical application of dipeptidyl peptidase-4 inhibition. *Clin. Sci. (Lond).* **118,** 31–41 (2010).

53. Zettl, H., Schubert-Zsilavecz, M. & Steinhilber, D. Medicinal chemistry of incretin mimetics and DPP-4 inhibitors. *ChemMedChem* **5,** 179–185 (2010).

54. Liu, Y., Hu, Y. & Liu, T. Recent advances in non-peptidomimetic dipeptidyl peptidase 4 inhibitors: Medicinal chemistry and preclinical aspects. *Curr. Med. Chem.* **19,** 3982–3999 (2012).

55. Ghate, M. & Jain, S. V. Structure based lead optimization approach in discovery of selective DPP4 inhibitors. *Mini Rev. Med. Chem.* **13,** 888–914 (2013).

56. Juillerat-Jeanneret, L. Dipeptidyl peptidase IV and its inhibitors: Therapeutics for type 2 diabetes and what else? *J. Med. Chem.* **57,** 2197–2212 (2014).

57. Patel, B. D. & Ghate, M. D. Recent approaches to medicinal chemistry and therapeutic potential of dipeptidyl peptidase-4 (DPP-4) inhibitors. *Eur J Med Chem* **74,** 574–605 (2014).

58. Liu, Y. & Hu, Y. Novel DPP-4 inhibitors against diabetes. *Future Med. Chem.* **6,** 793–808 (2014).

59. Smelcerovic, A. *et al.* An overview of recent dipeptidyl peptidase-IV inhibitors: Linking their structure and physico-chemical properties with SAR, pharmacokinetics and toxicity. *Curr. Top. Med. Chem.* **15,** 2342–2372 (2015).

60. Activity Miner; in: Forge, v10.4 , Cresset, Litlington, Cambridgeshire, UK, ; http://www.cresset-group.com/forge/

61. Ojeda, M. J., Cereto-Massagué, A., Valls, C. & Pujadas, G. DPP-IV, an important target for antidiabetic functional food design; in *Foodinformatics. Applications of Chemical Information to Food Chemistry* (eds. Martinez-Mayorga, K. & Medina-Franco, J. L.) 177–212 (Springer International Publishing, 2014).

62. Power, O., Nongonierma, A. B., Jakeman, P. & Fitzgerald, R. J. Food protein hydrolysates as a source of dipeptidyl peptidase IV inhibitory peptides for the management of type 2 diabetes. *Proc. Nutr. Soc.* **73,** 34–46 (2014).

63. Mentlein, R. Dipeptidyl-peptidase IV (CD26)-role in the inactivation of regulatory peptides. *Regul. Pept.* **85,** 9–24 (1999).

64. Thoma, R. *et al.* Structural basis of proline-specific exopeptidase activity as observed in human dipeptidyl peptidase-IV. *Structure* **11,** 947–959 (2003).

65. Nabeno, M. *et al.* A comparative study of the binding modes of recently launched dipeptidyl peptidase IV inhibitors in the active site. *Biochem. Biophys. Res. Commun.* **434,** 191–196 (2013).

66. Gwaltney, S. & Stafford, J. Inhibitors of dipeptidyl peptidase 4. *Annu. Rep. Med. Chem.* **40,** 149–165 (2005).

67. Chien, C. H. *et al.* One site mutation disrupts dimer formation in human DPP-IV proteins. *J. Biol. Chem.* **279,** 52338–52345 (2004).

68. Kuhn, B., Hennig, M. & Mattei, P. Molecular recognition of ligands in dipeptidyl peptidase IV. *Curr. Top. Med. Chem.* **7,** 609–619 (2007).

69. Mendieta, L., Tarrago, T. & Giralt, E. Recent patents of dipeptidyl peptidase IV inhibitors. *Expert Opin Ther Pat* **21,** 1693–1741 (2011).

70. Engel, M. *et al.* Rigidity and flexibility of dipeptidyl peptidase IV: Crystal structures of and docking experiments with DPIV. *J. Mol. Biol.* **355,** 768–783 (2006).

71. Li, C. *et al.* Possible ligand release pathway of dipeptidyl peptidase IV investigated by molecular dynamics simulations. *Proteins Struct. Funct. Bioinforma.* **79,** 1800–1809 (2011).

72. Schechter, I. & Berger, A. On the size of the active site in proteases. I. Papain. 1967. *Biochem. Biophys. Res. Commun.* **425,** 497–502 (2012).

73. Aertgeerts, K. *et al.* Crystal structure of human dipeptidyl peptidase IV in complex with a decapeptide reveals details on substrate specificity and tetrahedral intermediate formation. *Protein Sci.* **13,** 412–21 (2004).

74. Weber, A. E. Dipeptidyl peptidase IV inhibitors for the treatment of diabetes. *J. Med. Chem.* **47,** 4135–4141 (2004).

75. Wallace, M. B. *et al.* Structure-based design and synthesis of benzimidazole derivatives as dipeptidyl peptidase IV inhibitors. *Bioorg. Med. Chem. Lett.* **18,** 2362–2367 (2008).

76. Patel, B. & Ghate, M. Computational studies on structurally diverse dipeptidyl peptidase IV inhibitors: an approach for new antidiabetic drug development. *Med. Chem. Res.* **22,** 4505–4521 (2013).

77. Bjelke, J. R. *et al.* Tyrosine 547 constitutes an essential part of the catalytic mechanism of dipeptidyl peptidase IV. *J. Biol. Chem.* **279,** 34691–34697 (2004).

78. Yoshida, T. *et al.* Fused bicyclic heteroarylpiperazine-substituted l-prolylthiazolidines as highly potent DPP-4 inhibitors lacking the electrophilic nitrile group. *Bioorg. Med. Chem.* **20,** 5033–5041 (2012).

79. Yoshida, T. *et al.* Discovery and preclinical profile of teneligliptin (3-[(2S,4S)-4-[4-(3-methyl-1-phenyl-1H-pyrazol-5-yl)piperazin-1-yl]pyrrolidin-2-ylcarbonyl]thiazolidine): A highly potent, selective, long-lasting and orally active dipeptidyl peptidase IV inhibitor for the treatment of type 2 diabetes. *Bioorg. Med. Chem.* **20,** 5705–5719 (2012).

80. Safavi, M., Foroumadi, A. & Abdollahi, M. The importance of synthetic drugs for type 2 diabetes drug discovery. *Expert Opin. Drug Discov.* **8,** 1339–1363 (2013).

81. Kang, N. S., Ahn, J. H., Kim, S. S., Chae, C. H. & Yoo, S. E. Docking-based 3D-QSAR study for selectivity of DPP4, DPP8, and DPP9 inhibitors. *Bioorg. Med. Chem. Lett.* **17,** 3716–3721 (2007).

82. Jadav, P. *et al.* Long-acting peptidomimetics based DPP-IV inhibitors. *Bioorg. Med. Chem. Lett.* **22,** 3516–3521 (2012).

83. Fan, J., Johnson, M. H., Lila, M. A., Yousef, G. & De Mejia, E. G. Berry and citrus phenolic compounds inhibit dipeptidyl peptidase IV: Implications in diabetes management. *Evid Based*

Complement. Alternat. Med. **2013,** 479505 (2013).

84. Kim, H. J. *et al.* Dipeptidyl peptidase-4 inhibitor with β-amino amide scaffold: Synthesis, SAR and biological evaluation. *Bioorg. Med. Chem. Lett.* **22,** 5545–5549 (2012).

85. Jain, S. V. & Ghate, M. Atom-based pharmacophore modeling, CoMFA/CoMSIA-based 3D-QSAR studies and lead optimization of DPP-4 inhibitors for the treatment of type 2 diabetes. *Med. Chem. Res.* **23,** 3436-50 (2014).

86. Nojima, H. *et al.* Comprehensive analysis of the Co-structures of dipeptidyl peptidase IV and its inhibitor. *BMC Struct. Biol.* **16,** 11 (2016).

87. Sheehan, S. M. *et al.* Discovery of non-covalent dipeptidyl peptidase IV inhibitors which induce a conformational change in the active site. *Bioorg. Med. Chem. Lett.* **17,** 1765–1768 (2007).

88. Oefner, C. *et al.* High-resolution structure of human apo dipeptidyl peptidase IV/CD26 and its complex with 1-[({2-[(5-iodopyridin-2-yl)amino]-ethyl}amino)- acetyl]-2-cyano-(S)-pyrrolidine. *Acta Crystallogr. - Sect. D Biol. Crystallogr.* **59,** 1206–1212 (2003).

89. Gorrell, M. D. Dipeptidyl peptidase IV and related enzymes in cell biology and liver disorders. *Clin. Sci. (Lond).* **108,** 277–292 (2005).

90. Janardhan, S. & Sastry, G. N. Dipeptidyl peptidase IV inhibitors: A new paradigm in type 2 diabetes treatment. *Curr. Drug Targets* **1,** 600–621 (2014).

91. Van Goethem, S. *et al.* Structure-activity relationship studies on isoindoline inhibitors of dipeptidyl peptidases 8 and 9 (DPP8, DPP9): is DPP8-selectivity an attainable goal? *J. Med. Chem.* **54,** 5737–46 (2011).

92. Dubois, V. *et al.* Dipeptidyl peptidase 9 (DPP9) from bovine testes: identification and characterization as the short form by mass spectrometry. *Biochim. Biophys. Acta* **1804,** 781–8 (2010).

93. Pitman, M. R., Menz, R. I. & Abbott, C. A. Prediction of dipeptidyl peptidase (DP) 8 structure by homology modelling. *Adv. Exp. Med. Biol.* **575,** 33–42 (2006).

94. Rummey, C. & Metz, G. Homology models of dipeptidyl peptidases 8 and 9 with a focus on loop predictions near the active site. *Proteins Struct. Funct. Genet.* **66,** 160–171 (2007).

95. Park, J. *et al.* Reversible inactivation of human dipeptidyl peptidases 8 and 9 by oxidation. *Open Enzym. Inhib. J.* **1,** 52–60 (2008).

96. Yazbeck, R., Howarth, G. S. & Abbott, C. a. Dipeptidyl peptidase inhibitors, an emerging drug class for inflammatory disease? *Trends Pharmacol. Sci.* **30,** 600–607 (2009).

97. Janardhan, S. & Reddy, Y. P. Homology modeling and molecular docking studies of human DPP8 and DPP9. *Int. J. Pharma Res. Dev.* **2,** 131–146 (2011).

98. Pitman, M. R., Menz, R. I. & Abbott, C. a. Hydrophilic residues surrounding the S1 and S2 pockets contribute to dimerisation and catalysis in human dipeptidyl peptidase 8 (DP8). *Biol. Chem.* **391,** 959–972 (2010).

99. Nakajima, Y. *et al.* Dipeptidyl aminopeptidase IV from *Stenotrophomonas maltophilia* exhibits activity against a substrate containing a 4-hydroxyproline residue. *J. Bacteriol.* **190,** 7819–29 (2008).

100. Pieper, U. *et al.* ModBase, a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Res.* **42,** D336-46 (2014).

101. ModBase entry for human Dipeptidyl peptidase 8. Available at: http://goo.gl/1ixlkL. (Accessed: 7th October 2015)

102. ModBase entry for human Dipeptidyl peptidase 9. Available at: http://goo.gl/iHtf51. (Accessed: 7th October 2015)

103. Kim, D. *et al.* (2R)-4-oxo-4-[3-(trifluoromethyl)-5,6-dihydro[1,2,4]triazolo[4,3-a]pyrazin-7(8H)-yl]-1-(2,4,5-triflu orophenyl)butan-2-amine: a potent, orally active dipeptidyl peptidase IV inhibitor for the treatment of type 2 diabetes. *J. Med. Chem.* **48,** 141–151 (2005).

104. Jiaang, W. T. *et al.* Novel isoindoline compounds for potent and selective inhibition of prolyl dipeptidase DPP8. *Bioorg. Med. Chem. Lett.* **15,** 687–691 (2005).

105. Tanwar, O. *et al.* Novel hydrazine derivatives as selective DPP-IV inhibitors: findings from virtual screening and validation through molecular dynamics simulations. *J. Mol. Model.* **20,** 2118 (2014).

106. Friesner, R. A. *et al.* Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **47,** 1739–49 (2004).

107. Halgren, T. A. *et al.* Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **47,** 1750–9 (2004).

108. Gao, Z. *et al.* PDTD: a web-accessible protein database for drug target identification. *BMC Bioinformatics* **9,** 104 (2008).

109. Li, H. *et al.* TarFisDock: A web server for identifying drug targets with docking approach. *Nucleic Acids Res.* **34,** W219-24 (2006).

110. Specs.net. Available at: http://www.specs.net/. (Accessed: 19th October 2015)

111. 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.* **46,** 3–26 (2001).

112. Morris, G. M. *et al.* AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **30,** 2785–91 (2009).

113. Sherman, W., Day, T., Jacobson, M. P., Friesner, R. A. & Farid, R. Novel procedure for modeling ligand/receptor induced fit effects. *J. Med. Chem.* **49,** 534–53 (2006).

114. Wu, D. *et al.* Synthesis, Structure-activity relationship, and pharmacophore modeling studies of pyrazole-3-carbohydrazone derivatives as dipeptidyl peptidase IV inhibitors. *Chem. Biol. Drug Des.* **79,** 897–906 (2012).

115. Catalyst; Accelrys, Inc., San Diego, CA 92121, U.S.A. Available at: http://www.accelrys.com.

116. Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S. & Coleman, R. G. ZINC: A free tool to discover chemistry for biology. *J. Chem. Inf. Model.* **52,** 1757–1768 (2012).

117. 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* **9,** 396 (2008).

118. Hawkins, P. C. D. & Nicholls, A. Conformer generation with OMEGA: Learning from the data set and the analysis of failures. *J. Chem. Inf. Model.* **52,** 2919–2936 (2012).

119. Hawkins, P. C. D., Skillman, A. G., Warren, G. L., Ellingson, B. A. & Stahl, M. T. Conformer generation with OMEGA: Algorithm and validation using high quality structures from the protein databank and cambridge structural database. *J. Chem. Inf. Model.* **50,** 572–584 (2010).

120. Salam, N. K., Nuti, R. & Sherman, W. Novel method for generating structure-based pharmacophores using energetic analysis. *J. Chem. Inf. Model.* **49,** 2356–68 (2009).

121. Dixon, S. L. *et al.* 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,** 647–71 (2006).

122. 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.* **26,** 198–212 (2007).

123. EON 2.0.1: OpenEye Scientific Software, Santa Fe, NM. http://www.eyesopen.com.

124. Edmondson, S. D. *et al.* Fluoroolefins as amide bond mimics in dipeptidyl peptidase IV inhibitors. *Bioorg. Med. Chem. Lett.* **18,** 2409–2413 (2008).

125. Schrödinger LLC., Portland, USA. CombiGlide Diverse Side-chain Collection, v. 1.2.Available at: https://www.schrodinger.com/combiglide.

126. Guasch, L. *et al.* Identification of novel human dipeptidyl peptidase-IV inhibitors of natural origin (Part II): *In silico* prediction in antidiabetic extracts. *PLoS One* **7,** e44972 (2012).

127. Discovery Studio (DS) Visualizer; Accelrys, Inc., San Diego, CA 92121, U.S.A. Available at:

http://www.accelrys.com.

128. The National Cancer Institute - Development Therapeutics Program. Available at: https://dtp.cancer.gov/.

129. FILTER 2.0.1; OpenEye Scientific Software, Santa Fe, NM. Available at:http://www.eyesopen.com.

130. McGann, M. FRED pose prediction and virtual screening accuracy. *J. Chem. Inf. Model.* **51,** 578–96 (2011).

131. McGann, M. FRED and HYBRID docking performance on standardized datasets. *J. Comput. Aided. Mol. Des.* **26,** 897–906 (2012).

132. Discovery Studio (DS); Accelrys, Inc., San Diego, CA 92121, U.S.A. Available at: http://www.accelrys.com.

133. Venkatachalam, C. M., Jiang, X., Oldfield, T. & Waldman, M. LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. *J. Mol. Graph. Model.* **21,** 289–307 (2003).

134. Friesner, R. A. *et al.* Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* **49,** 6177–96 (2006).

135. Peters, J. U. *et al.* Aminomethylpyrimidines as novel DPP-IV inhibitors: A 10^5-fold activity increase by optimization of aromatic substituents. *Bioorg. Med. Chem. Lett.* **14,** 1491–1493 (2004).

136. Jones, G., Willett, P., Glen, R. C., Leach, A. R. & Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **267,** 727–48 (1997).

137. Xu, J. *et al.* Discovery of potent and selective phenylalanine based dipeptidyl peptidase IV inhibitors. *Bioorg. Med. Chem. Lett.* **15,** 2533–2536 (2005).

138. Edmondson, S. D. *et al.* Discovery of potent and selective orally bioavailable β-substituted phenylalanine derived dipeptidyl peptidase IV inhibitors. *Bioorg. Med. Chem. Lett.* **15,** 3048–3052 (2005).

139. Biftu, T. *et al.* (3*R*)-4-[(3*R*)-3-Amino-4-(2,4,5-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. *Bioorg. Med. Chem. Lett.* **17,** 49–52 (2007).

140. Pei, Z. *et al.* Discovery and structure - Activity relationships of piperidinone- and piperidine-constrained phenethylamines as novel, potent, and selective dipeptidyl peptidase IV inhibitors. *J. Med. Chem.* **50,** 1983–1987 (2007).

141. Wright, S. W. *et al.* (3R,4S)-4-(2,4,5-Trifluorophenyl)-pyrrolidin-3-ylamine inhibitors of dipeptidyl peptidase IV: synthesis, in vitro, in vivo, and X-ray crystallographic characterization. *Bioorg. Med. Chem. Lett.* **17,** 5638–5642 (2007).

142. Eckhardt, M. *et al.* 8-(3-(*R*)-aminopiperidin-1-yl)-7-but-2-ynyl-3-methyl-1-(4-methyl-quinazolin-2-ylmethyl)-3,7-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,** 6450–6453 (2007).

143. Nordhoff, S. *et al.* Discovery of β-homophenylalanine based pyrrolidin-2-ylmethyl amides and sulfonamides as highly potent and selective inhibitors of dipeptidyl peptidase IV. *Bioorg. Med. Chem. Lett.* **19,** 4201–4203 (2009).

144. Mattei, P. *et al.* Discovery of carmegliptin: a potent and long-acting dipeptidyl peptidase IV inhibitor for the treatment of type 2 diabetes. *Bioorg. Med. Chem. Lett.* **20,** 1109–13 (2010).

145. Banno, Y. *et al.* Identification of 3-aminomethyl-1,2-dihydro-4-phenyl-1-isoquinolones: a new class of potent, selective, and orally active non-peptide dipeptidyl peptidase IV inhibitors that form a unique interaction with Lys554. *Bioorg. Med. Chem.* **19,** 4953–70 (2011).

146. Sutton, J. M. *et al.* Novel heterocyclic DPP-4 inhibitors for the treatment of type 2 diabetes. *Bioorg. Med. Chem. Lett.* **22,** 1464–8 (2012).

147. Lam, B. *et al.* Structure-based design of pyridopyrimidinediones as dipeptidyl peptidase IV inhibitors. *Bioorg. Med. Chem. Lett.* **22,** 6628–31 (2012).

148. Sakashita, H. *et al.* Lead optimization of [(*S*)-γ-(arylamino)prolyl]thiazolidine focused on γ-substituent: Indoline compounds as potent DPP-IV inhibitors. *Bioorg. Med. Chem.* **15,** 641–55 (2007).

149. Lübbers, T. *et al.* 1,3-disubstituted 4-aminopiperidines as useful tools in the optimization of the 2-aminobenzo[a]quinolizine dipeptidyl peptidase IV inhibitors. *Bioorg. Med. Chem. Lett.* **17,** 2966–70 (2007).

150. Ikuma, Y. *et al.* Discovery of 3*H*-imidazo[4,5-*c*]quinolin-4(5*H*)-ones as potent and selective dipeptidyl peptidase IV (DPP-4) inhibitors. *Bioorg. Med. Chem.* **20,** 5864–5883 (2012).

151. Reaxys Medicinal Chemistry. Available at: http://www.reaxys.com.

152. Forge, v10.4 , Cresset, Litlington, Cambridgeshire, UK. Available at: http://www.cresset-group.com/forge/.

153. Aertgeerts, K. *et al.* Structural and kinetic analysis of the substrate specificity of human fibroblast activation protein α. *J. Biol. Chem.* **280,** 19441–4 (2005).

154. Bosshard, H. R., Marti, D. N. & Jelesarov, I. Protein stabilization by salt bridges: concepts, experimental approaches and clarification of some misunderstandings. *J. Mol. Recognit.* **17,**

1–16 (2004).

155. Abbott, C. A., McCaughan, G. W. & Gorrell, M. D. Two highly conserved glutamic acid residues in the predicted β propeller domain of dipeptidyl peptidase IV are required for its enzyme activity. *FEBS Lett.* **458,** 278–84 (1999).

156. Zhu, L. *et al.* Design and synthesis of 4-(2,4,5-Trifluorophenyl)butane-1,3-diamines as dipeptidyl peptidase IV inhibitors. *ChemMedChem* **8,** 1104–1116 (2013).

157. Metzler, W. J. *et al.* Involvement of DPP-IV catalytic residues in enzyme-saxagliptin complex formation. *Protein Sci.* **17,** 240–250 (2008).

158. Miyamoto, Y. *et al.* Discovery of a 3-pyridylacetic acid derivative (TAK-100) as a potent, selective and orally active dipeptidyl peptidase IV (DPP-4) inhibitor. *J. Med. Chem.* **54,** 831–850 (2011).

159. Miyamoto, Y. *et al.* Design and synthesis of 3-pyridylacetamide derivatives as dipeptidyl peptidase IV (DPP-4) inhibitors targeting a bidentate interaction with Arg125. *Bioorg. Med. Chem.* **19,** 172–85 (2011).

160. Longenecker, K. L. *et al.* Crystal structures of DPP-IV (CD26) from rat kidney exhibit flexible accommodation of peptidase-selective inhibitors. *Biochemistry* **45,** 7474–82 (2006).

161. Zhang, Z. *et al.* Design and synthesis of pyrimidinone and pyrimidinedione inhibitors of dipeptidyl peptidase IV. *J. Med. Chem.* **54,** 510–524 (2011).

162. Van der Veken, P. *et al.* Inhibitors of dipeptidyl peptidase 8 and dipeptidyl peptidase 9. Part 1: Identification of dipeptide derived leads. *Bioorg. Med. Chem. Lett.* **18,** 4154–4158 (2008).

163. Maezaki, H. *et al.* Discovery of potent, selective, and orally bioavailable quinoline-based dipeptidyl peptidase IV inhibitors targeting Lys554. *Bioorg. Med. Chem.* **19,** 4482–98 (2011).

164. Arulmozhiraja, S. *et al.* Comparative binding analysis of dipeptidyl peptidase IV (DPP-4) with antidiabetic drugs – An *ab initio* fragment molecular orbital study. *PLoS One* **11,** e0166275 (2016).

165. Maestro, version 10.3, Schrödinger, LLC, New York, NY, 2012. Available from: http://www.schrodinger.com/.

166. Sharma, M., Gupta, M., Singh, D., Kumar, M. & Kaur, P. Synthesis, evaluation and molecular docking of thiazolopyrimidine derivatives as dipeptidyl peptidase IV inhibitors. *Chem. Biol. Drug Des.* **80,** 918–28 (2012).

167. Hiramatsu, H. *et al.* Crystal structures of human dipeptidyl peptidase IV in its apo and diprotin B-complexed forms. *Acta Biochim. Biophys. Sin. (Shanghai).* **39,** 335–43 (2007).

168. Ceriello, A., Sportiello, L., Rafaniello, C. & Rossi, F. DPP-4 inhibitors: pharmacological

differences and their clinical implications. *Expert Opin. Drug Saf.* **13,** S57-68 (2014).

169. Cereto-Massagué, A. *et al.* The good, the bad and the dubious: VHELIBS, a validation helper for ligands and binding sites. *J. Cheminform.* **5,** 36 (2013).

170. Kleywegt, G. J. *et al.* The uppsala Electron-density server. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60,** 2240–2249 (2004).

171. Lu, G. *et al.* Molecular basis of binding between novel human coronavirus MERS-CoV and its receptor CD26. *Nature* **500,** 227–31 (2013).

172. Wang, N. *et al.* Structure of MERS-CoV spike receptor-binding domain complexed with human receptor DPP4. *Cell Res.* **23,** 986–93 (2013).

173. Edmondson, S. D. *et al.* (2*S*,3*S*)-3-Amino-4-(3,3-difluoropyrrolidin-1-yl)-N,N-dimethyl-4-oxo-2-(4-[1,2,4]triazolo[1,5-*a*]-pyridin-6-ylphenyl)butanamide: a selective α-amino amide dipeptidyl peptidase IV inhibitor for the treatment of type 2 diabetes. *J. Med. Chem.* **49,** 3614–3627 (2006).

174. Pei, Z. *et al.* Discovery of ((4*R*,5*S*)-5-amino-4-(2,4,5- trifluorophenyl)cyclohex-1-enyl)-(3-(trifluoromethyl)-5,6-dihydro- [1,2,4]triazolo[4,3-a]pyrazin-7(8*H*)-yl)methanone (ABT-341), a highly potent, selective, orally efficacious, and safe dipeptidyl peptidase IV inhibi. *J. Med. Chem.* **49,** 6439–6442 (2006).

175. Backes, B. J. *et al.* Pyrrolidine-constrained phenethylamines: The design of potent, selective, and pharmacologically efficacious dipeptidyl peptidase IV (DPP4) inhibitors from a lead-like screening hit. *Bioorg. Med. Chem. Lett.* **17,** 2005–2012 (2007).

176. Weihofen, W. a., Liu, J., Reutter, W., Saenger, W. & Fan, H. Crystal structure of CD26/dipeptidyl-peptidase IV in complex with adenosine deaminase reveals a highly amphiphilic interface. *J. Biol. Chem.* **279,** 43330–43335 (2004).

177. Wang, Q. *et al.* Bat origins of MERS-CoV supported by bat coronavirus HKU4 usage of human receptor CD26. *Cell Host Microbe* **16,** 328–37 (2014).

178. MarvinSketch 14.7.21.0, ChemAxon (2014). Available at: http://www.chemaxon.com/products/marvin/marvinsketch/.

Supplementary Material

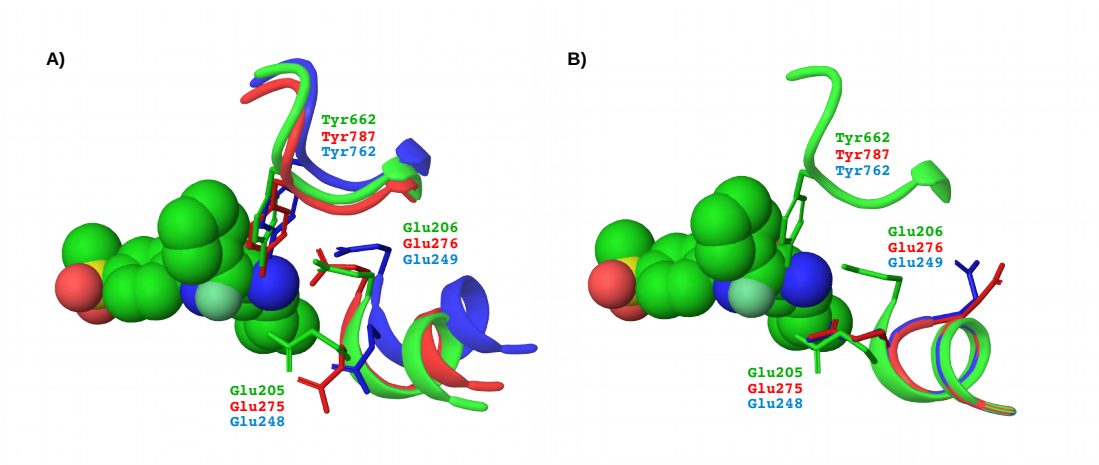
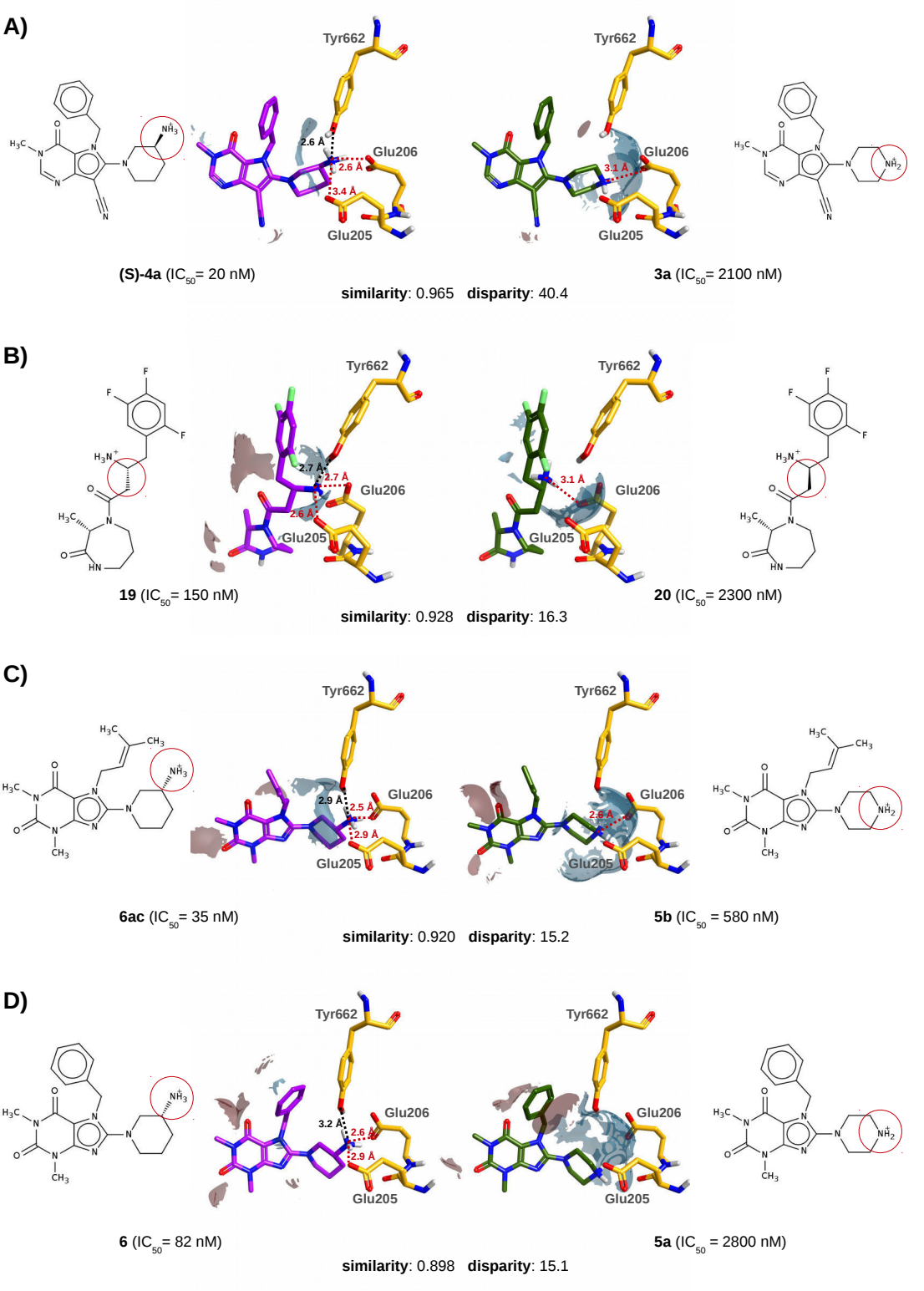


Figure S1. This figure compares the N-terminal recognition region in DPP-IV with the equivalent region in the ModBase[101,102] and Janardhan & Reddy[97] homology models for DPP8 and DPP9. Panel **A** corresponds to the superposition of the homology models downloaded from ModBase[101,102] relative to DPP-IV (PDB code: 3C45).[124] Panel **B** corresponds to the superposition of the homology models generated by Janardhan & Reddy[97] relative to DPP-IV (PDB code: 1X70).[103] DPP-IV, DPP8 and DPP9 are shown in ribbons and colored green, red and blue respectively (following the same color schema used in Figure 2). The DPP-IV ligand at 3C45[124] is shown in spacefill in both panels to reference the active site location. The Glu205, Glu206 and Tyr662 residues (and their equivalents in DPP8 and DPP9) are represented in wireframe format in the 3D structures. This figure was obtained with the help of the Maestro program.[165]



(S)-4a (IC_{50}= 20 nM) **3a** (IC_{50}= 2100 nM)

similarity: 0.965 **disparity**: 40.4

19 (IC_{50}= 150 nM) **20** (IC_{50}= 2300 nM)

similarity: 0.928 **disparity**: 16.3

6ac (IC_{50}= 35 nM) **5b** (IC_{50} = 580 nM)

similarity: 0.920 **disparity**: 15.2

6 (IC_{50}= 82 nM) **5a** (IC_{50} = 2800 nM)

similarity: 0.898 **disparity**: 15.1

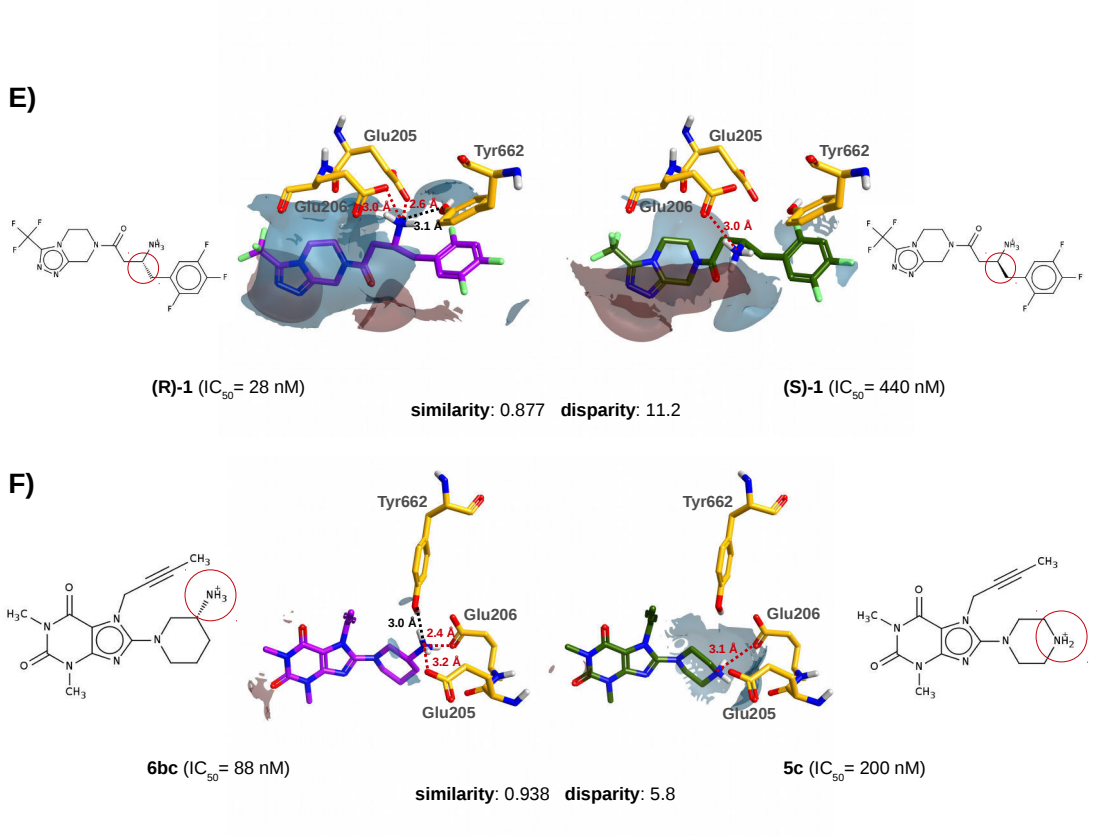
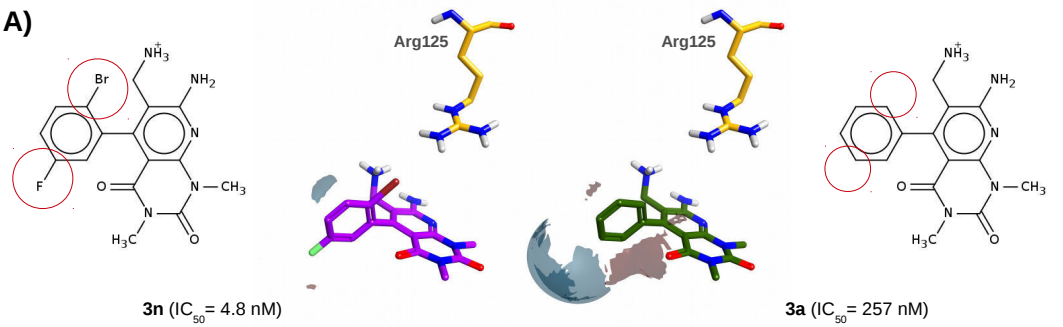


Figure S2. Comparison of the distribution of electrostatic surfaces between pairs of compounds that differ in their interactions with residues Glu205, Glu206 and Tyr662. For each panel the compound with the highest activity is shown in purple on the left and the compound with the lowest activity is shown in green on the right. Molecules are labeled with the same names that identify them in the corresponding paper.[103,139,142,146] The negative and positive electrostatic surface differences are shown in garnet and blue respectively (where the default value – *i.e.,* 2.0 – was used as the threshold for the surface difference between each pair). Dotted lines represent either donor and acceptor atoms with the potential to form hydrogen bonds (in black) or atom pairs with the potential to form salt bridges (in red). In the 2D representation of each ligand, the structural differences between the compared compounds are highlighted. The different panels are arranged in order of decreasing disparity and correspond to different situations: in panels **A**, **C**, **D** and **F** the NH_3^+ group is replaced by H and a charged secondary amine is introduced in the adjacent carbon;[142,146] in panels **B** and **E** the configuration of the carbon containing the NH_3^+ group is switched from (R) in the most potent compounds to (S).[103,139] The ligand orientations are the result of their superposition with co-crystallized ligands from the same or very similar chemical series (*i.e.,* 4A5S[146] for panel **A**, 2IIV[139] for panel **B**, 1X70[103] for panel **E** and 2RGU[142] for panels **C**, **D** and **F**; the residue locations in each panel are also taken from the corresponding PDB file). This figure was obtained with the help of the Forge[152] and MarvinSketch programs.[178]

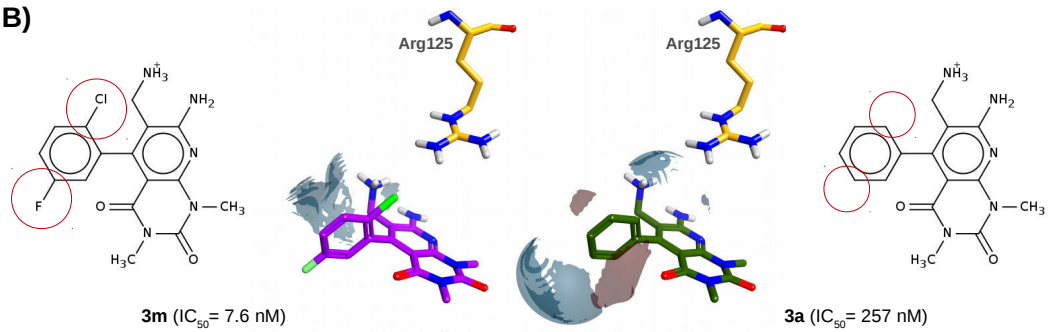
A)



3n (IC_{50}= 4.8 nM) **3a** (IC_{50}= 257 nM)

similarity: 0.925 **disparity**: 23.1

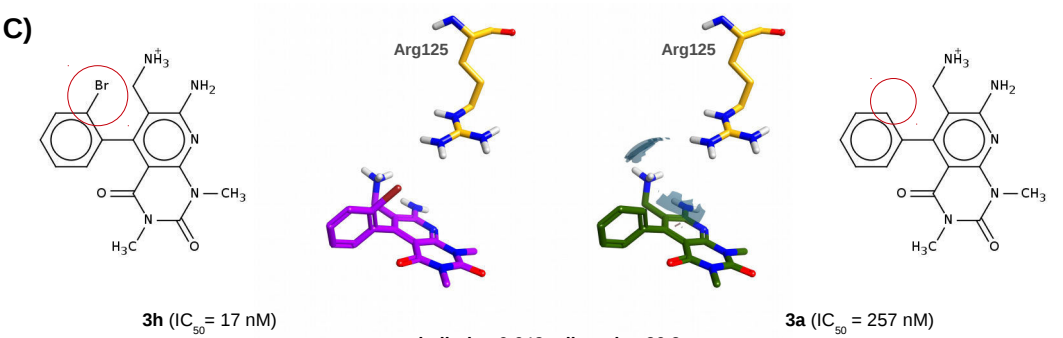
B)



3m (IC_{50}= 7.6 nM) **3a** (IC_{50}= 257 nM)

similarity: 0.931 **disparity**: 22.1

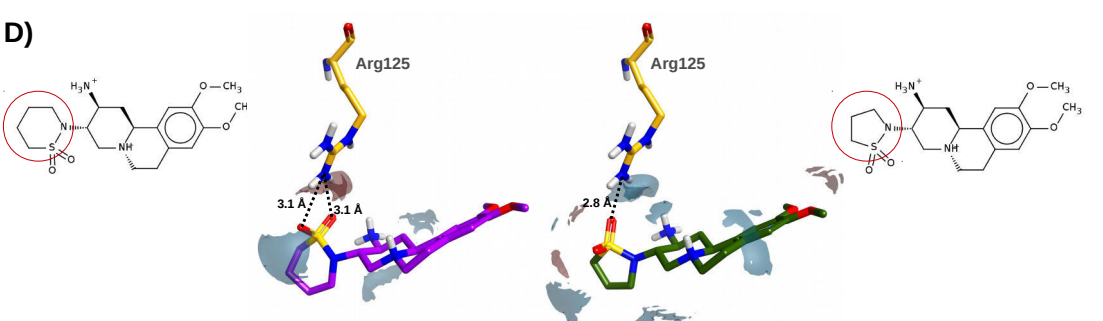
C)



3h (IC_{50}= 17 nM) **3a** (IC_{50} = 257 nM)

similarity: 0.943 **disparity**: 20.8

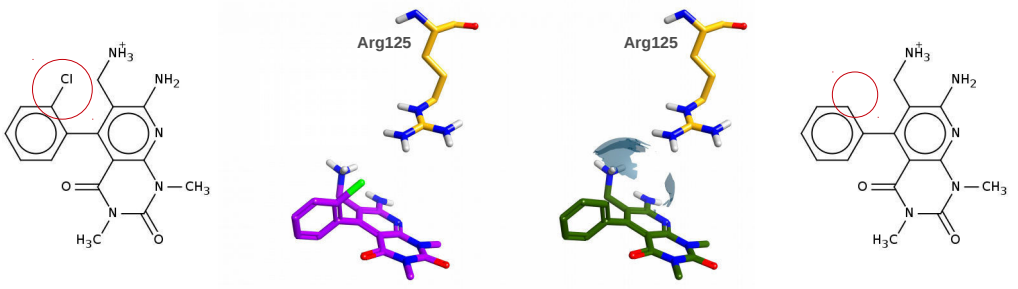
D)



8e (IC_{50}= 142 nM) **8d** (IC_{50}= 1330 nM)

similarity: 0.948 **disparity**: 18.8

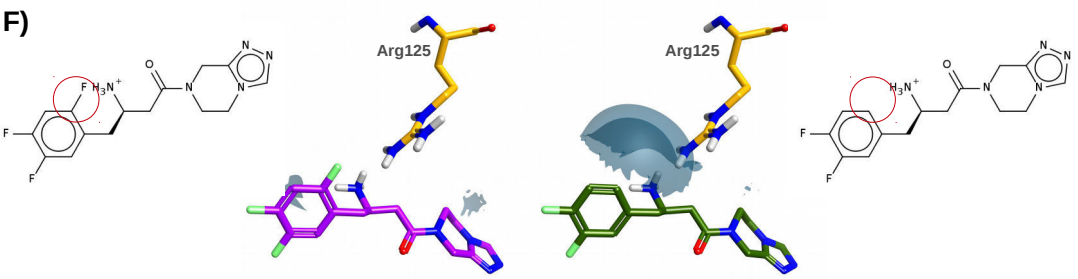
E)



3g (IC_{50}= 33 nM) **3a** (IC_{50} = 257 nM)

similarity: 0.952 **disparity**: 17.8

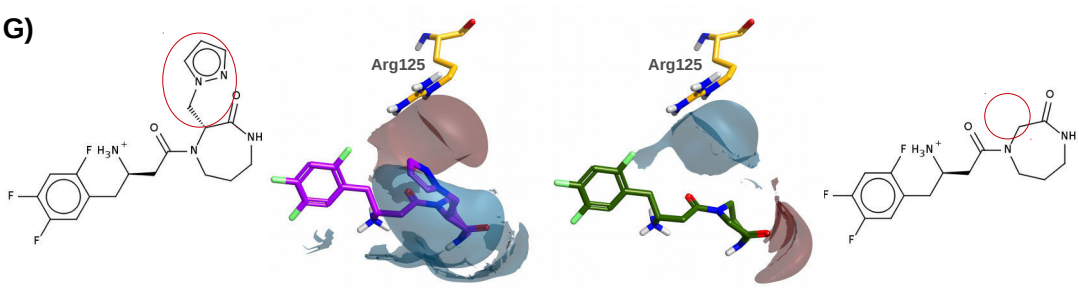
F)



26 (IC_{50}= 68 nM) **22** (IC_{50} = 455 nM)

similarity: 0.977 **disparity**: 16.6

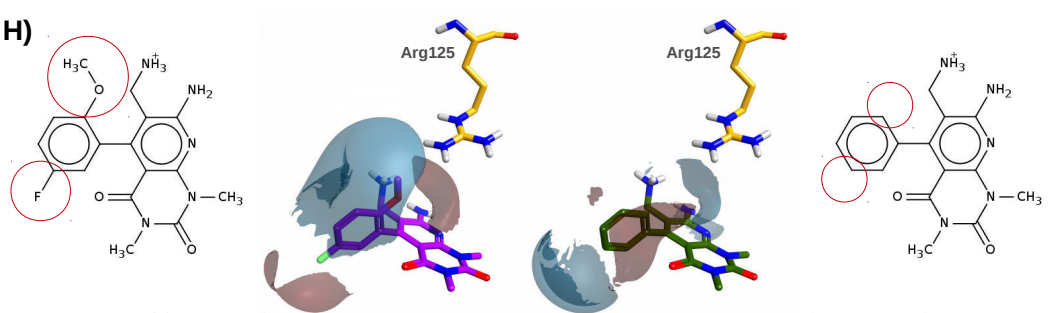
G)



26 (IC_{50}= 0.29 nM) **22** (IC_{50} = 140 nM)

similarity: 0.833 **disparity**: 16.1

H)



3l (IC_{50}= 20 nM) **3a** (IC_{50} = 257 nM)

similarity: 0.921 **disparity**: 14.0

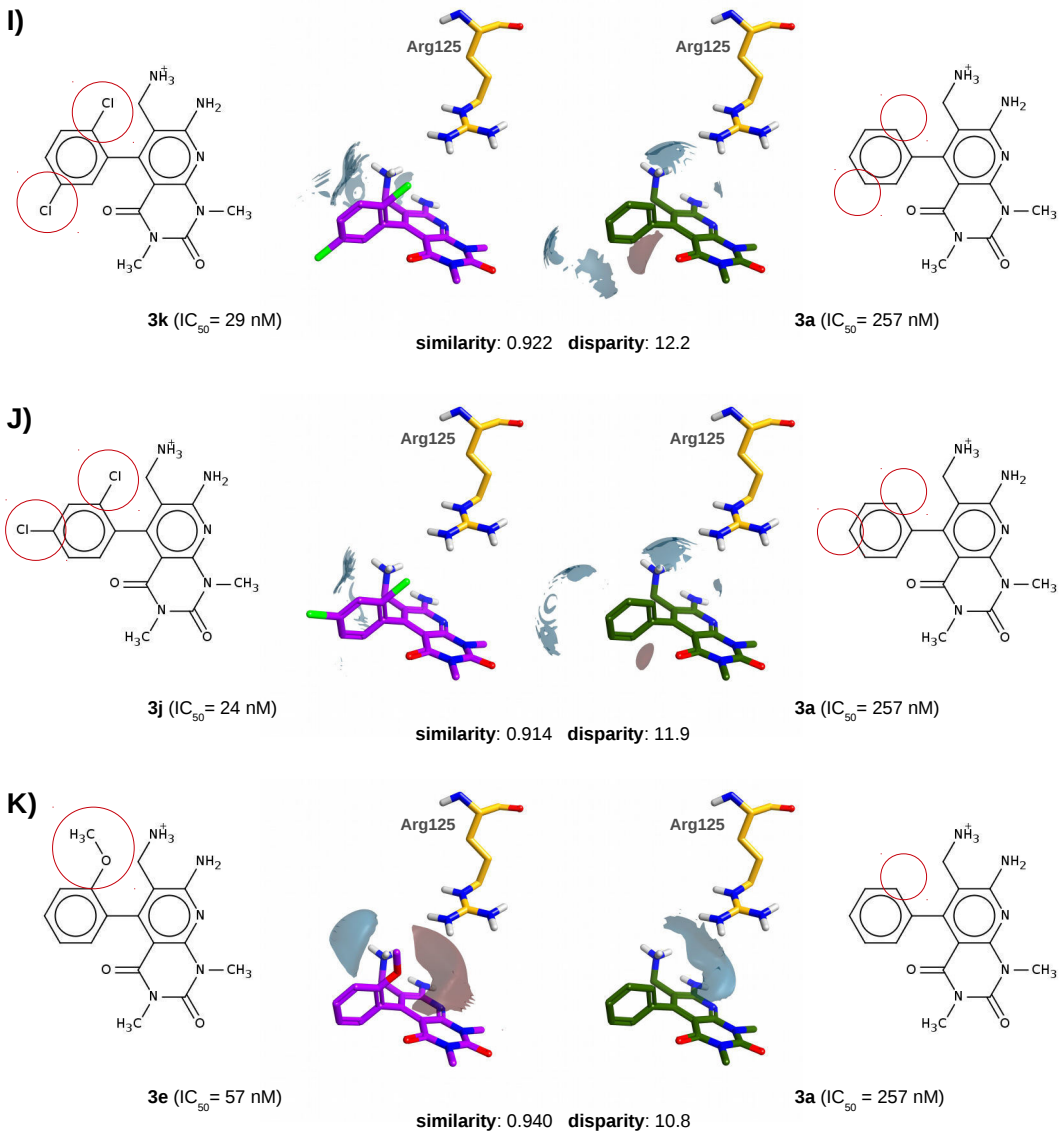
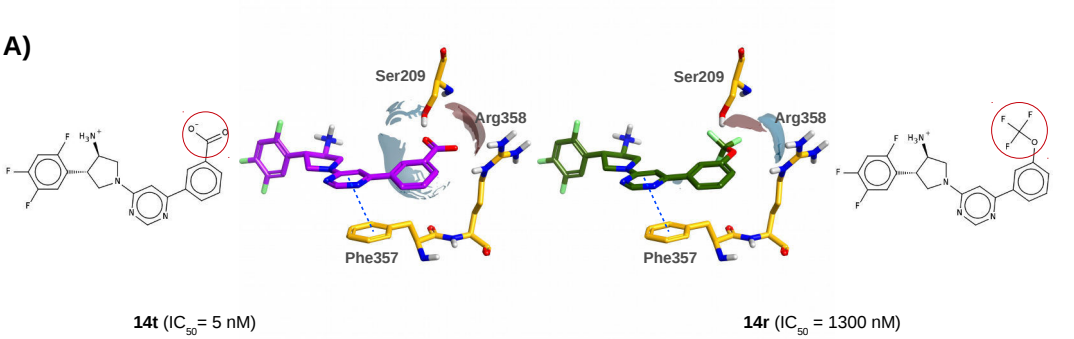
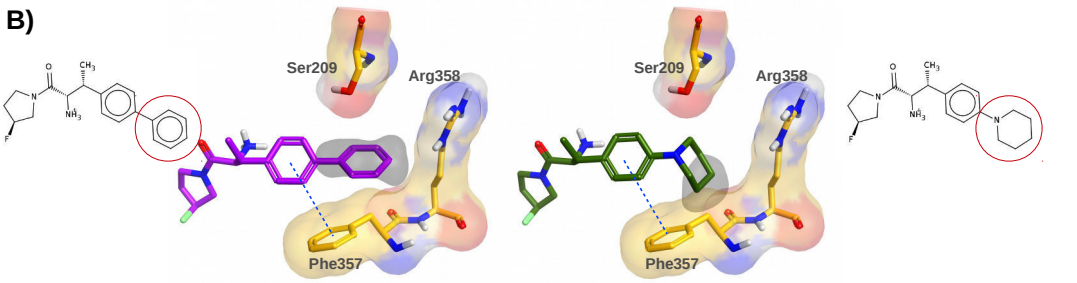


Figure S3. Comparison of the distribution of electrostatic surfaces between pairs of compounds that differ in their interactions with residue Arg125. For each panel the compound with the highest activity is shown in purple on the left and the compound with the lowest activity is shown in green on the right. Molecules are labeled with the same names that identify them in the corresponding paper.[103,139,144,147] The negative and positive electrostatic surface differences are shown in garnet and blue respectively (where the default value – *i.e.,* 2.0 – was used as the threshold for the surface difference between each pair). Dotted lines represent the distances between donor and acceptor atoms with the potential to form hydrogen bonds. In the 2D representation of each ligand, the structural differences between the compared compounds are highlighted. The different panels are arranged in order of decreasing disparity and correspond to different situations: in panels **A, B, C, E, F, I** and **J** a halogen is added in the ortho position of the phenyl ring;[103,147] in panels **G, H** and **K** a negative environment is placed around Arg125;[139,147] in panel **D** the ring size is increased.[144] The ligand orientations are the result of their superposition with co-crystallized ligands from the same or very similar chemical series (*i.e.,* 4G1F[147] for panels **A**, **B**, **C**, **E**, **H**, **I**, **J** and **K**, 3KWF[144] for panel **D**, 1X70[103] for panel **F** and 2IIV[139] for panel **G**; residue locations in each panel are also taken from the corresponding PDB file). This figure was obtained with the help of the Forge[152] and MarvinSketch programs.[178]



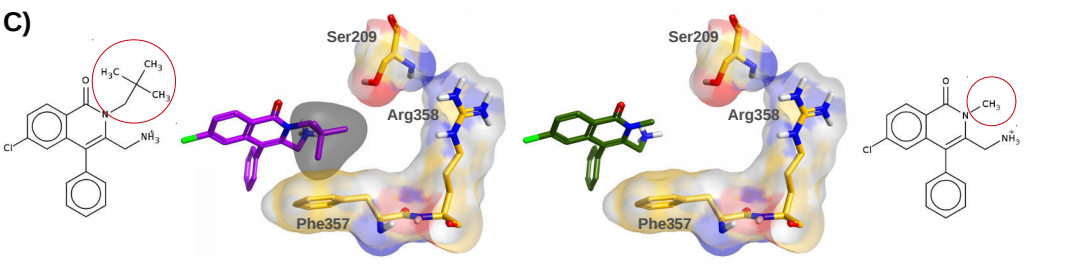
14t (IC_{50}= 5 nM) **14r** (IC_{50} = 1300 nM)

similarity: 0.936 **disparity**: 37.5



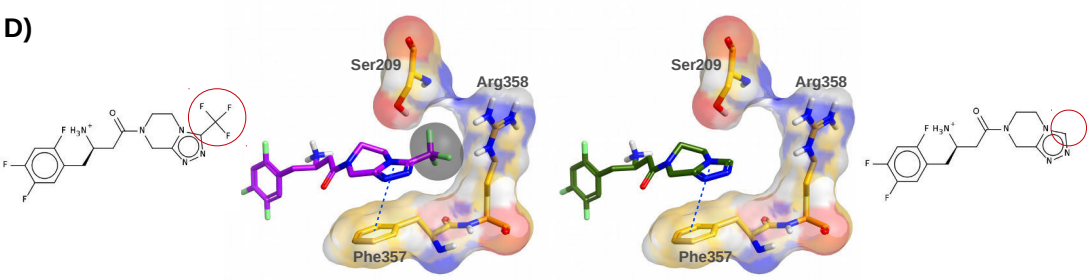
12q (IC_{50}= 120 nM) **12s** (IC_{50} = 7900 nM)

similarity: 0.883 **disparity**: 15.5



6g (IC_{50}= 1900 nM) **6a** (IC_{50} = 16000 nM)

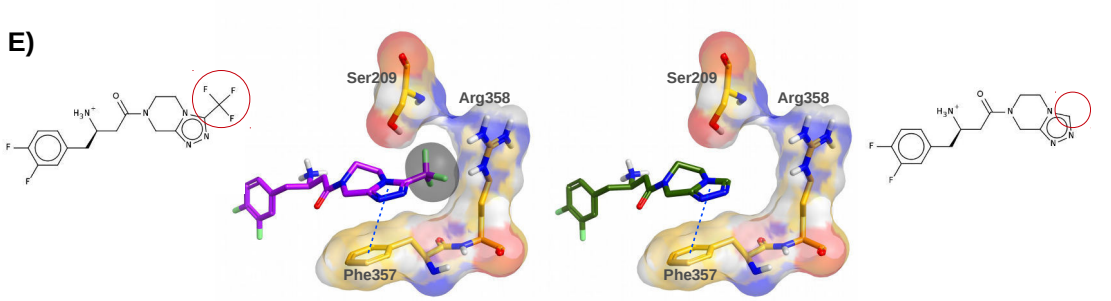
similarity: 0.904 **disparity**: 9.6



1 (IC_{50}= 18 nM) **26** (IC_{50} = 68 nM)

similarity: 0.924 **disparity**: 7.5

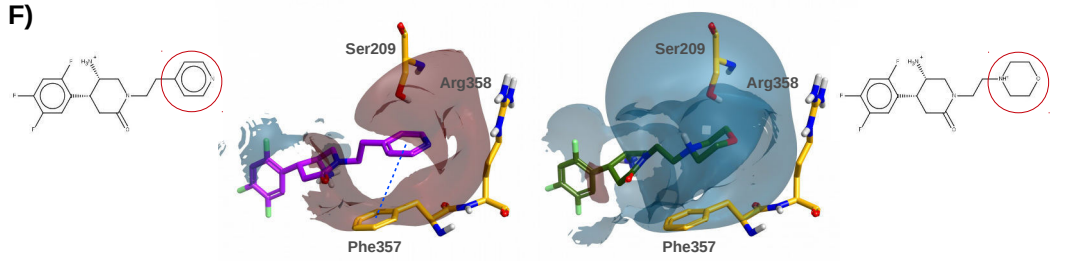
E)



24 (IC_{50}= 128 nM) **22** (IC_{50} = 455 nM)

similarity: 0.925 **disparity**: 7.3

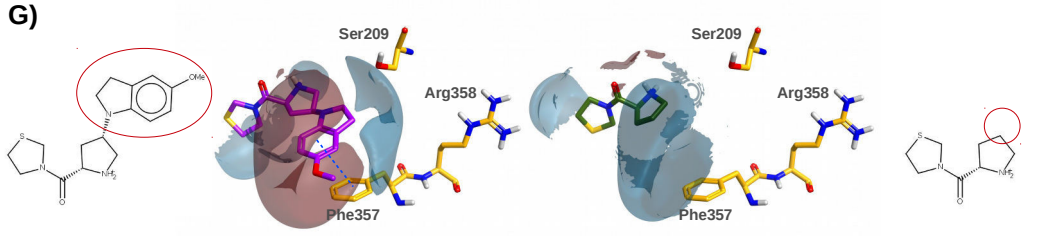
F)



24 (K_i = 24 nM) **23** (K_i = 590 nM)

similarity: 0.777 **disparity**: 6.2

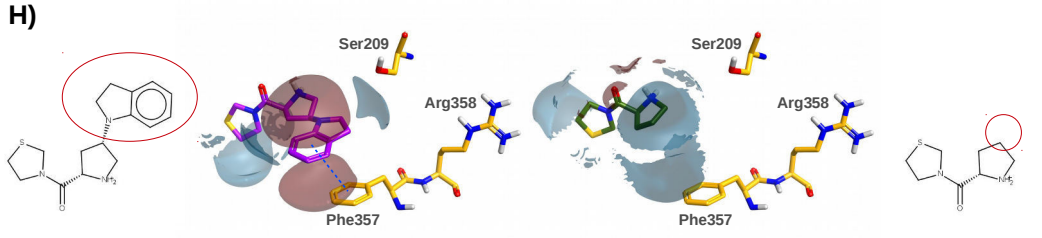
G)



22e (IC_{50}= 4.8 nM) **10** (IC_{50} = 558 nM)

similarity: 0.599 **disparity**: 5.2

H)



22c (IC_{50}= 11 nM) **10** (IC_{50} = 558 nM)

similarity: 0.651 **disparity**: 4.9

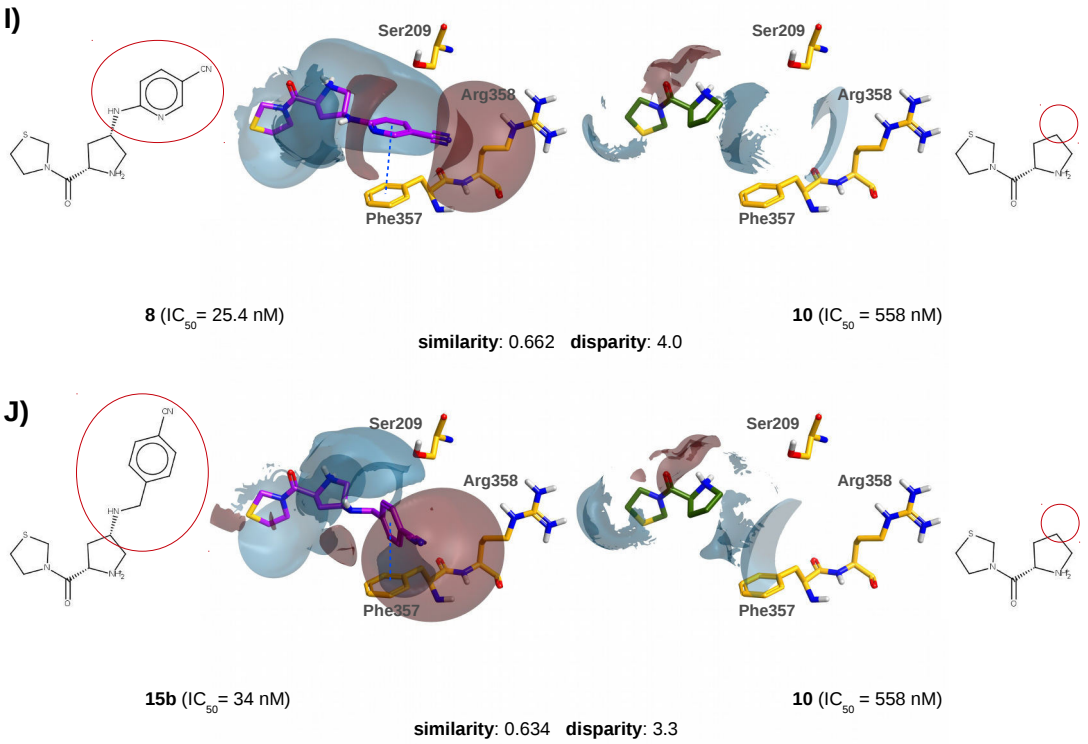
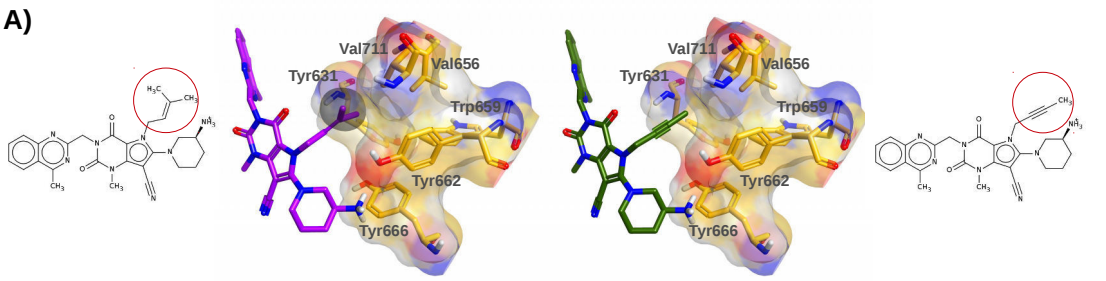


Figure S4. Comparison of the distribution of electrostatic and hydrophobic surfaces between pairs of compounds that differ in their interactions with residues Ser209, Phe357, Arg358. For each panel the compound with the highest activity is shown in purple on the left and the compound with the lowest activity is shown in green on the right. Molecules are labeled with the same names that identify them in the corresponding paper.[103,137,140,141,145,148] In panels **A**, **F**, **G**, **H**, **I** and **J** the negative and positive electrostatic surface differences are shown in garnet and blue respectively. In all the other panels the hydrophobic surface differences are shown in gray, while the protein surface has been colored according to atom color. The field surface difference is established by default at 2.0. In the 2D representation of each ligand, the structural differences between the compared compounds are highlighted. The different panels are arranged in order of decreasing disparity and correspond to different situations: in panel **A,** a negative environment is placed around Arg358;[141] in panels **B**, **D** and **E** a hydrophobic interaction is established with the S_2 pocket;[103,137] in panel **C**, a new hydrophobic interaction is established with Phe357;[145] and in panels **F**, **G, H, I** and **J** a new π-π interaction is established with Phe357.[140,148] The blue dashed lines show the π-π interactions between the corresponding ligand and Phe357 and were calculated with the help of Maestro[165] using default options. The ligand orientations are the result of their superposition with co-crystallized ligands from the same or very similar chemical series (*i.e.,* 2QJR for panel **A**, 2FJP[173] for panel **B**, 3OPM[145] for panel **C**, 1X70[103] for panels **D** and **E**, 2OQV[140] for panel **F** and 3VJK[79] for panels **G**, **H**, **I** and **J**; residue locations in each panel are also taken from the corresponding PDB file). This figure was obtained with the help of the Forge[152] and MarvinSketch programs.[178]

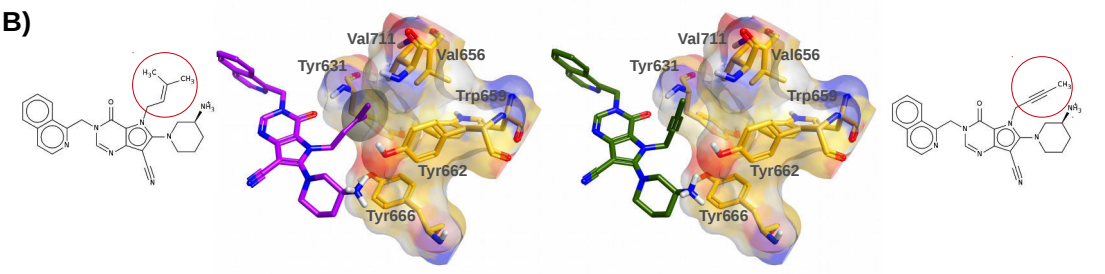
A)



(S)-11f (IC_{50}= 2 nM) **(S)-11k** (IC_{50}= 250 nM)

similarity: 0.980 **disparity**: 42.0

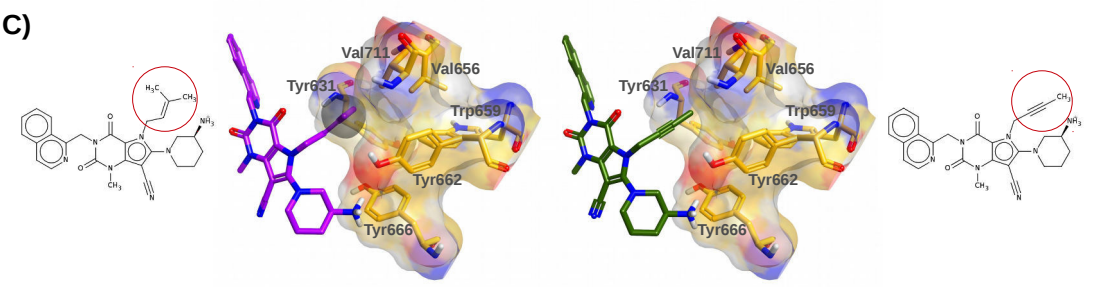
B)



(S)-4i (IC_{50}= 0.5 nM) **(S)-4f** (IC_{50}= 5 nM)

similarity: 0.928 **disparity**: 14.0

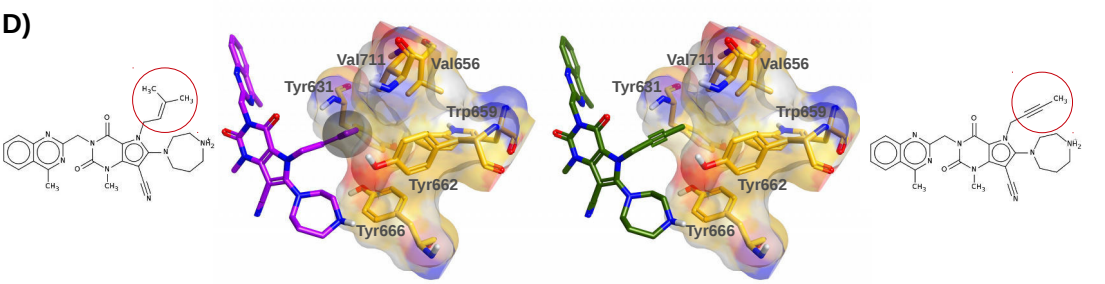
C)



(S)-11e (IC_{50}= 1 nM) **(S)-11j** (IC_{50} = 4 nM)

similarity: 0.980 **disparity**: 12.0

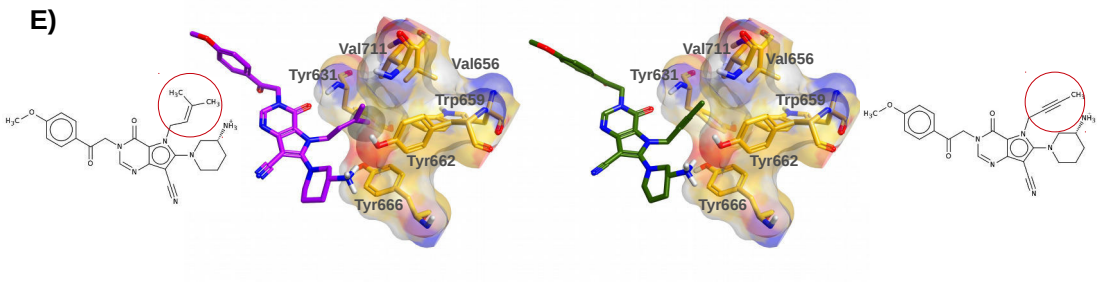
D)



12c (IC_{50}= 2 nM) **12h** (IC_{50} = 7.5 nM)

similarity: 0.958 **disparity**: 11.6

E)

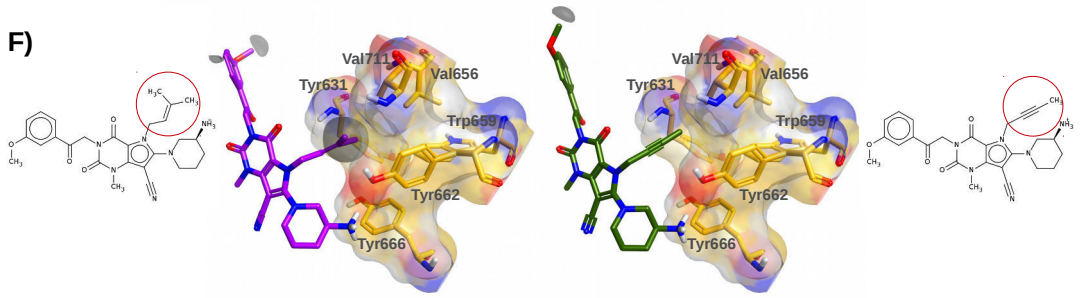


(R)-4g (IC_{50}= 4 nM)

similarity: 0.882 **disparity**: 10.0

(R)-4d (IC_{50}= 60 nM)

F)

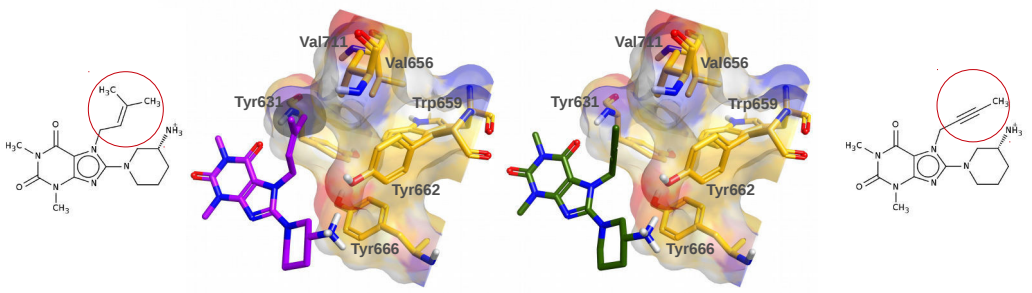


(S)-11d (IC_{50}= 0.5 nM)

similarity: 0.904 **disparity**: 6.2

(S)-11i (IC_{50}= 2 nM)

G)

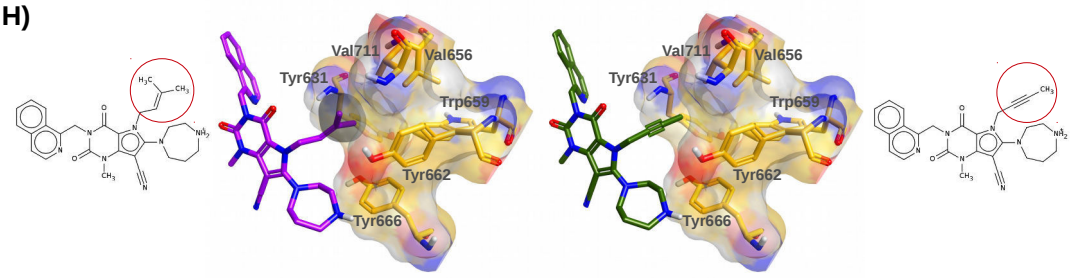


6ac (IC_{50}= 35 nM)

similarity: 0.916 **disparity**:4.7

6bc (IC_{50} = 88 nM)

H)



12b (IC_{50}= 3 nM)

similarity: 0.960 **disparity**: 4.4

12g (IC_{50} = 5 nM)

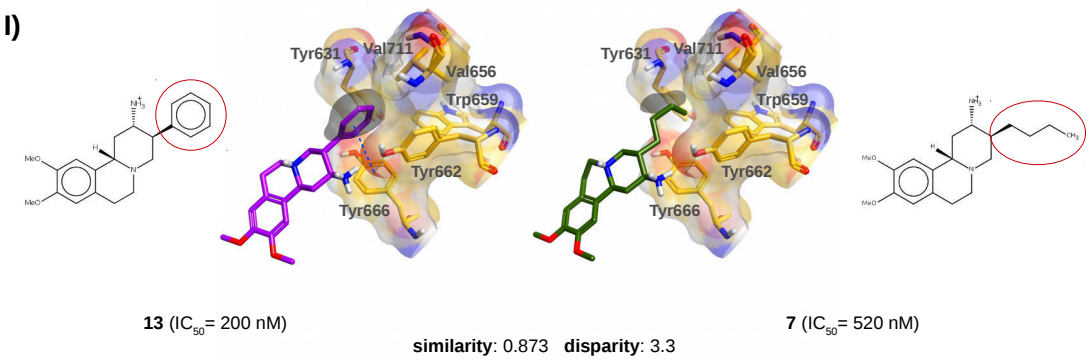
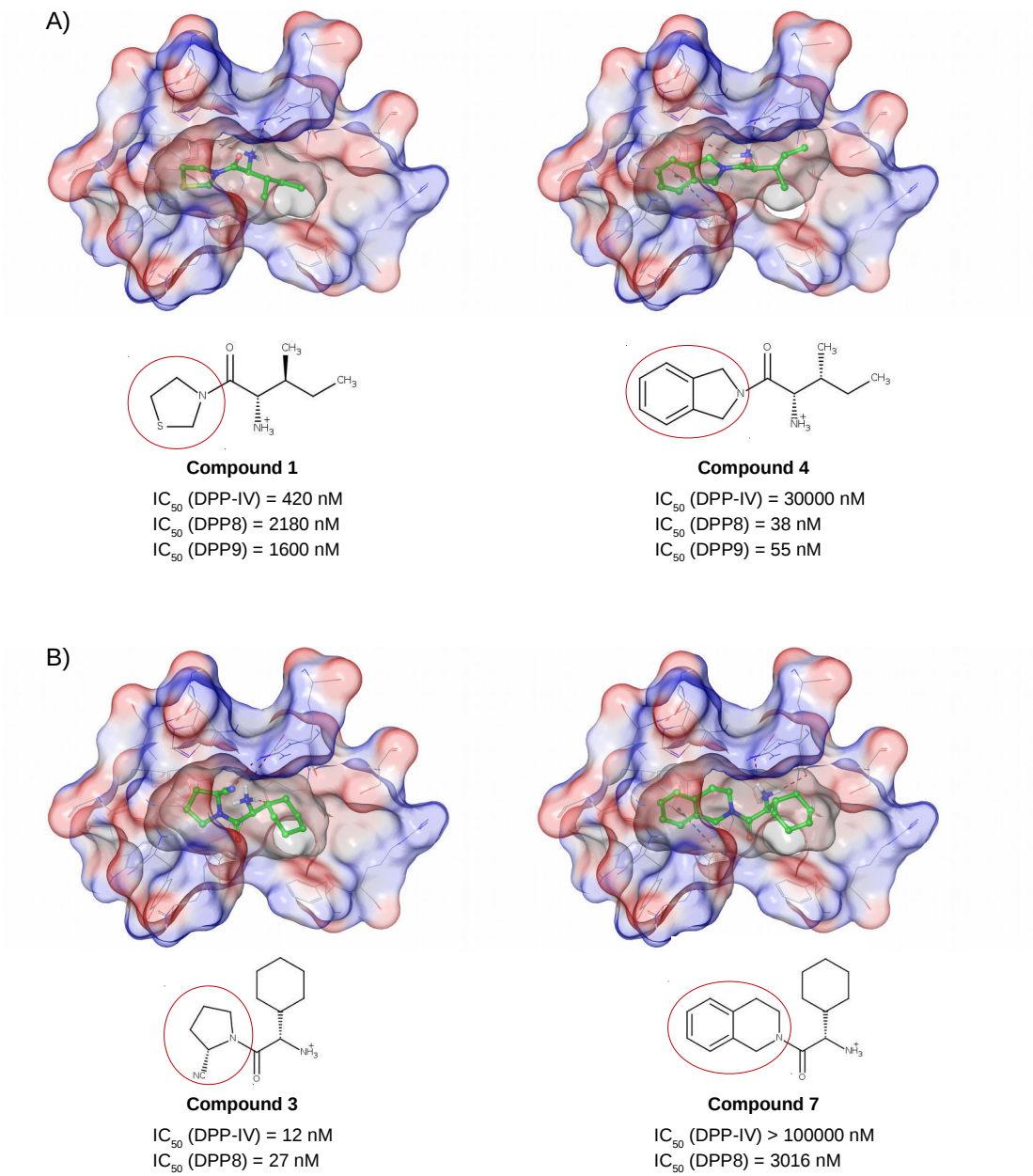


Figure S5 Comparison of the distribution of the hydrophobic surfaces between pairs of compounds that differ in their interactions with the S_1 subsite. For each panel the compound with the highest activity is shown in purple on the left and the compound with the lowest activity is shown in green on the right. Molecules are labeled with the same names that identify them in the corresponding paper.[142,146,149] The hydrophobic surface differences are shown in gray, while the protein surface has been colored according to atom color. The field surface difference is established by default at 2.0. In the 2D representation of each ligand, the structural differences between the compared compounds are highlighted. The different panels are arranged in order of decreasing disparity. The blue dashed line shows the π-π interaction between the **13** ligand and Tyr666[149] (calculated with the help of Maestro[165] using default options). In panels **A** to **H**, a but-2-yn-1-yl substituent in the S_1 subsite is replaced by a prenyl substituent,[142,146] while in panel **I** a monobutyl group is replaced by a phenyl group.[149] The ligand orientations are the result of their superposition with co-crystallized ligands from the same or very similar chemical series (*i.e.,* 4A5S[146] for panels **A** to **F**, 2RGU[142] for panel **G**, 4A5S[146] for panel **H** and 3KWF[144] for panel **I**; residue locations in each panel are also taken from the corresponding PDB file). This figure was obtained with the help of the Forge[152] and MarvinSketch programs.[178]



Compound 1

IC_{50} (DPP-IV) = 420 nM
IC_{50} (DPP8) = 2180 nM
IC_{50} (DPP9) = 1600 nM

Compound 4

IC_{50} (DPP-IV) = 30000 nM
IC_{50} (DPP8) = 38 nM
IC_{50} (DPP9) = 55 nM

Compound 3

IC_{50} (DPP-IV) = 12 nM
IC_{50} (DPP8) = 27 nM

Compound 7

IC_{50} (DPP-IV) > 100000 nM
IC_{50} (DPP8) = 3016 nM

Figure S6. Docked poses for pairs of compounds that basically differ in the moiety expected to bind to the S_1 pocket of DPP-IV (marked by a red circle in the corresponding 2D structure).[28,104] As shown by the corresponding IC_{50} values, when the size of this moiety is increased, the bioactivity for DPP-IV strongly decreases (whereas the DPP8/DPP-IV and/or DPP9/DPP-IV bioactivity relationships improve). The ligands were docked to 1X70[103] using GlideXP.[134] The molecular surfaces for the ligands and the binding site are shown and colored according to their electrostatic potential (where the surface for the binding site is more transparent than those for the corresponding ligands). This figure was obtained with the help of Maestro.[165]

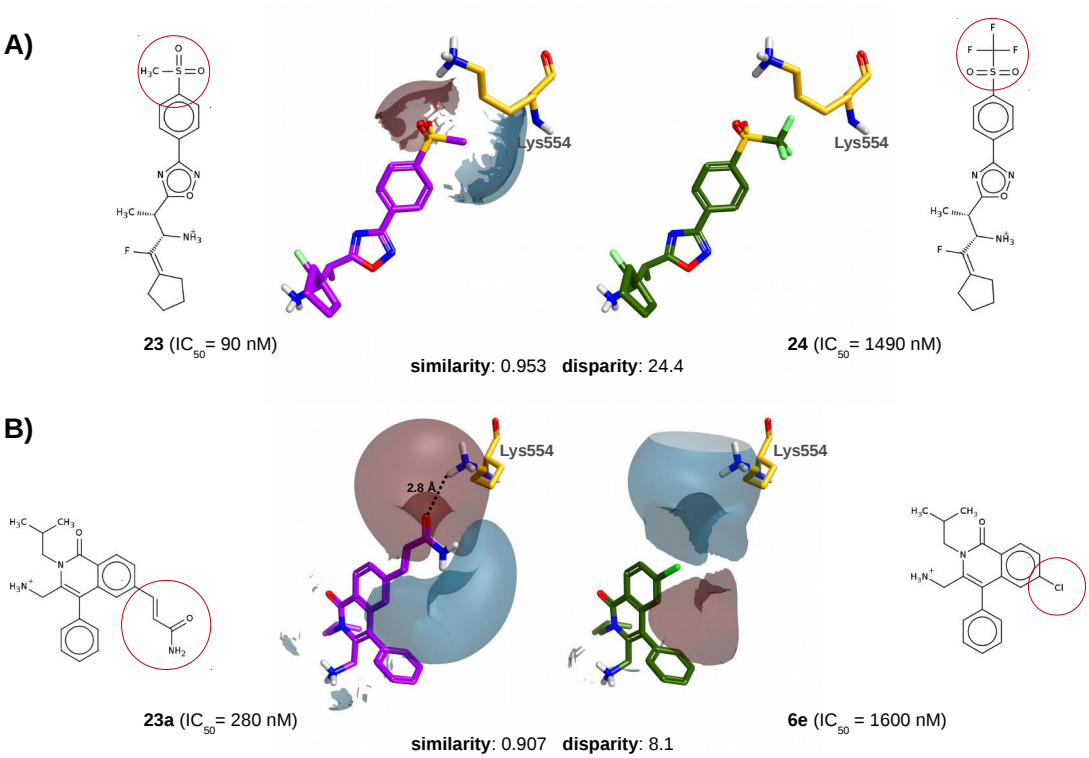
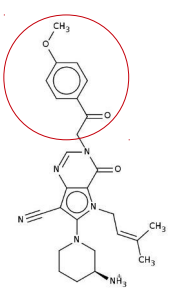
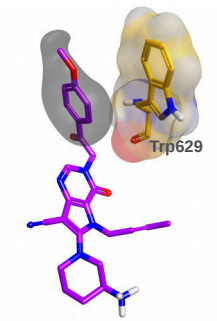


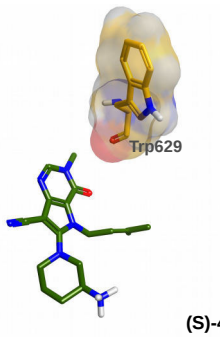
Figure S7. Comparison of the distribution of electrostatic surfaces between pairs of compounds that differ in their interactions with residue Lys554. For each panel the compound with the highest activity is shown in purple on the left and the compound with the lowest activity is shown in green on the right. Molecules are labeled with the same names that identify them in the corresponding paper.[124,145] The negative and positive electrostatic surface differences are shown in garnet and blue respectively (where the default value – *i.e.,* 2.0 – was used as the threshold for the surface difference between each pair). In the 2D representation of each ligand, the structural differences between the compared compounds are highlighted. The different panels are arranged in order of decreasing disparity and correspond to different situations: in panel **A** a methanesulfonyl group replaces a trifluoromethylsulfonyl substituent;[124] and in panel **B**, a 2-carbamoyleth-1-en-1-yl group replaces a **chlorine substituent**.[145] The hydrogen bond between Lys554 and compound **23a** in panel **B** is shown as a black dotted line. The ligand orientations are the result of their superposition with co-crystallized ligands from the same or very similar chemical series (*i.e.,* 3C45[124] for panel **A** and 3OPM[145] for panel **B**; residue locations in each panel are also taken from the corresponding PDB file). This figure was obtained with the help of the Forge[152] and MarvinSketch programs.[178]

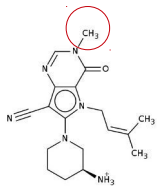
A)



(S)-4g (IC_{50}= 1 nM)



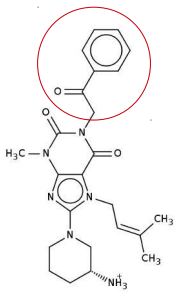




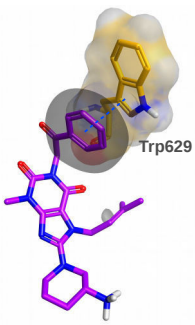
(S)-4l (IC_{50}= 9 nM)

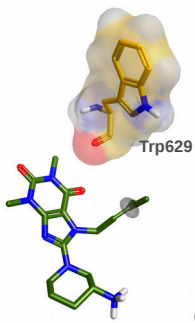
similarity: 0.830 **disparity**: 5.6

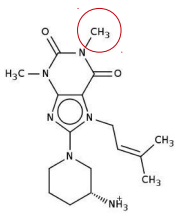
B)



6af (IC_{50}= 5 nM)



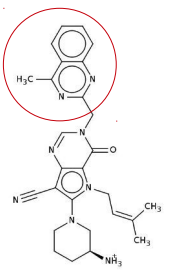




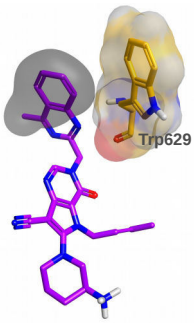
6ac (IC_{50}= 35 nM)

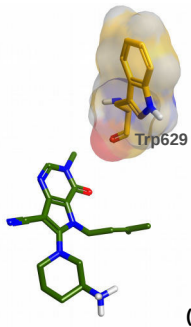
similarity: 0.771 **disparity**: 3.7

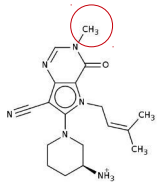
C)



(S)-4j (IC_{50}= 3 nM)



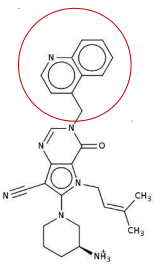




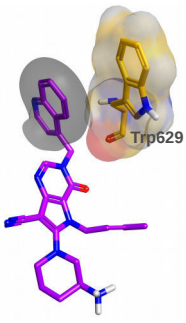
(S)-4l (IC_{50} = 9 nM)

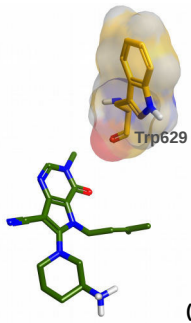
similarity: 0.830 **disparity**: 2.8

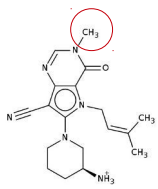
D)



(S)-4h (IC_{50}= 4 nM)







(S)-4l (IC_{50} = 9 nM)

similarity: 0.818 **disparity**: 1.9

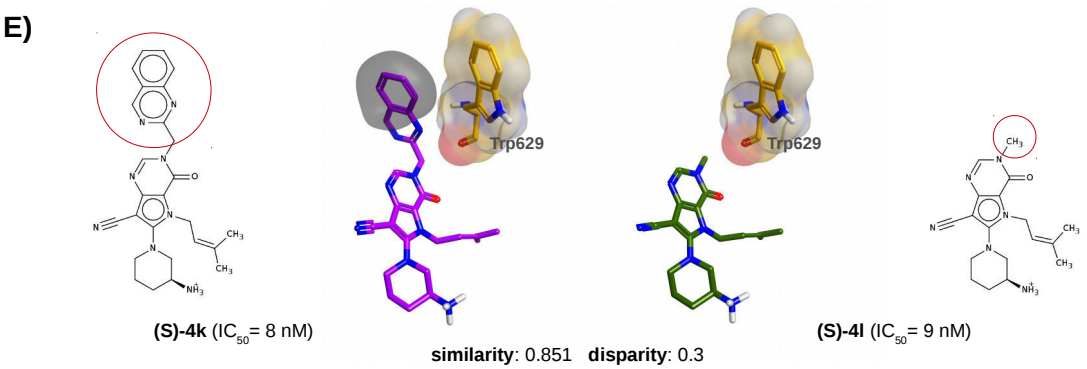
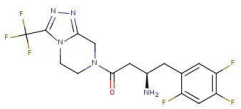
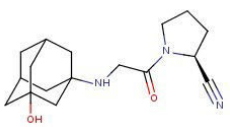
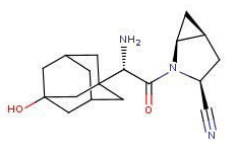
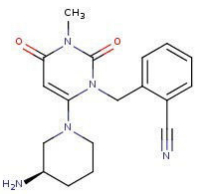
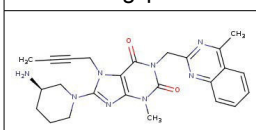
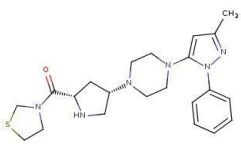
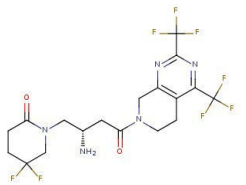
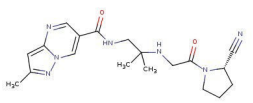
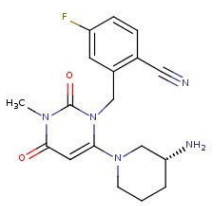
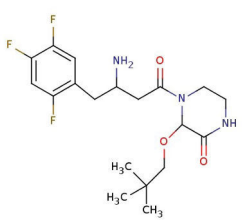
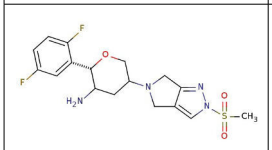


Figure S8. Comparison of the distribution of hydrophobic surfaces between pairs of compounds that differ in their interactions with the S_2' subsite. For each panel the compound with the highest activity is shown in purple on the left and the compound with the lowest activity is shown in green on the right. Molecules are labeled with the same names that identify them in the corresponding paper.[142,146] The hydrophobic surface differences are shown in gray, while the protein surface has been colored according to atom color. The field surface difference is established by default at 2.0. In the 2D representation of each ligand, the structural differences between the compared compounds are highlighted. The different panels are arranged in order of decreasing disparity. All panels correspond to the case where the extension to the S_2' pocket correlates with an improvement in DPP-IV activity.[142,146] The blue dashed line shows the π-π interaction between **6af** and Trp629 and was calculated with the help of Maestro[165] using default options. The ligand orientations are the result of their superposition with co-crystallized ligands from the same or very similar chemical series (*i.e.,* 4A5S[146] for panels **A**, **C**, **D** and **E** and 2RGU[142] for panel **B**; residue locations in each panel are also taken from the corresponding PDB file). This figure was obtained with the help of the Forge[152] and MarvinSketch programs.[178]

Table S1. Data on commercialized gliptins in order of year authorized by health agencies. IC_{50} values were taken from Reaxys Medicinal Chemistry and report data only for *in vitro* assays (*i.e.,* without using cells or tissues).

Compound	IC_{50} (nM)			Year of authorization by health agencies	Developed by
	DPP-IV	DPP8	DPP9		
 Sitagliptin	1.1 - 40	25220-48000	35600-100000	FDA, 2006 EMA, 2007	Merck & co.
 Vildagliptin*	3 - 120	900	680- 1300	EMA, 2007	Novartis
 Saxagliptin*	0.5 - 30	170 -244	61 -104	FDA, 2009 EMA, 2009	AstraZeneca and Bristol-Myers Squibb
 Alogliptin	2.63 - 24	> 100000	> 100000	Japan, 2010 FDA, 2013	Takeda Pharmaceutical Company
 Linagliptin	0.1 - 2	40000	10000	FDA, 2011	Boehringer Ingelheim
 Teneligliptin	0.29-0.37	260	540	Japan, 2012 Korea, 2014	Mitsubishi Tanabe Pharma

 Gemigliptin	16	169000	47000	Korea, 2012	LG Life Sciences
 Anagliptin	3.8	84700	56100	Japan, 2012	Sanwa Kagaku Kenkyusho Co., Ltd. and Kowa Company, Ltd
 Trelagliptin	4	100000	–	Japan, 2015	Takeda Pharmaceutical Company
 Evogliptin	0.98	6000-fold	6000-fold	Korea, 2015	Dong-A Pharmaceutical
 Omariglitpin	1.6	>67000	>67000	Japan, 2015	Merck & co.

* Vildagliptin and saxagliptin covalently bind to DPP-IV through Ser630.

Table S2. Summary of PDB codes for human DPP-IV (updated on 31.1.2017)

Apo form (10)

1J2E†, 1NU6†, 1PFQ, 1R9M†, 1TK3, 1U8E*, 1W1I, 4KR0, 4L72, 4QZV

DPP-IV/inhibitor complexes (87)

Complex with a ligand of oligopeptide nature (5)

1NU8†, 1R9N†, 1WCY†,2BGN, 2BGR

Complex with a ligand of drug-like nature (82)

Covalently bounded to DPP-IV through Ser630 (9)

1TKR, 2AJL, 2G5P, 2G5T, 2G63, 2I03, 2QKY†, 3BJM[1], 3W2T[2]

Non-covalently bounded to DPP-IV (73)

Goodness of fitting to the electron density map not validated because there are no structural factors available at the PDB (3)

1RWQ†, 2BUB†, 2JID†

Goodness of fitting to the electron density map not accomplished according to VHELIBS (7)

3OC0, **3Q8W**, **3VJK**[3], 4DSA, 4DSZ, 4J3J, 4PV7

Goodness of fitting to the electron density map validated according to VHELIBS (63)

1N1M, **1X70**[4], **2FJP**, 2HHA, 2I78, **2IIT**, **2IIV**, 2OAG, 2OGZ, 2OLE, 2ONC, 2OPH, 2OQI, 2OQV, 2P8S, **2QJR**, 2QOE, **2QT9**, **2QTB**, **2RGU**[5], 2RIP, 3C43, **3C45**, 3CCB, **3CCC**, 3D4L, 3EIO, 3F8S, **3G0B**[6], 3G0C, **3G0D**, **3G0G**, **3H0C**, **3HAB**, **3HAC**, **3KWF**[7], **3KWJ**, 3NOX, **3O95**, 3O9V, 3OPM, 3Q0T, 3QBJ, 3SWW, 3SX4, **3VJL**, **3VJM**, **3WQH**[8], 4A5S, 4DTC, 4G1F, 4JH0, 4LKO, 4N8D, 4N8E, **4PNZ**[9], **5T4B**, 5T4E, **5T4F**, **5T4H**, **5KBY**[10], **5I7U**, **5ISM**

Complexes with drug-like inhibitors with IC_{50} ≤ 10 nM that bind non-covalently to DPP-IV are shown in bold.

* PDB structures corresponding to mutants.

† No structural factors available in the PDB

[1] Complex with saxagliptin

[2] Complex with vildagliptin

[3] Complex with teneligliptin

[4] Complex with sitagliptin

[5] Complex with linagliptin

[6] Complex with alogliptin

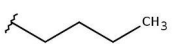
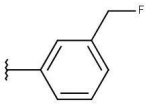
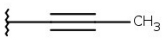
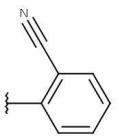
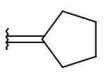
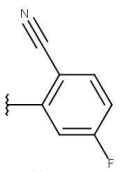
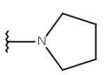
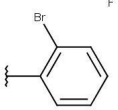
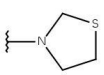
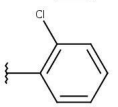
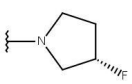
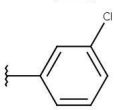
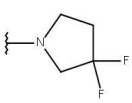
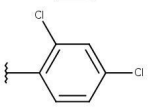
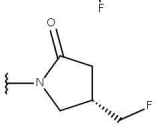
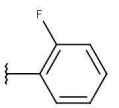
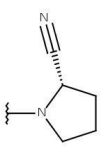
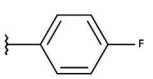
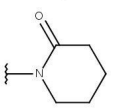
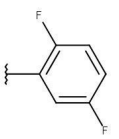
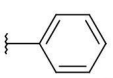
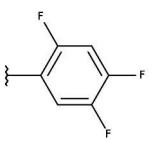
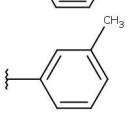
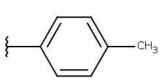
[7] Complex with carmegliptin

[8] Complex with anagliptin

[9] Complex with fluoroomarigliptin

[10] Complex with trelagliptin

Table S3. Ligand moieties that occupy the S_1 pocket in co-crystallized complexes between DPP-IV and non-covalently bound ligands of a drug-like nature.

Ligand moiety that binds at the S_1 pocket	PDB code	Ligand moiety that binds at the S_1 pocket	PDB code
	3OC0		3KWJ
	2RGU 5T4F 5T4B 5T4H 5T4E		2ONC 3G0D 3G0B 3G0G
	3C45		5I7U 5KBY
	1N1M 2HHA 2RIP		4G1F
	3VJK 3VJL 3VJM		3CCC 3G0C
	2FJP 2QT9 2OPH 3C43		3H0C 4N8E
	2QTB 3F8S		3NOX 4JH0 3Q0T 4LKO 3SWW 1RWQ 3SX4 4PV7
	3KWF		2BUB
	3WQH		2OGZ
	3QBJ		3D4L 5ISM
	3CCB 4A5S 3OPM 4N8D		1X70 2OLE 2QOE 4DTC 2I78 2OQI 3EIO 4DSA 2IIT 2OQV 3HAB 4DSZ 2IIV 2P8S 3HAC 4J3J 2OAG 2QJR 3Q8W 4PNZ
	2JID		
	3O9V 3O95		

Manuscript 2

A virtual screening strategy for mining large molecular databases to find new leads with low similarity to known actives: application to find new DPP-IV inhibitors

María José Ojeda-Montes[a], Àngela Casanova-Martí[b], Aleix Gimeno[a], Sarah Tomás-Hernández[a], Adrià Cereto-Massagué[a], Gerhard Wolber[c], Raúl Beltrán-Debón[a], Cristina Valls[a], Miquel Mulero[a], Montserrat Pinent[b], Gerard Pujadas[a,d,*] and Santiago Garcia-Vallvé[a,d]

[a]Research group in Cheminformatics & Nutrition, Departament de Bioquímica i Biotecnologia, Universitat Rovira i Virgili, Tarragona, Catalonia, Spain

[b]MoBioFood Research Group, Departament de Bioquímica i Biotecnologia, Universitat Rovira i Virgili, Tarragona, Catalonia, Spain

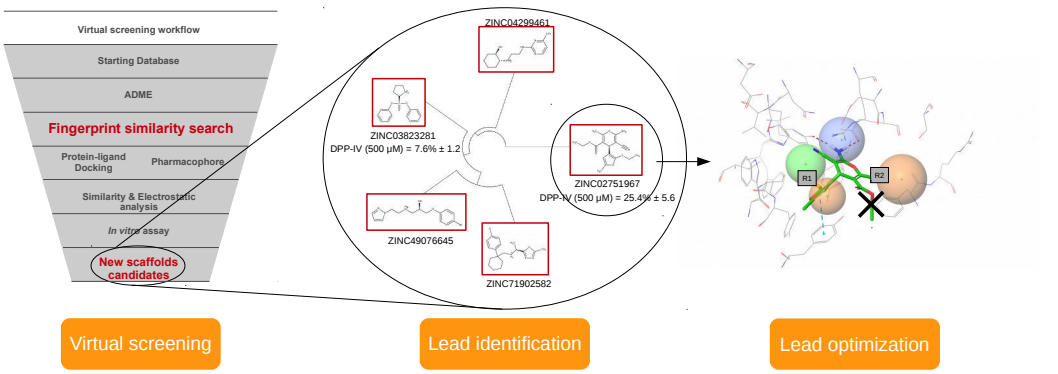
[c]Computer-Aided Drug Design, Institute of Pharmacy, Freie Universität Berlin, Berlin, Germany

[d]EURECAT, TECNIO, CEICS, Reus, Catalonia, Spain

**Correspondence to:*

Gerard Pujadas, Research group in Cheminformatics & Nutrition, phone: +34 977 55 95 65, fax: +34 977 55 82 32. Departament de Bioquímica i Biotecnologia, Facultat de Química, Universitat Rovira i Virgili, C/ Marcel·lí Domingo 1, Edifici N4, 43007 Tarragona, Catalonia, Spain. E-mail: gerard.pujadas@urv.cat

Abstract



Finding new leads is essential in projects to develop and discover new drugs. They can be found either by using *high throughput screening* (HTS) to experimental testing compound libraries or by using *virtual screening* (VS) to computationally predict the bioactivity of interest in files containing molecular databases. VS is cheaper than HTS but could lead to hits with high similarity to known actives and, therefore, of limited interest. Computational strategies to find new actives with low (or no) similarity to existing ones rely on either *fragment-based drug design* strategies or on *bioisosteric replacement*. Unfortunately, both computational approaches involve the synthesis of non-existing compounds (with the corresponding synthetic and purification difficulties) before experimentally it can be confirmed that the new compounds have the predicted bioactivity. Protein-ligand docking is a frequently used alternative but it can become computationally very expensive if large databases like ZINC need to be screened. Therefore, fast computational strategies need to be developed that can find completely new lead molecules in databases of purchasable compounds, thus removing the need to synthesize non-existing compounds before experimentally testing the bioactivity of the hits. Therefore, the main goal of this paper is the design of a computational strategy to look for new lead molecules with very low (or no) similarity to existing actives in databases of purchasable compounds and to apply it to a target of pharmacological interest (*i.e.*, DPP-IV). The bioactivity assays performed with the VS hits confirm that this strategy is able to quickly find completely new lead molecules with basal activity as DPP-IV inhibitors. Moreover, the computational strategy shown in this paper is sufficiently generic to be applicable to other targets of pharmacological interest.

1. Introduction

Finding new leads is an essential step in projects to develop and discover new drugs [1–4]. There are two alternatives for achieving this goal: **(a)** experimentally testing compound libraries to find molecules that show any level of the desired bioactivity (a process known as *high throughput screening*; HTS) [5–8]; and **(b)** computationally predicting the bioactivity of interest in files containing molecular databases (known as *virtual screening*; VS) [9–12]. Obviously, the VS alternative is significantly cheaper than HTS, but the fact that, frequently, the latter relies on several sequential filters that use characteristics of known actives to find new leads (*e.g.*, constrained protein-ligand docking, ligand or structure-based pharmacophores and shape/electrostatics similarity) is a major drawback because it can lead to VS hits with high chemical similarity to the known actives and, therefore, of limited interest.

There are a couple of computational strategies for finding new actives with low (or no) similarity to existing ones. The first consists of docking a fragment library at the target active site and then selecting the fragments with highest affinity and using linkers to join them (*i.e.*, *fragment-based drug design*) [13–15]. The second consists of replacing substructures involved in ligand-target interactions in a known active with other substructures that, although chemically different, can preserve equivalent intermolecular interactions with the target (*i.e.*, *bioisosteric replacement*) [16–18]. Unfortunately, both approaches involve the synthesis of non-existing compounds (with the corresponding difficulties associated to finding the proper synthetic plan and purifying the compound of interest) before it can be experimentally confirmed that the new compounds show the predicted bioactivity. Protein-ligand docking is a frequently used alternative but it can become computationally very expensive if large databases like ZINC [19] need to be screened. Therefore, fast computational strategies are needed to find new lead molecules with very low (or no) similarity to existing actives in databases of purchasable compounds. This would have the advantage not only of eliminating the need to synthesize non-existing compounds before experimentally testing the bioactivity of the hits, but also, that if one of these hits has the desired activity, their synthesis and purification is described and, therefore, this can be of help for using this lead molecule for performing the corresponding structure-activity

relationship studies by synthesizing new derivatives and finding which of them show improved bioactivity.

The goal of this paper is to design a computational strategy to look for new lead molecules with very low (or no) similarity to existing actives in databases of purchasable compounds and to apply it to a target of pharmacological interest (*i.e.*, DPP-IV) [20]. The strategy used initially discards those molecules in the database that show high similarity to known actives and applies a VS workflow to the remaining molecules (thus focusing computational resources only on those molecules that are of potential interest for finding new leads). The bioactivity assays performed with the VS hits obtained in this paper confirm that this strategy is able to quickly find completely new lead molecules with basal activity as DPP-IV inhibitors. Moreover, we use molecular modeling to suggest how the most potent VS hit could be used as a lead molecule to find derivatives with significantly improved bioactivity.

2. Results and Discussion

2.1. Description and validation of the virtual screening

A VS strategy was applied using sequential filters where the output molecules of one step were the input molecules for the next step and so on (see Figure 1). These steps mean that large libraries can be narrowed down to those molecules that are most likely to inhibit DPP-IV. Consequently, the following filters were applied in the present study: **(a)** ADME/tox filter (exclusively for VS purposes); **(b)** fingerprint similarity analyses with co-crystallized ligands (exclusively for VS purposes); **(c)** protein-ligand docking; **(d)** pharmacophore screening; and **(e)** shape/electrostatic analysis. It is important to note that the binding site of DPP-IV is quite rigid except for the residue Tyr547, which can adopt two different orientations [21]. Consequently, two different crystallized structures for DPP-IV (each with one of the two conformations for Tyr547; see Figure 2) were used during the protein-ligand docking step and the pharmacophore filter to take into account this flexibility.

The reliability of the VS workflow was evaluated using a starting database of 419 actives for DPP-IV and 15,084 decoys. Figure 1 shows how many actives and decoys remained after applying each VS step. The first filter applied to this set of actives and decoys — based on

Workflow steps	Validation			Virtual Screening
	Actives	Decoys	EF	ZINC Database
Starting Database	419	15,084		16,538,200
ADME QikProp v4.5	---	---		9,362,907
Fingerprint RDKit Torsion (Bottom 1%)	---	---		93,629
SP Docking Glide v6.8 **Pharmacophore** Phase v4.4	267	6,363	1.5	24,034
Similarity & Electrostatic analysis EON v2.2.0.5	101	137	10.5	404
***In vitro* assay**				5

Figure 1. The VS workflow used in the present study. The data corresponds to the number of molecules that remains after each VS step. The actives and decoys columns refer those molecules used during VS validation. The ZINC column refers to data obtained to find new leads for DPP-IV inhibition. During the validation of each step of the VS protocol, enrichment factors were calculated as the quotient between the fraction of actives in the sample that survived the VS step and the fraction of actives in the sample before the VS step.

docking and pharmacophore screening— placed the ligands in the binding site of the target requiring the most important interactions for the DPP-IV inhibition (*i.e.*, salt bridges/hydrogen bonds with the N-terminal recognition region, hydrophobic interactions with the S_1 pocket, interaction in S_2 pocket and π-π stacking with Tyr547) [22–25]. Thus, the numbers of actives and decoys were reduced to 267 and 6,363, respectively, but without a significant enrichment factor (*i.e.*, 1.5). However, the subsequent shape/electrostatic-potential comparison with co-crystallized ligands became a highly discriminative filter. Thus, Figures 3A and S1 show the distributions of the Tanimoto scores for actives and decoys provided by the different parameters calculated by EON [26]. A threshold for each parameter was established in order to recover the largest number of actives and remove the highest number of decoys. Then, a first cutoff of 0.7 for the coulombic part of the Poisson-Boltzmann electrostatics allowed us to remove nearly 70% of decoys and with almost no impact on the number of actives (*i.e.*, 2,039 and 236 molecules remain after the cutoff, respectively; see Figure 3A). In order to further concentrate the sample of active molecules, the Tanimoto scores for the remaining docked pose of each molecule were plotted again (see Figures 3B and S2). A second cutoff of 1.5 considering the combination of the coulombic part of the

Poisson-Boltzmann electrostatics and their shape fields resulted in 101 actives and 137 decoys (see Figure 3B). Therefore, this two-step shape/electrostatic-potential filter produces an enrichment factor of 10.5 relative to the previous pharmacophore screening because the number of decoys was strongly reduced (from 6,363 to only 137) in comparison with the number of actives (from 267 to 101). Overall, the enrichment factor of the complete VS workflow is 15.7 and, therefore, these results show that this VS protocol is able to discern those molecules that can inhibit DPP-IV from those that do not affect its activity.

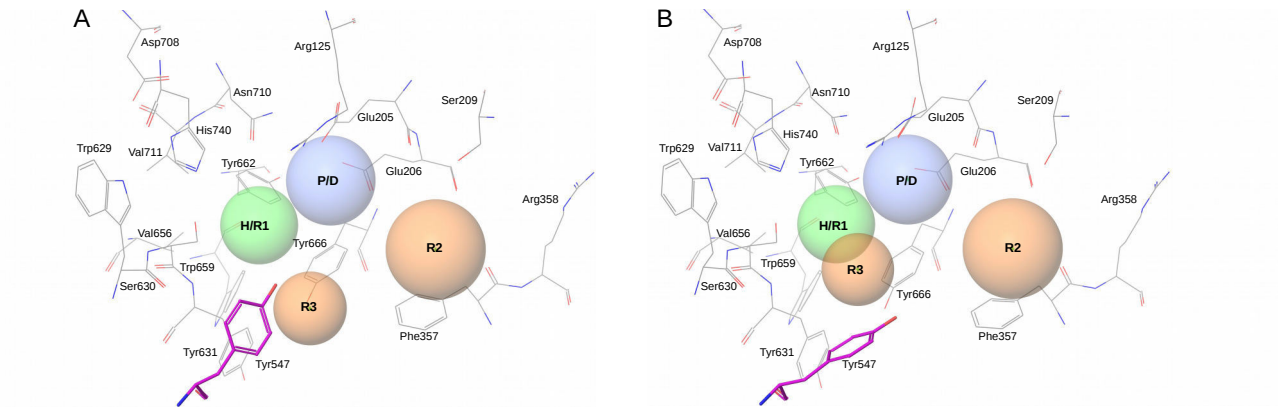
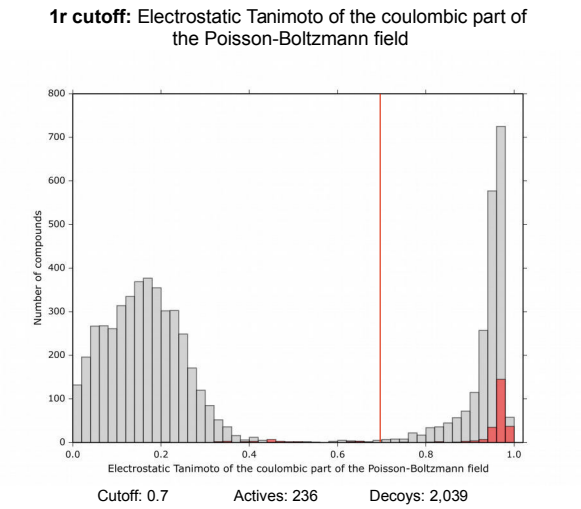


Figure 2. Structure-based pharmacophores used in this paper based on the crystal protein-ligand complex for the most important interactions. The difference between the two pharmacophores is due to the two different conformations of the residue Tyr547 (colored in pink) shown in the context of **(A)** the 1X70 active site and **(B)** the 3G0B active site. The pharmacophores are formed by a positive/hydrogen-bond donor feature (*i.e.*, **P/D**), a hydrophobic/aromatic ring site (*i.e.*, **H/R1**) and two aromatic ring sites (*i.e.*, **R2** and **R3**). The associated tolerances are 2.3Å for **P/D**, 2.0Å for **H/R1**, 2.5Å for **R2** and 1.8Å for **R3**. Two sites (*i.e.*, **P/D** and **H/R1**) together with a third site of the two remaining (*i.e.*, **R2** and **R3**) are required during the pharmacophore-based searches.

Once the VS workflow had been validated, it was decided to use the purchasable subset of the ZINC database [27] as the source of molecules for finding new lead structures with no (or very low) similarity to known DPP-IV inhibitors. To this end, the set of 16,538,200 molecules in that subset were first submitted to a filter aimed at discarding molecules that were either potentially toxic or exhibited poor ADME properties (see Figure 1). This filter reduced the number of molecules to be screened to 9,362,907 (see Figure 1). Then, the Tanimoto coefficient was calculated from the fingerprints in order to find those structures that could contribute to new scaffolds and that were significantly different from co-crystallized inhibitors. After sorting these 9,362,907 molecules in descending order of their Tanimoto coefficients,

A



B

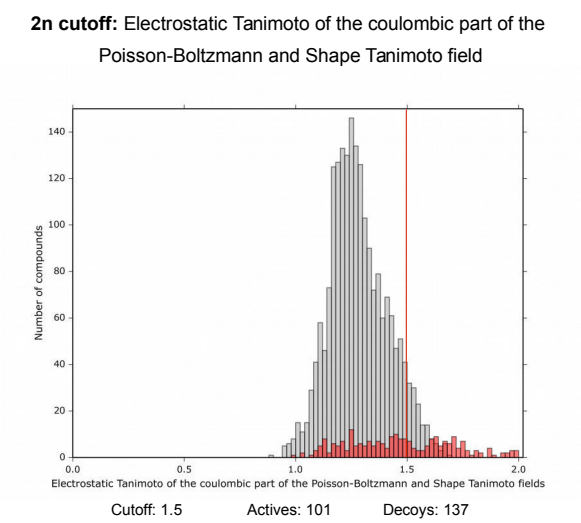


Figure 3. Histograms showing the distribution of the highest Tanimoto values for actives (shown in red) and decoys (shown in gray) for **(A)** the Electrostatic Tanimoto of the coulombic part of the Poisson-Boltzmann field and **(B)** the Electrostatic Tanimoto of the coulombic part of the Poisson-Boltzmann and Shape Tanimoto fields. Two consecutive cutoffs (red line) were applied to the set of actives and decoys by using these EON parameters in order to increase the enrichment factor of the VS validation.

only the bottom 1% of the resulting list (with values from 0 to 0.095 for that coefficient) was selected for the next filter (*i.e.*, 93,629 molecules). From the docking and pharmacophore screening, we obtained 24,034 compounds with at least one pose simultaneously filling the S_1 pocket and interacting with the Glu dyad (*i.e.*, **H/R1** and **P/D**, respectively) and either filling the S_2 pocket or interacting with Tyr547 (*i.e.*, **R2** and **R3**, respectively; see Figure 2). Finally, these poses were submitted to a shape/electrostatic comparison with known experimental poses for actives and only 404 structures were identified as VS hits with potential DPP-IV inhibitory bioactivity (see Figure 1).

2.2. Structure-activity relationship of selected compounds regarding the inhibition of DPP-IV

As mentioned before, the main goal of this study was to describe a computational strategy able to find new leads with no (or very low) similarity to known actives for DPP-IV. Thus, in order to select which of the 404 VS hits could be considered as new lead molecules and experimentally prove their bioactivity as DPP-IV inhibitors, we: **(a)** ensured that there were sufficient structural differences between the known DPP-IV inhibitors and the new lead molecules; **(b)** ensured that there were sufficient structural differences between the new lead molecules themselves; **(c)** visually inspected the docking poses; and **(d)** took into account commercial availability and cost. As result of these steps, the compounds ZINC04299461, ZINC03823281, ZINC02751967, ZINC49076645 and ZINC71902582 were selected in order to experimentally test their effects on the DPP-IV activity.

Figure 4 is a dendrogram in which the five selected hits are clustered according to their fingerprint similarity, thus revealing their structural diversity (*i.e.*, the maximum Tanimoto score among them is 0.0968 between ZINC49076645 and ZINC71902582). Moreover, a set of compounds (a total of 15,024 molecules) which have experimental bioactivity values for human DPP-IV were downloaded from Reaxys Medicinal Chemistry [28] for a fingerprint similarity analysis. After calculating the RDKit-Torsion [29] fingerprint for all of them, the highest Tanimoto coefficient was kept for comparison with the five VS lead candidates. As Figure 4 shows, the highest Tanimoto is for the compound ZINC03823281 (*i.e.*, Tanimoto value of 0.536) which shares part of its structure with XRN.24962630 while the remaining structures are sufficiently different from the associated molecule.

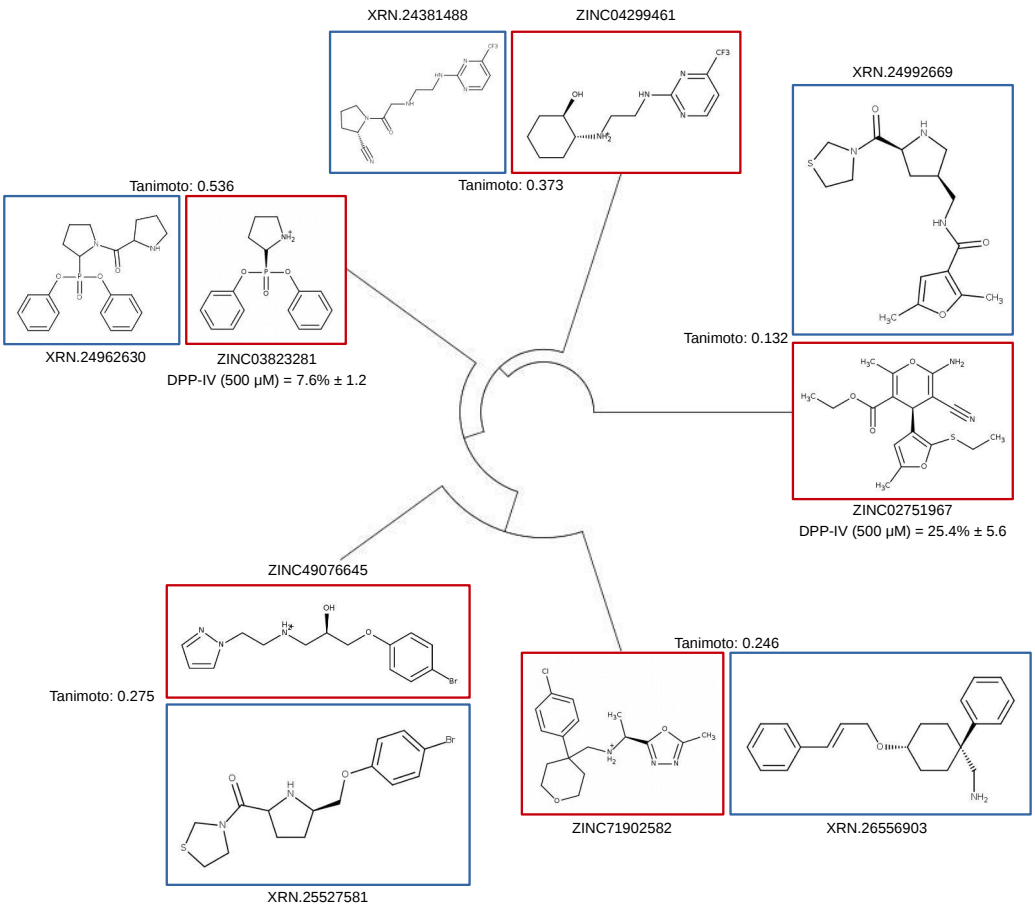


Figure 4. Dendrogram of the five hits selected for experimental testing as a result of the VS workflow (framed in red). The dendrogram shows the distances of the Tanimoto coefficient which represents the fingerprint similarity of the hits. Each hit is attached to a chemical structure downloaded from the Reaxys database which has experimental bioactivity values for human DPP-IV (framed in blue). This molecule is the most similar in terms of fingerprint similarity to the VS hit. Compounds ZINC02751967 and ZINC03823281 are the only ones which show significant *in vitro* DPP-IV inhibition.

The *in vitro* bioactivity assay shows that two out of these five selected compounds inhibit DPP-IV at 500 µM (*i.e.*, ZINC02751967 and ZINC03823281 significantly inhibited at 25.4% and 7.6%, respectively). Docking of these two compounds in the DPP-IV binding site (PDB entry 3G0B [30]) shows how these molecules match the main interactions determined by the pharmacophore (see Figure 5). From one side, both compounds use a primary or a secondary charged amine to interact with the N-terminal recognition region formed by the

residues Glu205, Glu206 and Tyr662 [31]. From the other side, the hydrophobic S_1 pocket (formed by the residues Tyr631, Val656, Trp659, Tyr662, Tyr666 and Val711) [32] is filled by different moieties (*i.e.*, ZINC02751967 places an ethylsulfanyl group whereas ZINC03823281 places a phenyl ring). Therefore, both ligands are able to match the main sites for DPP-IV inhibition, these being: **(a)** a salt bridge and/or hydrogen bond interactions with the N-terminal recognition region (*i.e.*, **P/D** site of the pharmacophore; see Figure 2); and **(b)** hydrophobic contacts with the S_1 pocket (*i.e.*, **H/R1** site of the pharmacophore; see Figure 2) [23–25]. Additionally, ZINC02751967 and ZINC03823281 place a 5-methylfuran-3-yl group and a phenyl ring at **R3**, respectively, which allows them to interact with Tyr547 through a π-π stacking (see Figure 5). Tyr547 has been reported to be essential for the catalytic activity of the enzyme. Moreover, due to Tyr547 flexibility in DPP-IV (which is not possible in DPP8/9) it has been suggested, to be also involved in inhibitor selectivity [33,34]. Finally, in the case of ZINC02751967, the cyano group is able to place a negative electrostatic environment around Arg125 (results not shown), whereas ZINC03823281 can interact with this residue by cation-π interaction (see Figure 5).

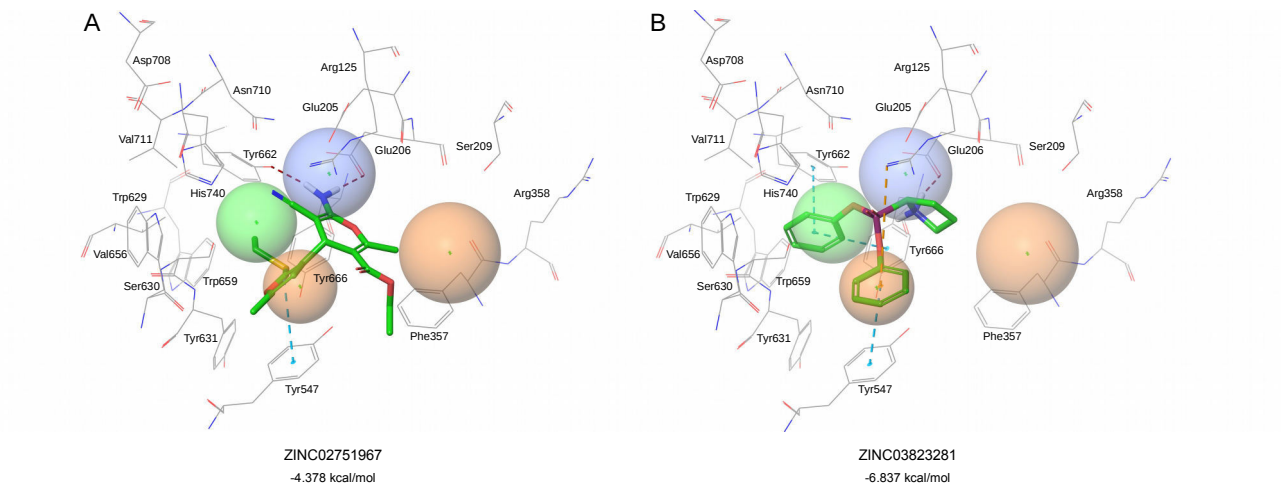


Figure 5. The best docked poses (with the corresponding XP GScore) for the compounds ZINC02751967 and ZINC03823281. Blue **and** orange dashed lines show π-π stacking and cation-π intermolecular interactions, respectively, whereas the red ones show either salt bridges (between the positively charged amine from ZINC03823281 and Glu**206**) or hydrogen bonds. Both panels are oriented the same way for easy comparison.

2.3. Lead optimization from ZINC02751967

The *in vitro* experiments demonstrated that, of the two VS hits that are bioactive as a DPP-IV inhibitors, ZINC02751967 is the more potent (see Figure 4). Consequently, this molecule is a promising lead compound for designing potent and selective DPP-IV inhibitors with very low similarity to existing actives. The lowest-energy docked pose of ZINC02751967 was used as the starting point for lead optimization (the XP GScore was -4.378 kcal/mol; see Figure 5A) with the purpose of improving the binding affinity with DPP-IV according to the most important interactions described for this target [22–25]. Therefore, an optimization process has been developed (see Figure 6) in order: **(a)** to improve the occupation of the hydrophobic S_1 pocket (*i.e.*, the **H/R1** site of the pharmacophore); and **(b)** to reach the S_2 extensive subsite (*i.e.*, the **R2** site of the pharmacophore).

Firstly, the ethoxycarbonyl group was removed from the original structure of ZINC02751967 (see Figure 6A) because this fragment is not able to interact with any residue (see Figure 5A) and because of its very low contribution to the XP GScore (from -4.378 for ZINC02751967 to -4.122 kcal/mol for ZINC02751967 without the ethoxycarbonyl moiety). Moreover, this new ZINC02751967 derivative (*i.e.*, ZINC02751967-dev) allowed us to attach bigger substituents in order to reach the pockets due to the reduction of its molecular weight (which is within the parameters of Lipinski's rule during the combinatorial screening). Next, the lead optimization process initially focused on introducing a moiety that could better fill the hydrophobic S_1 cavity than the initial ethylsulfanyl group (*i.e.*, **R1** label; see Figure 6A). As a result, ZINC02751967-dev-283 was selected on the basis of its XP GScore (*i.e.*, -7.072 kcal/mol; see Figure 6B). At this point, it is important to note that it has been experimentally shown that a better occupancy of the S_1 pocket results in higher bioactivities for DPP-IV inhibitors [35,36]. Consequently, the substitution of the original ethylsulfanyl substituent by a positively charged pyridin-4-yl moiety is expected to help increase the bioactivity of ZINC02751967-dev-283 relative to ZINC02751967-dev. Moreover, the docked pose of this new compound, not only maintains the original intermolecular interactions with the DPP-IV binding site (*i.e.*, two hydrogen bonds with residues of the N-terminal recognition region and the π-π stacking interaction with Tyr547) but also shows additional π-π and cation-π interactions with Tyr666 (which reinforces the hypothesis that ZINC02751967-dev-283 is a better DPP-IV inhibitor than ZINC02751967-dev). At this point, it is interesting to remark that

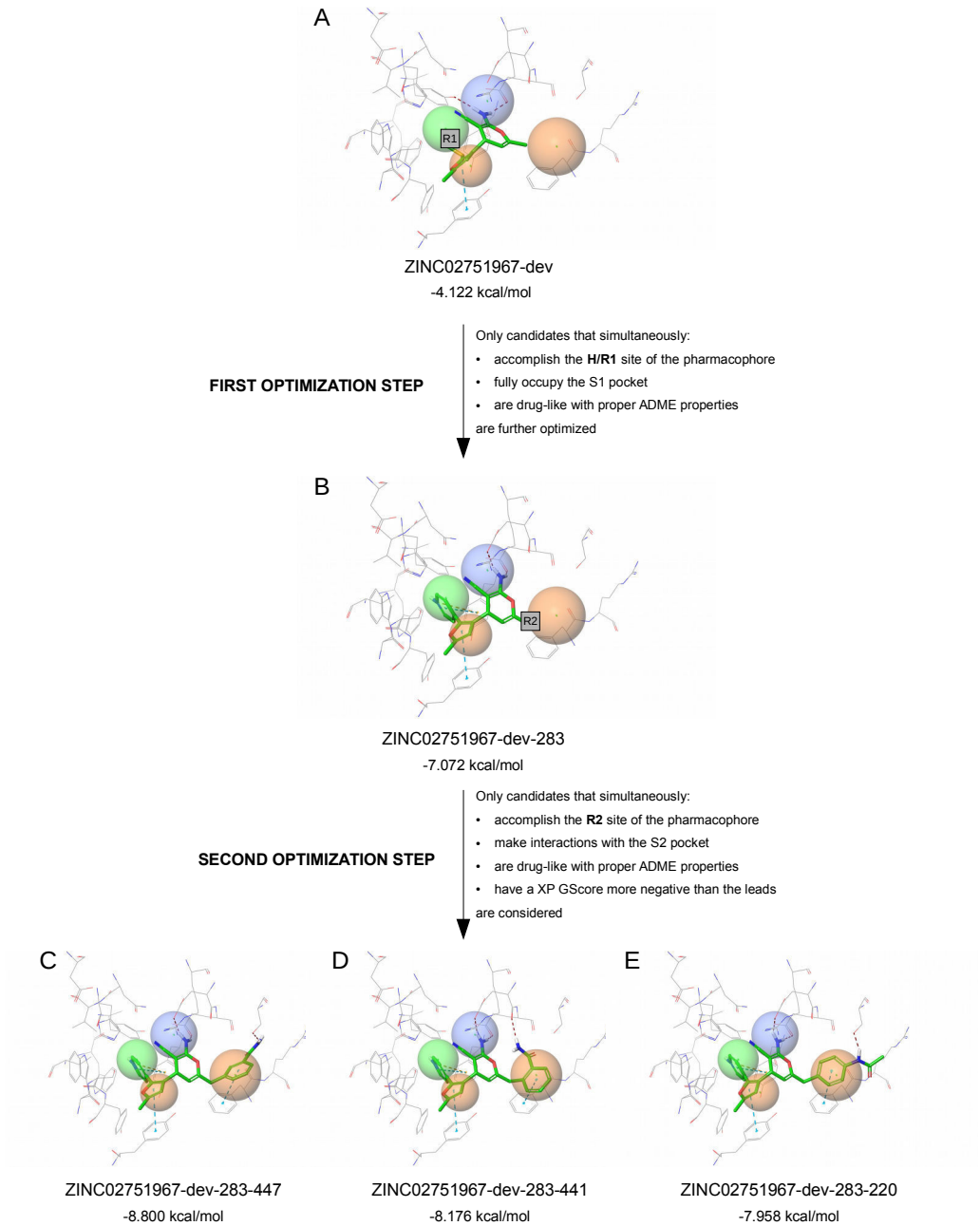


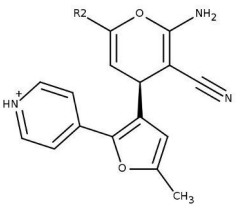
Figure 6. Lead optimization of a derivative of ZINC02751967 used with the aim of obtaining new molecules with improved potency and selectivity for DPP-IV. First, the ethoxycarbonyl group was removed from the initial ZINC02751967 (Figure 6A) because of its low contribution to the protein-ligand interaction (see Figure 5A). Then a substituent was attached to the ethylsulfanyl group of this ZINC02751967 derivative (*i.e.*, R1 label) in order to

improve the occupancy of S_1 pocket. The resulting derivative (Figure 6B) was selected for further optimization. Next, another point for attaching the substituents (*i.e.*, R2 label) was placed in these new derivatives in order to reach the S_2 extensive subsite. The docked poses of some of the most potent derivatives after this second optimization step are shown (Figure 6C-6E). The name for each derivative was built by adding the code of the attached fragment (according to the CombiGlide Diverse Side-chain Collection) to the lead name (see also in the Table 1 the 2D structure and XP GScore for the best ten derivatives obtained during the second optimization step). Blue and orange dashed lines show π-π stacking and cation-π intermolecular interactions, respectively, whereas the red ones show the hydrogen bonds.

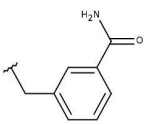
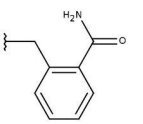
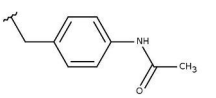
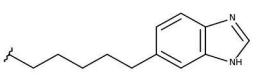
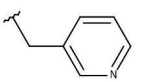
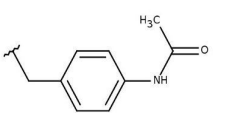
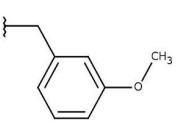
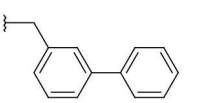
these results are coherent with those previously found by our group that show that improvements in the bioactivity of DPP-IV inhibitors can be obtained by replacing an alkyl substituent at the **H/R1** site by a group that is able to interact with the lipophilic atoms of the S_1 pocket either by producing the so-called hydrophobic enclosure reward (that would consist of enclosing the two sides of the substituent –at a 180º angle– on the hydrophobic environment of the S_1 pocket) or by making π-cation interactions with the aromatic side chains in this pocket [37]. In that sense, the positively charged pyridin-4-yl substituent of ZINC02751967-dev-283 would be able to perform, simultaneously, both kind of interactions.

A second optimization step was performed in order to reach the S_2 extensive subsite. This pocket has been shown to enhance the activity and selectivity of DPP-IV by interacting with Ser209, Phe357 and Arg358 [24,36,38]. Consequently, another point of attachment was placed in ZINC02751967-dev-283 (*i.e.*, **R2** label; see Figure 6B). The top ten derivatives of this second optimization step had docked poses that were able to further increase the XP GScore (*i.e.*, in the -8.800 to -7.319 Kcal/mol range; see Table 1). Most of them are expected to increase the binding affinity and selectivity for DPP-IV by either interacting with Ser209 through a hydrogen bond or by making a π-π stacking interation with Phe357 (see Figures 6C, 6D and 6E). The 2D structures for the ten best docked poses and their corresponding XP GScore values are shown in Table 1.

Table 1. Best ZINC02751967-dev-283 derivatives obtained in the second optimization step. Molecules are sorted according to XP GScore. The name for each derivative was built by adding the code of the attached fragment (according to the CombiGlide Diverse Side-chain Collection) to the lead name.



ZINC02751967-dev-283

ZINC02751967-dev	R2 substituent	XP GScore (Kcal/mol)
283-447		-8.800
283-441		-8.176
283-220		-7.958
283-312		-7.706
283-278		-7.651
283-500		-7.578
283-386		-7.418
283-236		-7.319

3. Conclusions

The design of the computational strategy used in this study has been demonstrated to be suitable for identifying new lead compounds in purchasable databases with very low (or no) similarity to known actives. Therefore, this VS workflow is a good alternative to other computational approaches such as bioisosteric replacement and fragment-based drug design because it reduces the cost and time of designing new potent actives; that is, by using this VS workflow, the synthetic effort focuses solely on improving a core structure with the desired basal bioactivity for the target of interest. Moreover, this computational strategy is significantly faster than protein-ligand docking (which can be also used for finding new leads in molecular databases of purchasable compounds). For instance, benchmarking studies have estimated that the fastest docking mode available in Glide (*i.e.*, HVTS) needs around 1.5 seconds to dock a ligand in a binding site by using a 2.2 GHz Opteron processor (*i.e.*, around 60,000 compounds per day) [39]. In contrast, by using a similar processor, we can reduce an initial sample of 9,362,907 molecules to the 1% of interest in only 6 hours (thus focusing computational resources only on those molecules that are potential candidates for finding new leads). Finally, the VS workflow described here is enough general to be valid to other targets of pharmacological interest.

4. Experimental

4.1. Set up of the starting databases for validating the virtual screening workflow and for lead discovery

The ability of our VS workflow to identify DPP-IV inhibitors was validated by applying it to an initial set of known 419 actives and 15,084 decoys (one active per 36 decoys; see Figure 1). This set of known DPP-IV inhibitors was formed by molecules with a high activity value [*i.e.*, $pX \geq 7$; where pX is calculated as the $-\log_{10}$ using the value for different activity measures (*e.g.*, IC_{50}, K_i, % of inhibition, etc.) with the goal of normalizing the bioactivity data from different experiments] and was obtained from Reaxys Medicinal Chemistry [28]. Before the VS, the 3D structures of the actives set were generated with Omega [40] allowing just one conformation for each input molecule. Furthermore, the decoys were obtained with a

modified version of DecoyFinder [41] that selected as decoys any molecule with a molecular weight (MW) within the range of the actives (resulting in a MW range of 215-586 Da for actives and of 300-440 Da for decoys).

Ligands for lead discovery purposes were downloaded from the purchasable subset of the ZINC12 database [27], which contains more than 16 million compounds. QikProp v4.5 [42] was then used to filter the ZINC molecules to discard those with bad ADME/Tox properties. Thus, only molecules that simultaneously fulfill the following drug-like properties were considered for the next step of the VS workflow: **(a)** MW at the 300-700 Da range; **(b)** only one violation for Lipinski's rule of five (predicted through the *RuleOfFive* parameter) [43]; **(c)** a maximum of 2 reactive functional groups (predicted through the *rtvFG* parameter); **(d)** a high human oral absorption (predicted through the *HumanOralAbsorption* parameter, which must have a value of 3) [44]; and **(e)** number of property or descriptor values that fall outside the 95% range of similar values for known drugs at the 0-5 range.

One of the most important challenges of this VS workflow is that it should be able to find new lead molecules with basal DPP-IV bioactivity but with extremely low similarity to known active compounds. Consequently, the RDKit-Torsion fingerprint [29,45] was used to *label* all ZINC molecules that fulfill the previously described ADME/Tox filter according to their chemical structure. The same procedure was used to label 33 known inhibitors co-crystallized with DPP-IV with an IC_{50} lower than 100 nM (see Table 2). Then the similarity

Table 2. Ligands from these PDB codes of DPP-IV complexes were used as references to calculate Tanimoto similarity with ZINC molecules based on RDKit-Torsion fingerprint. Only the bottom 1% of the sorted Tanimoto list was kept for the following VS steps.

PDB codes										
1X70	2FJP	2IIT	2IIV	2OGZ	2OLE	2ONC	2OPH	2P8S	2QJR	2QOE
2QT9	2QTB	2RGU	2RIP	3C43	3C45	3D4L	3F8S	3G0B	3G0D	3G0G
3H0C	3HAB	3HAC	3KWF	3KWJ	3O95	3VJL	3VJM	4A5S	4G1F	4J3J

between the RDKit-Torsion fingerprints of each co-crystallized inhibitor and each ZINC molecule was calculated using the Tanimoto coefficient. For each comparison between one specific ZINC molecule and the different co-crystallized inhibitors, only the highest Tanimoto

value was kept. Finally, ZINC molecules were sorted by decreasing the Tanimoto coefficient and only the bottom 1% of the sorted list was kept for the next step of the VS filter (see Figure 1).

4.2. Description and validation of the virtual screening

4.2.1. Ligand and protein setup

Before the docking filter, the 3D structure of all the remaining molecules (either actives or decoys or ZINC molecules for lead discovery) was prepared with LigPrep v3.5 [46] with the following settings: **(a)** the force field OPLS 2005 was used; **(b)** all possible ionization and tautomerization states at pH 7.0 ± 2.0 were generated with Epik; **(c)** the desalt option was activated; **(d)** chirality from input geometry was kept when generating stereoisomers; and **(e)** one low-energy ring conformation per ligand was generated.

The Protein Preparation Wizard (PPW) panel [47] was used to set up DPP-IV protein for use as a target in the following VS steps. Thus, chain A was prepared for two different PDB entries (*i.e.*, 1X70 and 3G0B [30,48]) in order to cover both possible positions of residue Tyr547 (for which the dihedral angle changes by 70° between the two orientations; see Figure 2) [21]. During the processing and refining steps of the PPW, all options were set to default with the exception of *remove original hydrogens*, *fill in missing side chains* and *cap termini* options, which were set to *on.*

4.2.2. Protein-ligand docking during the VS

During the VS workflow, protein-ligand docking studies were carried out using Glide v6.8 [49] with the following settings: **(a)** two different binding sites for DPP-IV were defined by using the previously curated coordinates of the two PDB files (*i.e.*, 1X70 and 3G0B) with the *Schrödinger's Grid Generation* panel (default options were used); **(b)** the standard precision mode (*i.e.*, SP) was used; **(c)** the maximum number of poses per ligand was increased to 32; and **(d)** the number of poses per ligand included in the post-docking minimization was increased to 32. The default values were used for the remaining docking parameters.

4.2.3. Structure-based pharmacophore screening

Docked poses were filtered through a couple of structure-based pharmacophores that were built to take into account the two different conformations that Tyr547 can adopt (see Figure 2). In this regard, docked poses obtained with 1X70 were filtered with the pharmacophore shown in Figure 2A whereas those obtained with 3G0B were filtered with the pharmacophore shown in Figure 2B. Both structure-based pharmacophores were designed by considering the most important interactions described for DPP-IV inhibition [22–25]. Thus, the two pharmacophores share three of the sites (*i.e.*, **P/D**, **H/R1** and **R2** that are associated, respectively, with interactions with the Glu205/Glu206 dyad, the S_1 pocket and the S_2 extensive subsite), whereas the location of the fourth site (*i.e.*, **R3**) depends on the conformation adopted by Tyr547 (see Figure 2). Associated tolerances are 2.3Å for **P/D**, 2.0Å for **H/R1**, 2.5Å for **R2** and 1.8Å for **R3**. Docked poses were filtered by the corresponding pharmacophore by using Phase v4.4 [50] and with the *score in place* option set to on (*i.e.*, no re-orientation of the docked poses was allowed during the search). Thus, only docked poses simultaneously matching at least three pharmacophore sites (*i.e.*, **P/D**, **H/R1** and either **R2** or **R3**) were kept for the next VS filter.

4.2.4. Electrostatic and shape similarity screening

The software EON v2.2.0.5 [26] compares the poses for two different compounds by calculating the Tanimoto coefficients associated with either their the electrostatic potentials (*i.e.*, the Poisson-Boltzmann electrostatics and the coulombic part of the Poisson-Boltzmann electrostatics), their shape or the combination of the Poisson-Boltzmann electrostatics and their shape. Thus, for the Poisson-Boltzmann electrostatic and for the coulombic part of the Poisson-Boltzmann electrostatic, the Tanimoto score is in the -⅓ to 1 range (where a value of 1 corresponds to identical electrostatic potential overlap whereas negative values correspond to the overlap of opposite charges between the two poses). For the shape, the Tanimoto score is a quantitative measure of three-dimensional overlap between the two poses where 1 corresponds to a perfect overlap (*i.e.*, the same shape) and 0.

Fifteen complexes between DPP-IV and potent and selective non-covalent inhibitors (*i.e.*, 1X70, 2FJP, 2HHA, 2OLE, 2OPH, 2P8S, 2QOE, 2RGU, 3G0B, 3HAB, 3HAC, 3KWF, 3O95,

3WQH, 4PNZ) were superposed to 1X70 and 3G0B with the help of the PPW panel [47]. Then the experimental poses of their ligands were used during EON comparisons with the docked poses of the actives and decoys that passed the previous VS filter. This allows to find which of the Tanimoto scores provided by EON (and considering too the combination of the coulombic part of the Poisson-Boltzmann electrostatics and their shape) and which threshold for these scores produces a better enrichment factor (see Figures 3, S1 and S2) and to determine the influence of the combination of the coulombic part of the Poisson-Boltzmann electrostatics and their shape). The highest value obtained for each Tanimoto score was kept from each comparison of one docked pose with the set of fifteen experimental poses.

4.3. Hit selection for further experimental assays on DPP-IV activity

The RDKit-Torsion fingerprints [29] of all the ZINC molecules that passed all the VS workflow filters were clustered on the basis of their Tanimoto similarities. Five structurally different compounds were then selected for *in vitro* assays of DPP-IV inhibitory activity on the basis of their commercial availability, cost and low chemical similarity to any molecule that has been experimentally shown to be bioactive as a human DPP-IV inhibitor (see Figure 4). These compounds were ZINC04299461, ZINC03823281, ZINC02751967, ZINC49076645 and ZINC71902582 and they were purchased from either Ambinter c/o Greenpharma (Orléans, France) or Epsilon Chimie (Brest-Guipavas, France).

4.4. In vitro assay of selected compounds' inhibition of DPP-IV

The DPP-IV enzyme purified from porcine kidney (product number 317640, Merck Millipore Corporation) was used to evaluate the effect of the selected compounds on DPP-IV activity. Stock solutions of the assayed compounds were made in DMSO and diluted in buffer (50 mM Tris-HCl) to a final concentration of 500 µM. The final concentration of DMSO in the assay was 1%. The DPP-IV enzyme (diluted with 100 mM Tris HCl buffer pH 8.0 to 0.26 mU per well) and 10 µL of a test sample with a different concentration was pre-incubated for 10 min at 37°C using 96-well microplates to allow compound/enzyme interaction. Next, the enzymatic assay was initiated by the addition of 50 µL of the fluorimetric substrate H-Gly-Pro-AMC [product number I-1225, purchased from Bachem (Bubendorf, Switzerland)] at a final concentration of 0.01mM. Fluorescence was measured in a Biotek FLx800

Fluorescence Microplate Reader at Ex:380nm/Em:460nm and 37°C for 30 min. Sitagliptin, a well-known DPP-IV inhibitor (that non-covalently binds to DPP-IV), was used as reference inhibitor and positive control. At least three independent assays were performed, each with two technical replicates. DPP-IV inhibition is expressed as a percentage, which is the difference of the activity in presence of test compounds versus the total activity of enzyme. Significant results showed $p<0.05$ with a Student's T test (SPSS software; SPSS, Chicago, USA).

4.5. Lead-optimization from ZINC02751967

Lead optimization was performed with CombiGlide v3.9 [51] by using the Virtual Combinatorial Screening workflow. The core-containing molecule was a derivative of the lowest-energy docked pose for ZINC02751967 with the ethoxycarbonyl moiety of the original molecule removed (compare Figures 5A and 6A) and the substituents were obtained from the Schrödinger CombiGlide Diverse Side-chain Collection v1.2 [52] (which contains all reasonable ionization and tautomeric states for a collection of 817 representative functional groups commonly found in pharmaceuticals, with linkers of variable lengths). In two consecutive steps we established the points where the substituents had to be attached with the aim of improving the interactions at the: **(a)** S_1 pocket; and **(b)** S_2 extensive subsite of DPP-IV.

Thus, the following parameters were used for a single-position docking run: **(a)** the receptor grid for PDB code 3G0B was the same as the one previously used at protein-ligand docking step; **(b)** the *apply Glide core constraints* option was used within a maximum RMSD of 1.0Å; **(c)** the *Fully enumerated* option was selected; and **(d)** the *CombiGlide XP docking* mode was used. After this process, the resulting derivatives were filtered with the ADME filter set as *Druglike*. During this ADME filter, only two violations of the following criteria were allowed: **(a)** molecular weight less than 500 Da; **(b)** a maximum of 5 hydrogen bond donors; **(c)** a maximum of 10 hydrogen bond acceptors; **(d)** a predicted octanol/water partition coefficient (*i.e.*, logP) less than 5; **(e)** 10 or fewer rotatable bonds; and **(f)** a 150 $Å^2$ or less Van der Waals surface area. Next, the derivatives were filtered with the same pharmacophore, but a fourth compulsory site was required during the S_2 site enlargement (*i.e.*, **R2**). Finally, these poses were re-docked with the following settings: **(a)** the receptor grid for PDB code 3G0B

was the same as the one previously used for the protein-ligand docking step; **(b)** the extra precision mode (*i.e.*, XP) was used; **(c)** the *Refine* option was used; and **(d)** the maximum number of poses per ligand was increased to 10.

Declaration of interest

Conflicts of interest: none

Contributions

MJ.OM, G.P, G.W and S.GV designed the research. MJ.OM, A.G and A.CM performed the computational research and wrote the manuscript, A.CM and M.P performed the experimental research. All authors read the final manuscript, suggested changes/improvements and approved the manuscript.

Acknowledgements

This study was supported by research grants 2014PFR-URV-B2-67 and 2015PFR-URV-B2-67 from our University. AG's contract is supported by grant 2015FI_B00655 from the Catalan Government. MP and MM are Serra Húnter research fellows. We thank OpenEye Scientific Software Inc. for generously providing us with an academic license to use their programs. This manuscript has been edited in accordance with the English language usage of our University.

References

[1] S. Jain, M. Jacob, L. Walker, B. Tekwani, Screening North American plant extracts in vitro against *Trypanosoma brucei* for discovery of new antitrypanosomal drug leads, BMC Complement. Altern. Med. 16 (2016) 131. doi:10.1186/s12906-016-1122-0.

[2] M.M. Ghorab, M.S. Alsaid, Novel 4-aminoquinazoline derivatives as new leads for anticancer drug discovery, Acta Pharm. 65 (2015) 299–309. doi:10.1515/acph-2015-0021.

[3] D.G. Brown, T. Lister, T.L. May-Dracka, New natural products as new leads for antibacterial drug discovery, Bioorg. Med. Chem. Lett. 24 (2014) 413–8. doi:10.1016/j.bmcl.2013.12.059.

[4] A.M. Pohlit, R.B.S. Lima, G. Frausin, L.F.R.E. Silva, S.C.P. Lopes, C.B. Moraes, P. Cravo, M.V.G. Lacerda, A.M. Siqueira, L.H. Freitas-Junior, F.T.M. Costa, Amazonian plant natural products: perspectives for discovery of new antimalarial drug leads, Molecules. 18 (2013) 9219–40. doi:10.3390/molecules18089219.

[5] S.J. Martins da Silva, S.G. Brown, K. Sutton, L. V. King, H. Ruso, D.W. Gray, P.G. Wyatt, M.C. Kelly, C.L.R. Barratt, A.G. Hope, Drug discovery for male subfertility using high-throughput screening: a new approach to an unsolved problem, Hum. Reprod. 32 (2017) 974–984. doi:10.1093/humrep/dex055.

[6] J. Alonso-Padilla, A. Rodríguez, High throughput screening for anti-*Trypanosoma cruzi* drug discovery, PLoS Negl. Trop. Dis. 8 (2014) e3259. doi:10.1371/journal.pntd.0003259.

[7] S. Kessel, S. Cribbes, O. Déry, D. Kuksin, E. Sincoff, J. Qiu, L.L.-Y. Chan, High-throughput 3D tumor spheroid screening method for cancer drug discovery using celigo image cytometry, J. Lab. Autom. (2016) 1–12. doi:10.1177/2211068216652846.

[8] F. Annang, G. Pérez-Moreno, R. García-Hernández, C. Cordon-Obras, J. Martín, J.R. Tormo, L. Rodríguez, N. de Pedro, V. Gómez-Pérez, M. Valente, F. Reyes, O. Genilloud, F. Vicente, S. Castanys, L.M. Ruiz-Pérez, M. Navarro, F. Gamarro, D. González-Pacanowska, High-throughput screening platform for natural product-based drug discovery against 3 neglected tropical diseases: human African trypanosomiasis, leishmaniasis, and Chagas disease, J. Biomol. Screen. 20 (2015) 82–91. doi:10.1177/1087057114555846.

[9] S. Chen, Z. Feng, Y. Wang, S. Ma, Z. Hu, P. Yang, Y. Chai, X. Xie, Discovery of novel ligands for TNF-α and TNF receptor-1 through structure-based virtual screening and biological assay,

J. Chem. Inf. Model. 57 (2017) 1101–1111. doi:10.1021/acs.jcim.6b00672.

[10] T.Q. Froes, M.C.C. Melo, G.E.P. Souza, M.S. Castilho, D.M. Soares, Virtual screening and biological evaluation of novel antipyretics compounds, Chem. Biol. Drug Des. 38 (2017) 42–49. doi:10.1111/cbdd.12995.

[11] W. Cui, W. Lv, Y. Qu, R. Ma, Y.-W. Wang, Y.-J. Xu, D. Wu, X. Chen, Discovery of 2-((3-cyanopyridin-2-yl)thio)acetamides as human lactate dehydrogenase A inhibitors to reduce the growth of MG-63 osteosarcoma cells: Virtual screening and biological validation, Bioorg. Med. Chem. Lett. 26 (2016) 3984–7. doi:10.1016/j.bmcl.2016.06.083.

[12] F. Shirgahi Talari, K. Bagherzadeh, S. Golestanian, M. Jarstfer, M. Amanlou, Potent human telomerase inhibitors: Molecular dynamic simulations, multiple pharmacophore-based virtual screening, and biochemical assays, J. Chem. Inf. Model. 55 (2015) 2596–610. doi:10.1021/acs.jcim.5b00336.

[13] V. Mendes, T.L. Blundell, Targeting tuberculosis using structure-guided fragment-based drug design, Drug Discov. Today. 22 (2017) 546–554. doi:10.1016/j.drudis.2016.10.003.

[14] F. Benmansour, I. Trist, B. Coutard, E. Decroly, G. Querat, A. Brancale, K. Barral, Discovery of novel dengue virus NS5 methyltransferase non-nucleoside inhibitors by fragment-based drug design, Eur. J. Med. Chem. 125 (2017) 865–880. doi:10.1016/j.ejmech.2016.10.007.

[15] B.C. Doak, R.S. Norton, M.J. Scanlon, The ways and means of fragment-based drug design, Pharmacol. Ther. 167 (2016) 28–37. doi:10.1016/j.pharmthera.2016.07.003.

[16] M. Tuyishime, R. Lawrence, S. Cocklin, Core chemotype diversification in the HIV-1 entry inhibitor class using field-based bioisosteric replacement, Bioorg. Med. Chem. Lett. 26 (2016) 228–34. doi:10.1016/j.bmcl.2015.10.080.

[17] N. Chandna, S. Kumar, P. Kaushik, D. Kaushik, S.K. Roy, G.K. Gupta, S.M. Jachak, J.K. Kapoor, P.K. Sharma, Synthesis of novel celecoxib analogues by bioisosteric replacement of sulfonamide as potent anti-inflammatory agents and cyclooxygenase inhibitors, Bioorg. Med. Chem. 21 (2013) 4581–90. doi:10.1016/j.bmc.2013.05.029.

[18] Z. Jiang, Y. Wang, W. Wang, S. Wang, B. Xu, G. Fan, G. Dong, Y. Liu, J. Yao, Z. Miao, W. Zhang, C. Sheng, Discovery of highly potent triazole antifungal derivatives by heterocycle-benzene bioisosteric replacement, Eur. J. Med. Chem. 64 (2013) 16–22. doi:10.1016/j.ejmech.2013.04.025.

[19] T. Sterling, J.J. Irwin, ZINC 15 – ligand discovery for everyone, J. Chem. Inf. Model. 55 (2015) 2324–37. doi:10.1021/acs.jcim.5b00559.

[20] R.N. Kushwaha, W. Haq, S.B. Katti, Sixteen-years of clinically relevant dipeptidyl peptidase-IV (DPP-IV) inhibitors for treatment of type-2 diabetes: a perspective, Curr. Med. Chem. 21 (2014) 4013–45. doi:10.2174/09298673216661409151433 09.

[21] S.M. Sheehan, H.-J. Mest, B.M. Watson, V.J. Klimkowski, D.E. Timm, A. Cauvin, S.H. Parsons, Q. Shi, E.J. Canada, M.R. Wiley, G. Ruehter, B. Evers, S. Petersen, L.C. Blaszczak, S.R. Pulley, B.J. Margolis, G.N. Wishart, B. Renson, D. Hankotius, M. Mohr, J.-C. Zechel, J. Michael Kalbfleisch, E. a Dingess-Hammond, A. Boelke, A.G. Weichert, Discovery of non-covalent dipeptidyl peptidase IV inhibitors which induce a conformational change in the active site, Bioorg. Med. Chem. Lett. 17 (2007) 1765–8. doi:10.1016/j.bmcl.2006.12.074.

[22] B. Patel, M. Ghate, Computational studies on structurally diverse dipeptidyl peptidase IV inhibitors: An approach for new antidiabetic drug development, Med. Chem. Res. 22 (2013) 4505–4521. doi:10.1007/s00044-012-0455-6.

[23] H. Nojima, K. Kanou, G. Terashi, M. Takeda-Shitaka, G. Inoue, K. Atsuda, C. Itoh, C. Iguchi, H. Matsubara, Comprehensive analysis of the Co-structures of dipeptidyl peptidase IV and its inhibitor, BMC Struct. Biol. 16 (2016) 11. doi:10.1186/s12900-016-0062-8.

[24] A. Smelcerovic, F. Miljkovic, A. Kolarevic, J. Lazarevic, A. Djordjevic, G. Kocic, M. Anderluh, An overview of recent dipeptidyl peptidase-IV inhibitors: Linking their structure and physico-chemical properties with SAR, pharmacokinetics and toxicity, Curr. Top. Med. Chem. 15 (2015) 2342–72.

[25] Y. Liu, Y. Hu, Novel DPP-4 inhibitors against diabetes, Future Med. Chem. 6 (2014) 793–808. doi:10.4155/fmc.14.39.

[26] EON 2.2.0.5: OpenEye Scientific Software, Santa Fe, NM. http://www.eyesopen.com.

[27] J.J. Irwin, T. Sterling, M.M. Mysinger, E.S. Bolstad, R.G. Coleman, ZINC: a free tool to discover chemistry for biology, J. Chem. Inf. Model. 52 (2012) 1757–68. doi:10.1021/ci3001277.

[28] Reaxys Medicinal Chemistry. www.reaxys.com.

[29] RDKit, Open-Source Cheminformatics. http://www.rdkit.org.

[30] Z. Zhang, M.B. Wallace, J. Feng, J. a Stafford, R.J. Skene, L. Shi, B. Lee, K. Aertgeerts, A. Jennings, R. Xu, D.B. Kassel, S.W. Kaldor, M. Navre, D.R. Webb, S.L. Gwaltney, Design and synthesis of pyrimidinone and pyrimidinedione inhibitors of dipeptidyl peptidase IV, J. Med. Chem. 54 (2011) 510–24. doi:10.1021/jm101016w.

[31] M.J. Ojeda,A. Cereto-Massagué, C. Valls, G. Pujadas, DPP-IV, an important target for antidiabetic functional food design, in: Foodinformatics, applications of chemical information to food chemistry, Springer (2014) 177-212. doi:10.1007/978-3-319-10226-9.

[32] B. Kuhn, M. Hennig, P. Mattei, Molecular recognition of ligands in dipeptidyl peptidase IV, Curr. Top. Med. Chem. 7 (2007) 609–19.

[33] G. Scapin, Structural biology and molecular modeling in the design of DPP4 inhibitors, in: Multifaceted roles of crystallography in modern drug discovery, Springer (2015) 53–67. doi:10.1007/978-94-017-9719-1_5.

[34] J.R. Bjelke, J. Christensen, S. Branner, N. Wagtmann, C. Olsen, A.B. Kanstrup, H.B. Rasmussen, Tyrosine 547 constitutes an essential part of the catalytic mechanism of dipeptidyl peptidase IV, J. Biol. Chem. 279 (2004) 34691–7. doi:10.1074/jbc.M405400200.

[35] Y. Liu, Y. Hu, T. Liu, Recent advances in non-peptidomimetic dipeptidyl peptidase 4 inhibitors: medicinal chemistry and preclinical aspects, Curr. Med. Chem. 19 (2012) 3982–99.

[36] B.D. Patel, M.D. Ghate, Recent approaches to medicinal chemistry and therapeutic potential of dipeptidyl peptidase-4 (DPP-4) inhibitors, Eur. J. Med. Chem. 74 (2014) 574–605. doi:10.1016/j.ejmech.2013.12.038.

[37] L. Guasch, M.J. Ojeda, N. González-Abuín, E. Sala, A. Cereto-Massagué, M. Mulero, C. Valls, M. Pinent, A. Ardévol, S. Garcia-Vallvé, G. Pujadas, Identification of novel human dipeptidyl peptidase-IV inhibitors of natural origin (part I): virtual screening and activity assays, PLoS One. 7 (2012) e44971. doi:10.1371/journal.pone.0044971.

[38] C. Rummey, G. Metz, Homology models of dipeptidyl peptidases 8 and 9 with a focus on loop predictions near the active site, Proteins. 66 (2007) 160–71. doi:10.1002/prot.21138.

[39] Schrödinger - Knowledge Base. https://www.schrodinger.com/kb/1012.

[40] OMEGA 2.5.1.4: OpenEye Scientific Software, Santa Fe, NM. http://www.eyesopen.com.

[41] A. Cereto-Massagué, L. Guasch, C. Valls, M. Mulero, G. Pujadas, S. Garcia-Vallvé, DecoyFinder: an easy-to-use python GUI application for building target-specific decoy sets, Bioinformatics. 28 (2012) 1661–1662. doi:10.1093/bioinformatics/bts249.

[42] QikProp, version 4.5, Schrödinger, LLC, New York, NY, (2015).

[43] C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, Adv. Drug Deliv. Rev. 46 (2001) 3–26.

[44] J. Kelder, P.D. Grootenhuis, D.M. Bayada, L.P. Delbressine, J.P. Ploemen, Polar molecular surface as a dominating determinant for oral absorption and brain penetration of drugs, Pharm. Res. 16 (1999) 1514–9. doi:10.1023/A:1015040217741.

[45] R. Nilakantan, N. Bauman, J.S. Dixon, R. Venkataraghavan, Topological torsion: A new molecular descriptor for SAR applications. Comparison with other descriptors, J. Chem. Inf. Model. 27 (1987) 82–85. doi:10.1021/ci00054a008.

[46] LigPrep, version 3.5, Schrödinger, LLC, New York, NY, (2015).

[47] Schrödinger Suite 2015-3 Protein Preparation Wizard; Epik; Impact; Prime, Schrödinger, LLC, New York, NY, (2015).

[48] D. Kim, L. Wang, M. Beconi, G.J. Eiermann, M.H. Fisher, H. He, G.J. Hickey, J.E. Kowalchick, B. Leiting, K. Lyons, F. Marsilio, M.E. McCann, R.A. Patel, A. Petrov, G. Scapin, S.B. Patel, R.S. Roy, J.K. Wu, M.J. Wyvratt, B.B. Zhang, L. Zhu, N.A. Thornberry, A.E. Weber, (2*R*)-4-oxo-4-[3-(trifluoromethyl)-5,6-dihydro[1,2,4]triazolo[4,3-a]pyrazin-7(8*H*)-yl]-1-(2,4,5-triflu orophenyl)butan-2-amine: a potent, orally active dipeptidyl peptidase IV inhibitor for the treatment of type 2 diabetes, J. Med. Chem. 48 (2005) 141–51. doi:10.1021/jm0493156.

[49] Glide, version 6.8, Schrödinger, LLC, New York, NY, (2015).

[50] Phase, version 4.4, Schrödinger, LLC, New York, NY, (2015).

[51] CombiGlide, version 3.9, Schrödinger, LLC, New York, NY, (2015).

[52] CombiGlide Diverse Side-chain Collection, v. 1.2. https://www.schrodinger.com/combiglide.

Supplementary Material

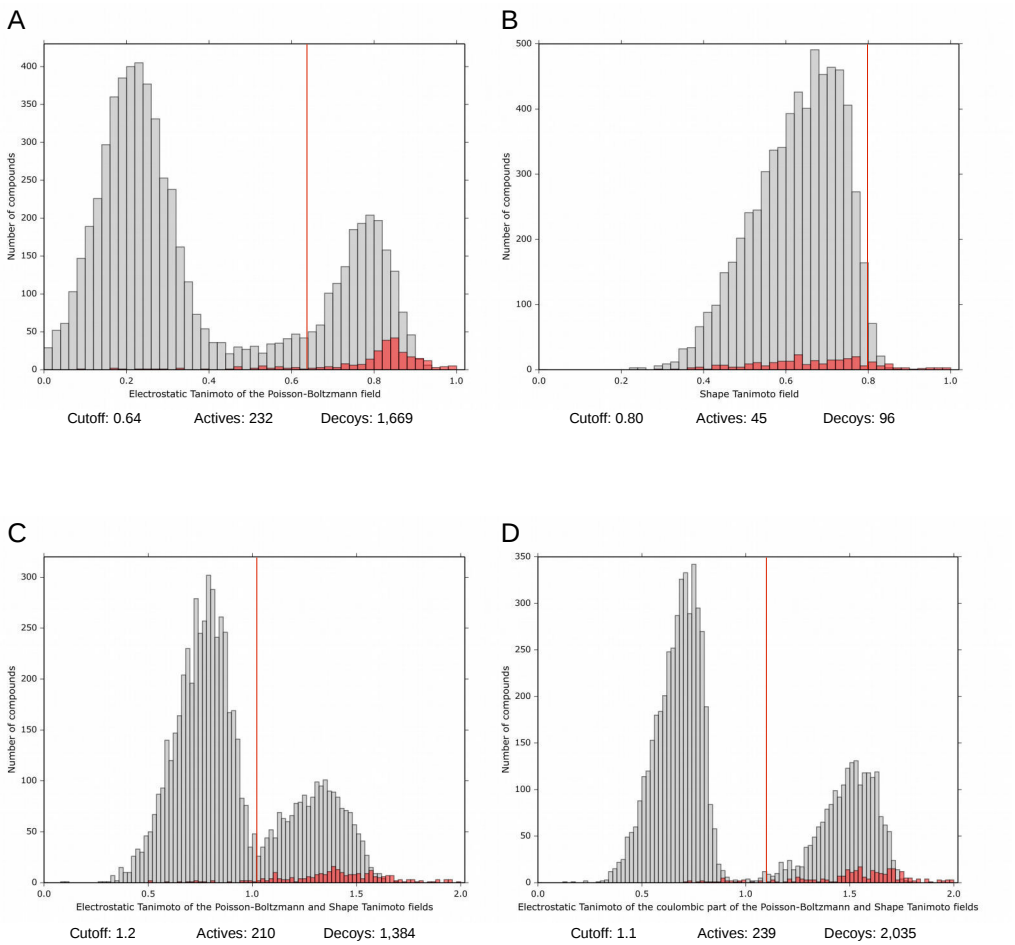


Figure S1. Histograms showing the distribution of the highest Tanimoto values for actives (shown in red) and decoys (shown in gray) for **(A)** the Electrostatic Tanimoto of the Poisson-Boltzmann field; **(B)** the Shape Tanimoto field; **(C)** the Electrostatic Tanimoto of the Poisson-Boltzmann and Shape Tanimoto fields and **(D)** the Electrostatic Tanimoto of the coulombic part of the Poisson-Boltzmann and Shape Tanimoto fields. Different cutoffs (red line) were applied to the set of actives and decoys by using these EON parameters in order to increase the enrichment factor of the VS validation.

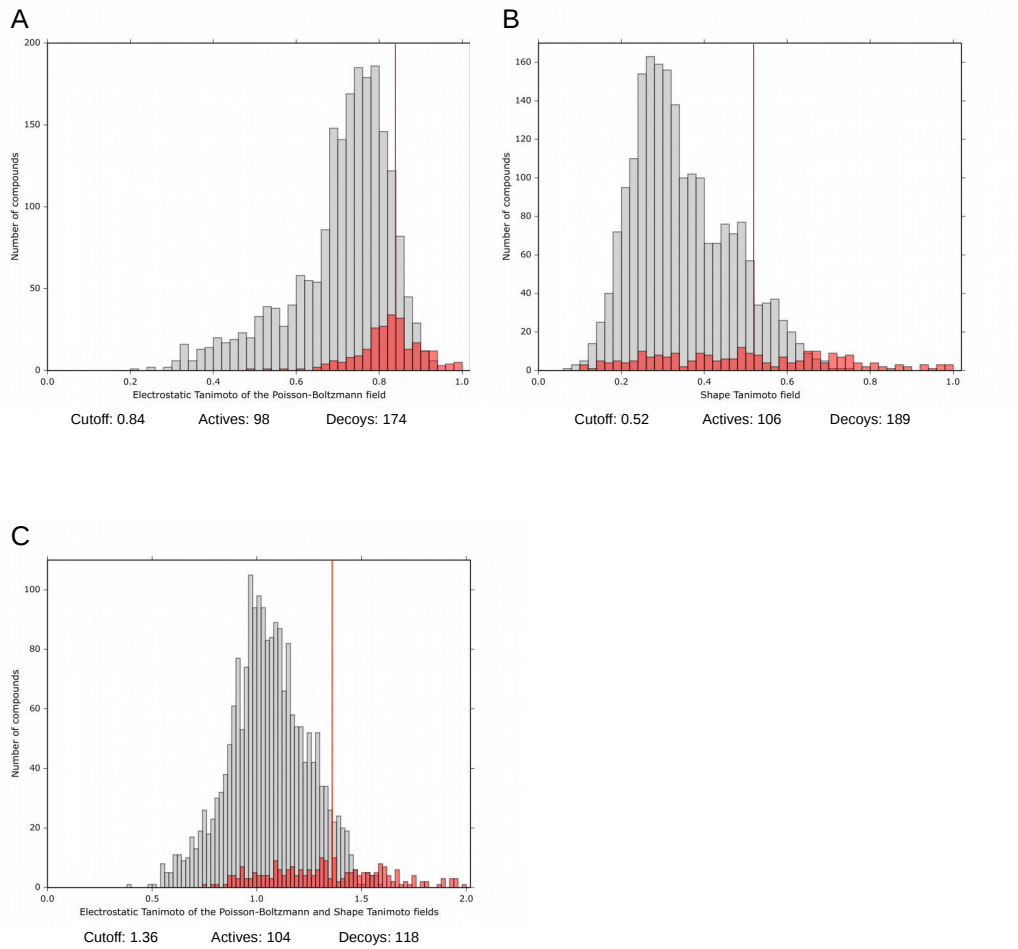


Figure S2. Histograms showing the distribution of the highest Tanimoto values for the pose of actives (shown in red) and decoys (shown in gray) that are above the first cutoff of 0.7 for the Electrostatic Tanimoto of the coulombic part of the Poisson-Boltzmann (see Figure 3A). The panels represent **(A)** the Electrostatic Tanimoto of the Poisson-Boltzmann field; **(B)** the Shape Tanimoto field and **(C)** the Electrostatic Tanimoto of the Poisson-Boltzmann and Shape Tanimoto fields. Different cutoffs (red line) were applied for the set of actives and decoys by using these EON parameters in order to increase the enrichment factor of the VS validation.

Manuscript 3

Ephedrine as a lead compound for the development of new DPP-IV inhibitors

María José Ojeda-Montes[a,1], Andrea Ardid-Ruiz[a,1], Sarah Tomás-Hernández[a], Aleix Gimeno[a], Adrià Cereto-Massagué[a], Raúl Beltrán-Debón[a], Miquel Mulero[a], Santiago Garcia-Vallvé[a,b], Gerard Pujadas[a,b,*] and Cristina Valls[a]

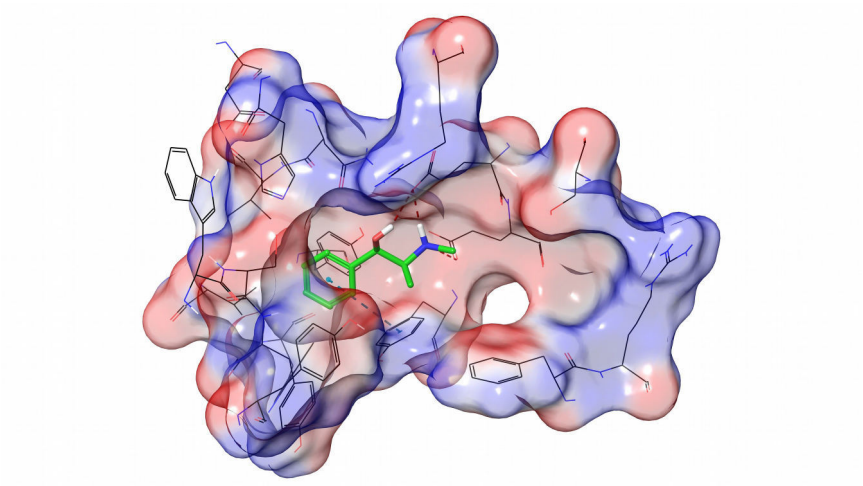
[a]Research group in Cheminformatics & Nutrition, Departament de Bioquímica i Biotecnologia, Universitat Rovira i Virgili, Tarragona, Catalonia, Spain

[b]EURECAT, TECNIO, CEICS, Reus, Catalonia, Spain

**Correspondence to:*

Gerard Pujadas, Research group in Cheminformatics & Nutrition, phone: +34 977 55 95 65, fax: +34 977 55 82 32. Departament de Bioquímica i Biotecnologia, Facultat de Química, Universitat Rovira i Virgili, C/ Marcel·lí Domingo 1, Edifici N4, 43007 Tarragona, Catalonia, Spain. E-mail: gerard.pujadas@urv.cat

Abstract



Background: Extracts from *Ephedra* species have been used in traditional medicine to treat hyperglycemia. A previous *in silico* study predicted that ephedrine and five ephedrine derivatives could contribute to the described antidiabetic effect of *Ephedra* extracts by inhibiting dipeptidyl peptidase IV (DPP-IV). Finding selective DPP-IV inhibitors is a current therapeutic strategy for type 2 diabetes mellitus management. Therefore, the main aim of this work is to experimentally determine whether these alkaloids are DPP-IV inhibitors.

Results: Our results show that all six molecules are DPP-IV inhibitors, with IC_{50} ranging from 124 μM for ephedrine to 28.890 mM for N-methylpseudoephedrine.

Conclusions: Further computational analysis shows how *Ephedra*'s alkaloids are promising lead molecules for developing more potent and selective DPP-IV inhibitors.

1. Introduction

Nature has been an essential resource in folk medicine for the treatment of a wide variety of diseases. Moreover, in recent decades, medicinal chemists have used the complexity and molecular diversity of natural products to discover and develop new drugs [1,2]. Of all small-molecule approved drugs between 1981 and 2014, 33% are either natural products or have been derived from them [3]. Plants in particular have played an important role in healthcare for different cultures. For instance, about 800 medicinal plants have been reported to show hypoglycemic properties [2,4].

Plants from the *Ephedra* family, whose traditional preparation is called *ma huang*, have been used in Chinese medicine for over 5000 years as a circulatory stimulant, a diaphoretic and an antipyretic remedy [5]. Indeed, *Ephedra* was considered a spiritual drink for rejuvenation, immortality and resurrection in ancient Aryan culture [6]. *Ephedra* species contain alkaloids of biological relevance such as ephedrine, pseudoephedrine, norephedrine, norpseudoephedrine, N-methylephedrine and N-methylpseudoephedrine [7,8]; all which have a similar molecular structure. These alkaloids are adrenergic agonists that stimulate the release of endogenous catecholamines (see Table S1) [9–11]. For this reason, the *Ephedra* herb has been commonly used for increasing blood pressure and stimulating the central nervous system [9,10]. Likewise, it has benefits for the treatment of asthma, coughs, colds and mild forms of bronchospasm [11–13]. Unfortunately, consumption of this plant has been associated with various adverse effects including palpitations, hypertension, insomnia, anxiety and dysuria [10–13]. Because of these adverse cardiovascular effects and the fact that it is a major source of methamphetamine [5,8], *Ephedra* extracts have been banned in food supplements in several European countries [14] and by the American Food & Drug Administration (FDA) [5,10].

Extracts from some *Ephedra* species (*i.e.*, *Ephedra distachya* and *Ephedra alata*) have been also reported to be effective as antidiabetics [4,15,16] because they regenerate atrophied pancreatic islets and restore insulin secretion [4,17]. Interestingly, a recent computational study has predicted that ephedrine and other derivatives also present in different *Ephedra* species can inhibit the dipeptidyl peptidase-IV enzyme (DPP-IV) [18]. DPP-IV is a ubiquitous

aminopeptidase that selectively removes N-terminal dipeptides from peptides with proline or alanine in the second position [19,20]. DPP-IV is involved in glycemia, cardiovascular function and inflammation through the proteolytic inactivation of a wide variety of biological regulatory peptides, including incretin hormones [21]. Incretins, secreted as a response to nutrient ingestion, play a key role in the regulation of glucose homeostasis and pancreatic islet functions [22]. Therefore, a good mechanism for treating or preventing type 2 diabetes mellitus (T2DM) could consist of inhibiting DPP-IV and thus increasing the half-lives of incretin hormones [19,22]. Consequently, DPP-IV inhibitors are of considerable interest to the pharmaceutical industry [20,23,24], and intensive research in this area has resulted in the recent launch of several gliptins onto the market for T2DM treatment (see Table S2). Interestingly, DPP-IV inhibitors can also inhibit the apoptosis of β-cells and thus help them to regenerate and differentiate, which would have the same positive effect on islet pancreatic regeneration [25,26] that has also been described for *Ephedra* extracts [17]. The selectivity of the DPP-IV inhibitors is also important for establishing an optimal safety profile for this antihyperglycemic treatment. Thus, the inhibition of DPP8 and DPP9 by DPP-IV inhibitors has been associated with multi-organ toxicities in preclinical species (*e.g.* alopecia, thrombocytopenia, reticulocytopenia, enlarged spleen, multi-organ histopathological changes, mortality in rats and gastrointestinal toxicity in dogs) [27].

Therefore, the main goals of the present study are **(1)** to demonstrate that, at least partly, the described antidiabetic effect of different *Ephedra* species extracts is the result of the DPP-IV inhibitory bioactivity of ephedrine and the ephedrine-derivatives found in these extracts; and **(2)** to suggest how one of these alkaloids can be used as lead compound for developing new DPP-IV inhibitors with better potency and selectivity.

2. Experimental

2. 1. Chemicals and biochemicals

The compounds (1R,2S)-(-)-ephedrine (product number 134910), (1S,2S)-(+)-pseudoephedrine (product number 287636), (1S,2R)-(+)-norephedrine (product number 317500), (1R,2R)-(-)-norpseudoephedrine (product number 670561), (1R,2S)-(-)-N-methylephedrine (product number 235210) and

(1S,2S)-(+)-N-methylpseudoephedrine (product number 290041) (purity ≥ 98%) were purchased from Sigma Aldrich (St Louis, MO). The DPP-IV Drug Discovery Kit-AK499 was purchased from Enzo Life Sciences International, Inc. (Farmingdale, NY) whereas the fluorogenic DPP8 (catalog number 80208) and DPP9 (catalog number 80209) assay kits were bought from BPS Bioscience, Inc. (San Diego, CA). Vildagliptin (purity ≥ 98%) was purchased from Cayman Chemical Company (Ann Arbor, MI).

2.2. Pure compound preparation

The first stock solution for pseudoephedrine and norpseudoephedrine was made with 99.9% ethanol, N-methylephedrine and N-methylpseudoephedrine were prepared with 99.9% methanol and ephedrine and norephedrine were prepared with 1M HCl solution. The rest of the dilution banks were made with 50 mM Trizma hydrochloride at pH 7.4.

2.3. In vitro assay of inhibition of DPP-IV by ephedrine and ephedrine-derivatives

The DPP-IV drug discovery Kit-AK499 was used to measure the DPP-IV inhibitory activity of ephedrine and ephedrine-derivatives. The DPP-IV activity was monitored fluorimetrically by using the fluorogenic substrate (*i.e.*, H-Gly-Pro-AMC) provided by the kit. This fluorimetric assay is based on the cleavage of the 7-amino-4-methylcoumarin moiety (*i.e.*, AMC) from the C-terminus of the peptide substrate, which intensifies its fluorescence. The procedure was performed in a 96-well microplate in 100 µL of reaction volume. After the addition of 10 µL of each compound tested to the human recombinant DPP-IV enzyme, the microplate was incubated for 10 minutes at room temperature to allow compound/enzyme interaction. Next, the reaction was started by the addition of 50 µL (diluted 50-fold from stock) of the fluorimetric substrate H-Gly-Pro-AMC. Finally, sample triplicates were read each minute for 40 minutes in a Biotek FLx800 Fluorescence Microplate Reader at Ex:380nm/Em:460nm. Four different conditions were used: **(1)** the blank (*i.e.*, no DPP-IV present); **(2)** the negative control (*i.e.*, no inhibitor present); **(3)** the positive control (*i.e.*, the P32/98 inhibitor diluted 10-fold from stock provided by the kit); and **(4)** the ephedrine and ephedrine-derivatives at different concentrations (see Figure S1). Data was plotted as time versus Relative Fluorescence Units, and only the slope where the reaction was linear was used to obtain the activity of the enzyme. The remaining enzyme activity for DPP-IV in the presence of each

tested compound was calculated as a percentage of activity relative to negative control wells.

2.4. In vitro assay of inhibition of DPP8 and DPP9 by ephedrine and ephedrine-derivatives

Similar fluorogenic assays were performed to measure DPP8 and DPP9 activity under the effect of ephedrine and ephedrine-derivatives. The procedure was performed in a 96-well black microplate in 100 µL of reaction volume. Briefly, after the addition of 10 µL of each compound tested to the human recombinant DPP8 and DPP9 enzyme, the microplate was incubated for 10 minutes at room temperature to allow compound/enzyme interaction. Next, the reaction was started by the addition of 20 µL (diluted 20-fold from stock) of the fluorigenic substrate. Finally, sample triplicates were read in a Biotek FLx800 Fluorescence Microplate Reader at Ex:380nm/Em:460nm. Four different conditions were used: **(1)** the blank (*i.e.*, no DPP8/DPP9 present); **(2)** the negative control (*i.e.*, no inhibitor present); **(3)** the positive control [*i.e.*, vildagliptin (with IC_{50} = 900 and 680 nM for DPP8 and DPP9, respectively) at 2.5 nM]; and **(4)** the ephedrine and ephedrine-derivatives at concentrations equal to their IC_{50} for DPP-IV. The remaining enzyme activity for DPP8 or DPP9 in the presence of each tested compound was calculated as a percentage of activity relative to negative control wells.

2.5. Statistical analysis and IC_{50} calculation

Statistical calculations were carried out with XLSTAT v2016.7 for Windows. Results were expressed as the mean ± SD of three replicates. The DPP-IV IC_{50} was obtained through non-linear regression (curve fit) followed by sigmoidal dose-response curve (variable slope). One-Way ANOVA followed by Bonferroni's Multiple Comparison Test were applied to observe significant differences ($p < 0.05$).

2.6. Target structure setup for protein-ligand docking

The protein structure used during the protein-ligand docking studies has the Protein Data Bank (PDB) code 1X70. Default options with VHELIBS were used to evaluate the correctness of the fit between the 1X70 binding site coordinates and the electron density map [28]. According to the results of this analysis, the binding site coordinates for this

structure are reliable and, therefore, suitable for use as the target during protein-ligand docking.

In order to set up 1X70 for use as a target during the protein-ligand docking with Glide v6.9 (Schrödinger, LLC, New York, NY, 2015), it was prepared with the *Protein Preparation Wizard* panel (PPW; Protein Preparation Wizard 2015-4, Schrödinger, LLC, New York, NY, 2015) available at the Schrödinger 2015-4 release. During the PPW's pre-process step all options were set to default with the exception of *remove original hydrogens*, *fill in missing side chains* and *cap termini* options that were set to on. During the PPW's *review and modify* step, chain B was removed and states for the 1X70's ligand were generated with default parameters. Then the S2 state for the ligand was chosen because it has a lower state penalty (*i.e.*, 0.83 Kcal/mol) and a higher H-bound count (*i.e.*, 2) compared with those of the original ligand state (*i.e.*, 4.00 Kcal/mol and 1, respectively). Finally, during the PPW's *refine* step: **(a)** hydrogen bond assignment was performed with default options with the *automatically optimize* tool; **(b)** waters with less than 3 hydrogen bonds to non-waters were removed; and **(c)** the resulting structure was submitted to a restrained minimization with default options.

The resulting 1X70 structure was then used to obtain the grid that it is needed for identifying the DPP-IV binding site most suitable for further docking ephedrine and ephedrine-derivatives. The receptor grid was obtained around the 1X70's ligand (*i.e.*, sitagliptin) by using the *Schrödinger's Grid Generation* panel with default options.

2.7. Ephedrine and ephedrine-derivatives setup for protein-ligand docking

In order to obtain their 3D structures, ephedrine and ephedrine-derivatives were downloaded from Reaxys Medicinal Chemistry (http://www.reaxys.com). Then, all six molecules were processed with LigPrep v3.6 (Schrödinger, LLC, New York, NY, 2015) as follow: **(a)** the force field OPLS 2005 was used; **(b)** all possible ionization and tautomerization states at pH 7.0 ± 2.0 were generated with Epik; **(c)** the desalt option was activated; **(d)** chiralities were respected from input geometry when generating stereoisomers; and **(e)** one low-energy ring conformation per ligand was generated.

2.8. Structure-based pharmacophore generation for DPP-IV inhibition

The structure-based pharmacophore used in the present study (see Figure S2A) is a modified version of a previous one developed by our group (see Figure S2B) [29]. Thus, Figure S2 shows that: **(a)** both pharmacophores share two compulsory sites (*i.e.*, **P/D** and **H/R1**) showing interactions with the Glu205/Glu206 dyad (*i.e.*, **P/D**) and with the S_1 pocket (*i.e.*, **H/R1**); **(b)** the optional sites **H/R2** and **H/R4** are related with sites **R2** and **R3** in the new pharmacophore (where only aromatic rings have been required to allow π-π stacking interaction with Phe357 and Tyr547, respectively); and **(c)** the rest of the sites in the old pharmacophore are not shared either because of its preference to form covalent bonds (*i.e.*, **H/R3**) or because of their low energy contribution to the protein-ligand interaction energy (*i.e.*, **A1** and **A2**). The tolerances for the different sites were also readjusted using as reference more co-crystallized DPP-IV inhibitor complexes than in the original pharmacophore. The associated tolerances were thus 2.3Å for **P/D** (instead of 1.8Å), 2.0Å for **H/R1** (like in the original one), 2.5Å for **R2** (instead of 3.3Å for **H/R2**) and 1.8Å for **R3** (instead of 2.0Å for **H/R4**).

2.9. Protein-ligand docking and structure-based pharmacophore screening

Docking studies for the six assayed alkaloids were carried out with Glide v6.9 as follows: **(a)** the extra precision mode (*i.e.*, XP) was used; **(b)** the aromatic hydrogens were considered as donors; **(c)** the maximum number of poses per ligand was increased to 32; and **(d)** the number of poses per ligand to include in the post-docking minimization was increased to 320.

Next, the docked poses of ephedrine and ephedrine-derivatives were filtered through the structure-based pharmacophore by using Phase v4.5 (Schrödinger, LLC, New York, NY, 2015) with the *score in place* option set to on (*i.e.*, no re-orientation of the docked poses was allowed during the search). Only those docked poses that matched the two compulsory sites of the structure-based pharmacophore were kept.

2.10. Lead-optimization from the (1S,2R)-(+)-norephedrine

Lead optimization was performed with CombiGlide v3.9 (Schrödinger, LLC, New York, NY, 2015) by using the *Virtual Combinatorial Screening* workflow. The core-containing molecule was the lowest-energy docked pose for norephedrine and the substituents were obtained

from the Schrödinger CombiGlide Diverse Side-chain Collection v1.2 (https://www.schrodinger.com/combiglide) (which contains all reasonable ionization and tautomeric states for a collection of 817 representative functional groups commonly found in pharmaceuticals, with linkers of variable lengths).

After defining the point where the substituents were attached, we established the following parameters for a single-position docking run: **(a)** the receptor grid was the same as during the previous protein-ligand docking step; **(b)** the *apply core constraints* option was used within a maximum RMSD of 1.0Å; **(c)** the *single-position and combinatorial* option was selected; and **(d)** the CombiGlide XP docking mode was used. After this process, the resulting norephedrine derivatives were filtered according to their ADME properties using *Druglike* filter set. Only two violations of the following criteria were allowed for these derivatives to be considered: **(a)** molecular weight less than 500 Da; **(b)** maximum 5 hydrogen bond donors; **(c)** maximum 10 hydrogen bond acceptors; **(d)** predicted octanol/water partition coefficient (*i.e.,* logP) less than 5; **(e)** 10 or less rotatable bonds; and **(f)** 150 $Å^2$ or less Van der Waals surface area. Then, the resulting derivatives were refined with Glide XP by using default parameters. Finally, the resulting derivatives were filtered again with the same structure-based pharmacophore and conditions as above but with the exception that a third compulsory site was now required (*i.e.*, **R2**). The same process was followed again with the aim of further optimizing the derivatives (this time, by using two different norephedrine derivatives as the core-containing molecule where two significantly different substituents were included during the first optimization step). All previous steps for CombiGlide were performed following the same parameters and selection criteria as the first optimization step with the exception of the structure-based pharmacophore, in which a fourth compulsory site was now required (*i.e.*, **R3**).

3. Results and Discussion

3.1. DPP-IV inhibitory bioactivity for selected compounds

Previous *in silico* work by our research group allowed us to predict which molecules in natural extracts with a known antidiabetic effect were DPP-IV inhibitors [18]. Two of these molecules, ephedrine and pseudoephedrine, have been isolated from two *Ephedra* species

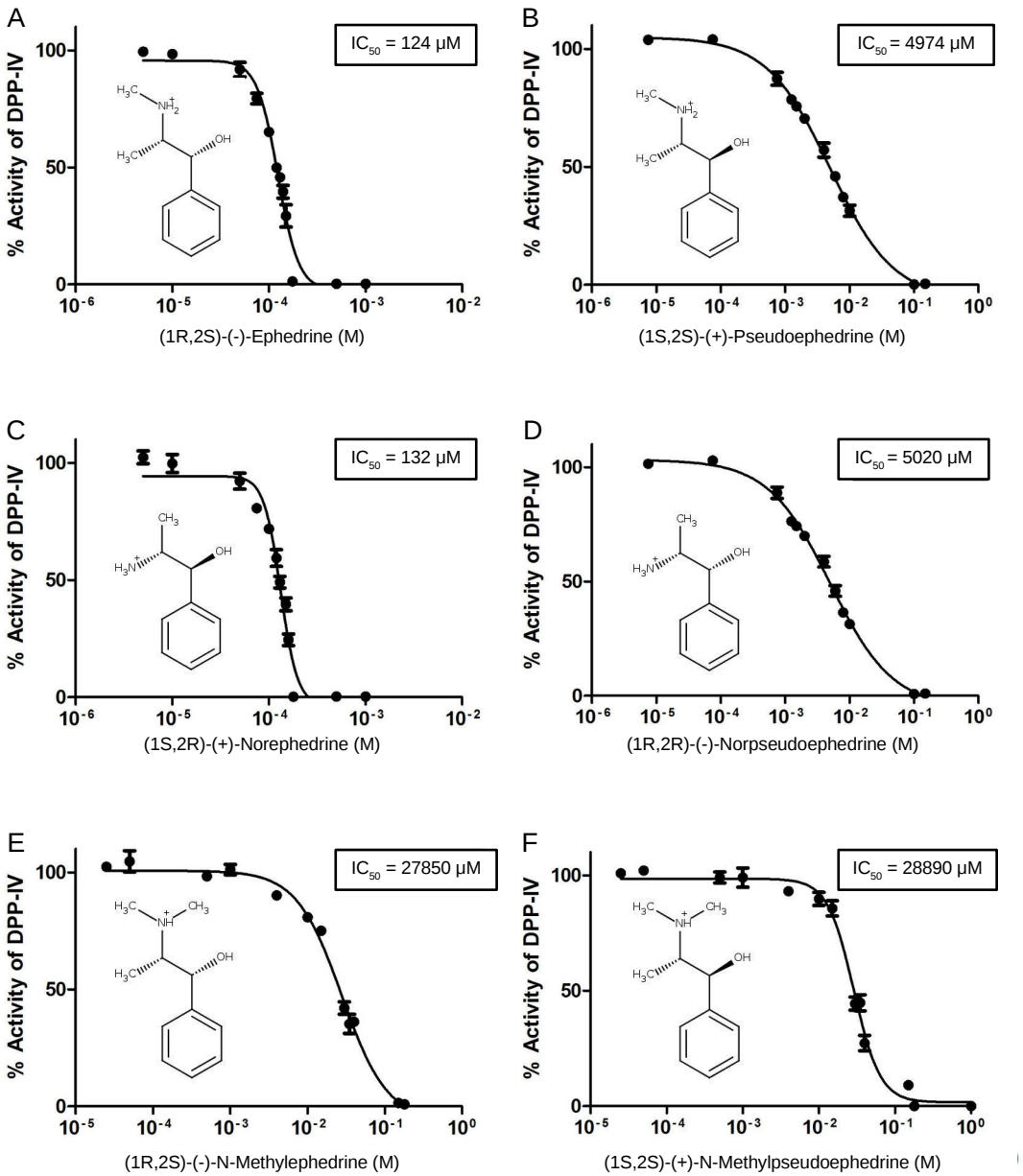


Figure 1. Percentage of DPP-IV activity relative to the concentration (in M) for each assayed compound. DPP-IV inhibitory dose-response curve was obtained via a competitive binding assay. The data represent the mean ± SD of an experiment performed in triplicate. A non-linear regression (curve fit) followed by sigmoidal dose-response curve (variable slope) was made to find the IC_{50}. Each panel includes also the 2D structure of the ionization state of the corresponding molecule at pH 7.

(*i.e.*, *Ephedra distachya* and *Ephedra alata)* with described antidiabetic bioactivity [15,16]. Another four ephedrine alkaloids (*i.e.*, norephedrine, norpseudoephedrine, N-methylephedrine and N-methylpseudoephedrine) were also predicted to be DPP-IV inhibitors [18] and, although no antidiabetic activity had been reported before for the natural sources from which they were isolated, they have been identified in species that share the same genus as plants with known antidiabetic properties [8]. For this reason, the six molecules were experimentally assayed to determine their bioactivity on DPP-IV and the results are shown in Figure 1 and Figure S1.

The bioactivity assay shows that all six molecules are able to inhibit DPP-IV at relatively high concentrations (*i.e.*, IC_{50} values in the 124 μM to 28.890 mM range) with a dose-response behavior (see Figure S1). Therefore, all six molecules are significantly less active as DPP-IV inhibitors than commercially available ones (*e.g.*, the IC_{50} for vildagliptin, sitagliptin, alogliptin and saxagliptin is in the 3.5 to 18 nM range). Unfortunately, these six alkaloids also inhibit DPP8 and DPP9 (see Table 1) at a concentration equivalent to its IC_{50} for DPP-IV. This absence of selectivity could be because of the similar pattern of intermolecular interactions shared by these six compounds with conserved residues in all the three enzymes (*e.g.*, the Glu205/Glu206 dyad [30] and the hydrophobic residues of the S_1 pocket [31]; see Figure 2). These compounds are therefore expected to present the secondary effects of the DPP8 and DPP9 inhibitors [27].

Table 1. Bioactivity data of the *Ephedra* alkaloids on DPP-IV, DPP8 and DPP9.IC_{50} has been calculated for DPP-IV, whereas for DPP8 and DPP9 the value corresponds to the percentage of inhibition using a concentration of each compound equivalent to its IC_{50} for DPP-IV.

***Ephedra* alkaloid**	**DPP-IV IC_{50} (μM)**	**% of inhibition for DPP8 (% ± SD)**	**% of inhibition for DPP9 (% ± SD)**
(1R,2S)-(-)-Ephedrine	124	75.87 ± 1.57	90.19 ± 0.39
(1S,2S)-(+)-Pseudoephedrine	4974	30.30 ± 1.68	52.21 ± 3.41
(1S,2R)-(+)-Norephedrine	132	74.53 ± 2.76	90.30 ± 0.53
(1R,2R)-(-)-Norpseudoephedrine	5020	46.71 ± 3.27	56.93 ± 5.30
(1R,2S)-(-)-N-Methylephedrine	27850	69.49 ± 1.62	86.61 ± 2.59
(1S,2S)-(+)-N-Methylpseudoephedrine	28890	73.84 ± 0.35	88.92 ± 0.68

Figure 2. The best docked poses (with the corresponding XP GScore) for the six studied *Ephedra* alkaloids at the DPP-IV binding site. Blue dashed lines show π-π stacking intermolecular interactions whereas the red ones show either salt bridges (between the positively charged amine and the Glu dyad) or hydrogen bonds. All panels have the same orientation to allow easy comparison between them.

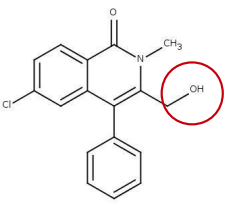
3.2. Structure-activity relationship as DPP-IV inhibitors

Figure 2 shows protein-ligand docked poses for the six assayed alkaloids and suggests how they bind to the binding site. Overall, due to the low molecular weight of the tested alkaloids (at the 151-179 Da range), only a reduced number of interactions with the DPP-IV binding site are exhibited. These are **(1)** either one or two salt bridges (depending on the alkaloid) between the ligand amino group and the Glu205/Glu206 side chains (*i.e.*, **P/D** site of the pharmacophore; see Figure S2A); **(2)** a hydrophobic interaction between the phenyl ring and the S_1 pocket (*i.e.*, **H/R1** site of the pharmacophore; see Figure S2A); and **(3)** hydrogen bonds with either Glu205, Tyr662 or Asn710.

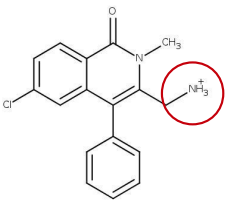
The Glu dyad has been highly conserved throughout the DPP-IV protein family for different species [30,32] and the interactions with these two residues have been shown to be essential for DPP-IV inhibition by single amino acid point mutations [30]. Moreover, their relevance has been demonstrated experimentally in different structure-activity relationship (SAR) studies that show how the bioactivity of DPP-IV inhibitors is strongly influenced by: **(a)** their different capacity to form salt bridges and/or hydrogen bonds with the N-terminal recognition region (which includes the Glu dyad and Tyr662); and **(b)** the electrostatic surfaces they create in this area. For instance, SAR studies have shown that a replacement of a hydroxyl group in compound **5a** by a positively charged primary amine in compound **6a** leads to a 6-fold increase in the bioactivity (*i.e.*, from IC_{50} > 100 μM to IC_{50} = 16 μM; see Figure 3A) [33] because it allows the formation of intermolecular salt bridges with Glu205/Glu206. Also, if the configuration of the carbon containing the positively charged amino group in sitagliptin is switched from **R** to **S**, there is a 24-fold decrease in the bioactivity of the resulting enantiomer (*i.e.*, from IC_{50} = 18 nM to IC_{50} = 440 nM; see Figure 3B) due to a reduced number of interactions with the N-terminal recognition region [34] [a similar effect has been shown for compounds **18** and **19** in relation to their corresponding stereoisomers (*i.e.*, compounds **21** and **20**; see Figures 3C and 3D) with a 10,000-fold and 15-fold decrease in bioactivity, respectively] [35].

The S_1 pocket is formed by the side chains of residues Tyr631, Val656, Trp659, Tyr662, Tyr666 and Val711 [19,29] and is a rigid cavity that is also highly conserved among the DPP-IV gene family [31]. SAR studies indicate that an aromatic ring is commonly used by

A

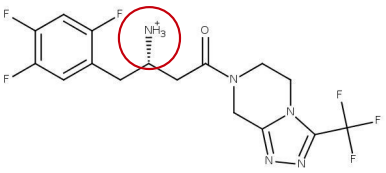


5a [33]
IC_{50} > 100 µM

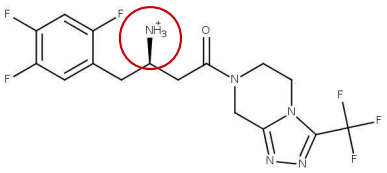


6a [33]
IC_{50} = 16 µM

B

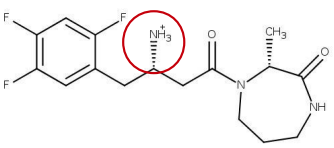


(S)-Sitagliptin [34]
IC_{50} = 440 nM

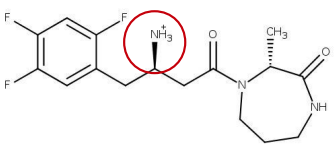


(R)-Sitagliptin [34]
IC_{50} = 18 nM

C

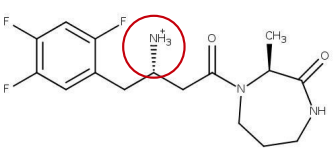


21 [35]
IC_{50} = 67000 nM

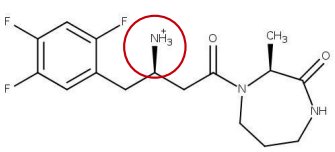


18 [35]
IC_{50} = 6.6 nM

D

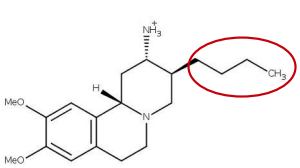


20 [35]
IC_{50} = 2300 nM

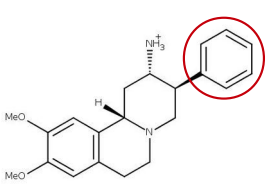


19 [35]
IC_{50} = 150 nM

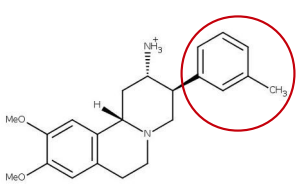
E



89/7 [37,38]
IC_{50} = 520 nM

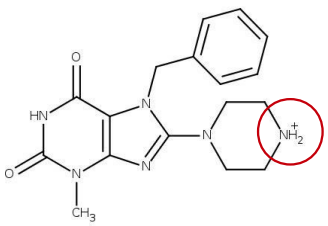


13 [38]
IC_{50} = 200 nM

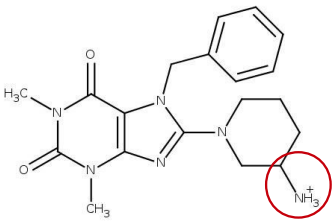


90 [37]
IC_{50} = 4.6 nM

F

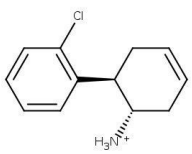


1 [43]
IC_{50} =3900 nM

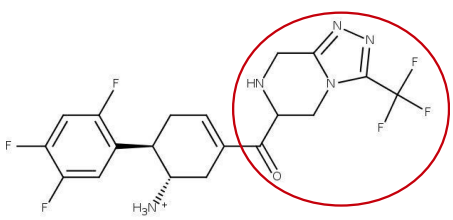


2 [43]
IC_{50} = 82 nM

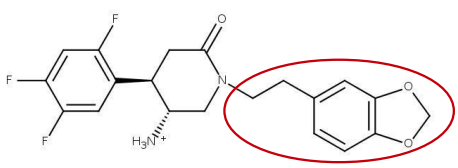
G



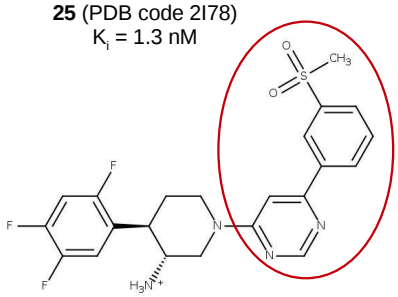
6 [45]
K_i = 0.82 μM



25 (PDB code 2I78)
K_i = 1.3 nM

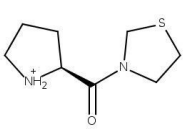


31 (PDB code 2OQI)
K_i = 3.2 nM

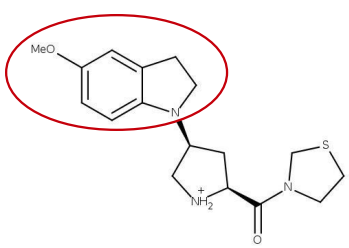


42 (PDB code 2OQV)
K_i = 4.0 nM

H



10 [46]
IC_{50} = 538 nM



22e [46]
IC_{50} = 4.8 nM

Figure 3. Chemical structures from SAR studies that are discussed in this paper. Compounds in the same panel show important changes in their DPP-IV bioactivity as a consequence of localized changes in their structures (highlighted with a red line surrounding them). Compounds in each panel are classified by increasing bioactivity from left to right.

non-peptidomimetic inhibitors to fill the S_1 pocket optimally [20,36]. For instance, the replacement of an *n*-butyl group (compound **89**/**7**) [37,38] by either a phenyl substituent (compound **13** [38]) or a 3-methylphenyl substituent (compound **90** [37]) increases the bioactivity (from IC_{50} = 520 nM to IC_{50} = 200 nM for compound **13** and to IC_{50} = 4.6 nM for compound **90**; see Figure 3E).

Therefore, the docking poses of ephedrine and ephedrine derivatives highlight the importance of these two recognition motifs, namely salt bridge and/or hydrogen bond interactions with the N-terminal recognition region and hydrophobic interactions with the S_1 pocket. The fact that the IC_{50} values for all six molecules were in the µM/mM range is also consistent with previous findings that associated these interactions with basal levels ofDPP-IV inhibition, but not necessarily with high activity/potency [39,40]. In this regard, three co-crystallized inhibitors [*i.e.*, valine-pyrrolidide (PDB code 1N1M), 1-methylamine-1-benzyl-cyclopentane (PDB code 2BUA) and 1-biphenyl-2-ylmethanamine (PDB code 3CCB) show that small molecules (170.25 Da, 175.27 Da and 183.25 Da, respectively) are able to inhibit DPP-IV, albeit with a low micromolar affinity (K_i = 2 µM, IC_{50} = 33 µM, and IC_{50} = 30 µM, respectively), by accomplishing interactions with only these two recognition motifs. This is also supported by virtual screening studies with fragments [39] and low molecular weight structures [40] that found micromolar bioactivity for chemical starting points able to simultaneously bind to the S_1 pocket and to the N-terminal recognition region.

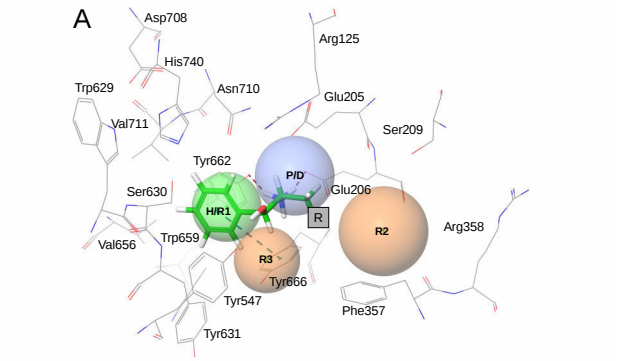
A comparison of the bioactivities for the six alkaloids shows that there is a large loss of bioactivity for N-methylephedrine and N-methylpseudoephedrine relative to the other ephedrine derivatives (see Figure 1). This difference could be explained by the fact that the positively charged amine group is not able to make as many salt bridges with the Glu dyad. Thus, in the case of the N-methylephedrine and N-methylpseudoephedrine pair, the amino group corresponds to a tertiary amine (see Figures 1E and 1F), while for the norephedrine/norpseudoephedrine and for the ephedrine/pseudoephedrine pairs, a primary and a secondary amine are placed, respectively (see Figures 1A, 1B, 1C and 1D). The introduction of a primary amino group can have an impact on potency [23,41,42] because it can contribute with three instead of one or two interactions. For instance, substituting the piperazine ring in compound **1** (IC_{50} = 3900 nM) with a 3-aminopiperidine in compound **2** (IC_{50} = 82 nM) leads to a 48-fold gain in bioactivity (see Figure 3F) [43]. Unfortunately, the

docked poses for the six compounds do not allow us to explain their bioactivity differences. For instance, although the bioactivity of ephedrine and norephedrine are higher than their corresponding pseudo partners, their docked poses show no differences in terms of their intermolecular interactions with the DPP-IV binding site. Likewise, the intermolecular interactions of the docked poses for norephedrine and N-methylephedrine are not sufficiently different (see Figures 1C and 1E) to explain the differences in bioactivity between the two compounds.

3.3. Using norephedrine as a lead compound for designing new DPP-IV inhibitors

Our results show that ephedrine derivatives have a basal DPP-IV inhibitory activity and, consequently, that these natural products are interesting to medicinal chemistry as promising lead compounds for designing potent and selective DPP-IV inhibitors. With the aim of improving the binding affinity and the selectivity for DPP-IV, we have used as the starting point for lead optimization the lowest-energy docked pose of norephedrine (with a Glide XP GScore of -8.378 kcal/mol; see Figure 4A). In this sense, the optimization process has been designed with the goal of building norephedrine derivatives that reach the S_2 extensive subsite (*i.e.*, the **R2** site of the pharmacophore) and interact with Tyr547 (*i.e.*, the **R3** site of the pharmacophore).

The reason for this selection is that π-π stacking intermolecular interaction between the ligand and Phe357 (one of the residues of the S_2 extensive subsite) has recently been demonstrated to increase bioactivity in DPP-IV inhibitors [23,44]. For instance, compounds **25**, **31** and **42** (co-crystallized with DPP-IV at the PDB structures encoded as 2I78, 2OQI and 2OQV; respectively) were developed in a lead optimization process from a low MW compound (*i.e.*, **6** with 207.70 Da; see Figure 3G) [45]. Thus, the corresponding SAR studies show how compound **6** (K_i = 0.82 μM) can be modified to become compounds **25**, **31** and **42** for reaching the S_2 extensive subsite by establishing either hydrophobic or π-π interactions that lower K_i to the 1.3-4.0 nM range for DPP-IV. Another example is compound **22e** [46], whose π-π interaction between its indolinyl moiety and Phe357 improves 116-fold its bioactivity relative to compound **10** (which has the same structure as **22e** with the exception that in **10**, no substituent fills the S_2 extensive subsite because **10** lacks a substituent equivalent to the indolinyl moiety in **22e**; see Figure 3H) [46]. Moreover, the S_2 extensive



Norephedrine

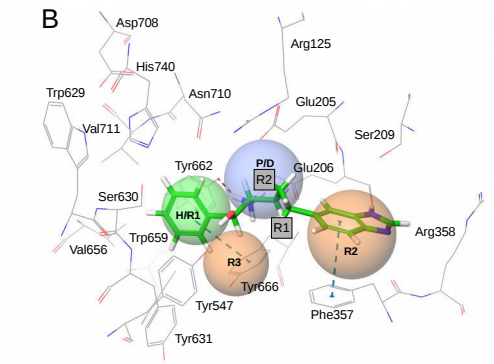
-8.378 kcal/mol

FIRST OPTIMIZATION STEP

Only candidates that simultaneously:

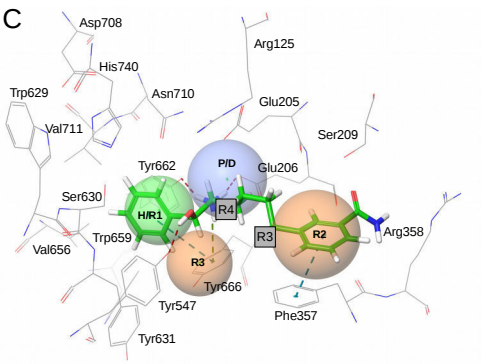
- accomplish the **R2** site of the pharmacophore
- make π-π interaction with Phe357
- are drug-like with proper ADME properties
- include significantly different substituents

are further optimized



Norephedrine-309

-9.847 kcal/mol



Norephedrine-449

-10.481 kcal/mol

SECOND OPTIMIZATION STEP

Only candidates that simultaneously:

- accomplish the **R3** site of the pharmacophore
- make π-π interaction with Tyr547
- are drug-like with proper ADME properties
- have a XP GScore more negative than the leads

Norephedrine-309 derivatives		
R1	**R2**	
309-220	309-432	309-330
	309-316	309-306
	309-318	309-280
	309-288	309-437
	309-287	309-321
	309-281	

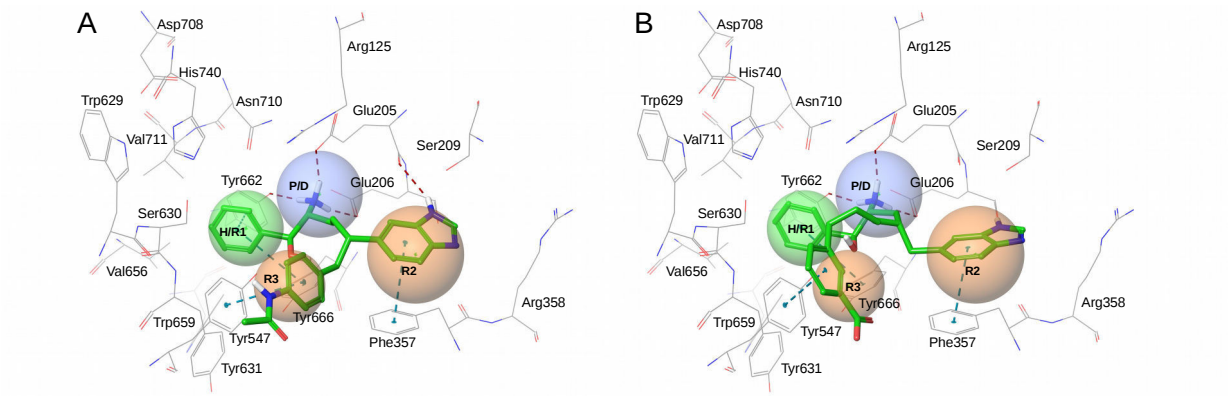
Norephedrine-449 derivatives	
R3	**R4**
449-103	449-661
	449-660
	449-734

Figure 4. Lead optimization of norephedrine that aims to obtain derivatives with improved potency and selectivity for DPP-IV. Firstly, a substituent was attached to the methyl substituent (Figure 4A) in order to reach the S_2 extensive subsite (*i.e.,* the **R2** site at the pharmacophore) and allow a π-π stacking interaction with Phe357. As a result of the first optimization and the application of the corresponding selection criteria, two derivatives (*i.e.*, norephedrine-309 and norephedrine-449) were selected for further optimization (see Figures 4B and 4C, respectively). Furthermore, two points for the attachment of new substituents were also selected on each derivative in order to allow π-π stacking interactions with Tyr547 (reaching the **R3** site of the pharmacophore). Names for the derivatives obtained after this second optimization step are classified according to the attachment position and sorted by their XP GScore (see also their 2D structure and XP GScore at Tables 2 and 3).

subsite has been described as governing the selectivity for DPP-IV over DPP8 and DPP9 [23,24]. For these reasons, a point of attachment for norephedrine substituents was placed on the methyl group of the lead (see the **R** label at Figure 4A) with the aim of establishing π-π intermolecular stacking with Phe357. Therefore, only derivatives that met all the following criteria were selected: **(a)** reaching the **R2** site at the pharmacophore; **(b)** making the π-π interaction with Phe357; and **(c)** being drug-like with proper ADME properties. The resulting derivatives were then visually inspected and only those with significantly different substituents (*i.e.*, norephedrine-309 and norephedrine-449) were considered for further optimization (see Figures 4B and 4C).

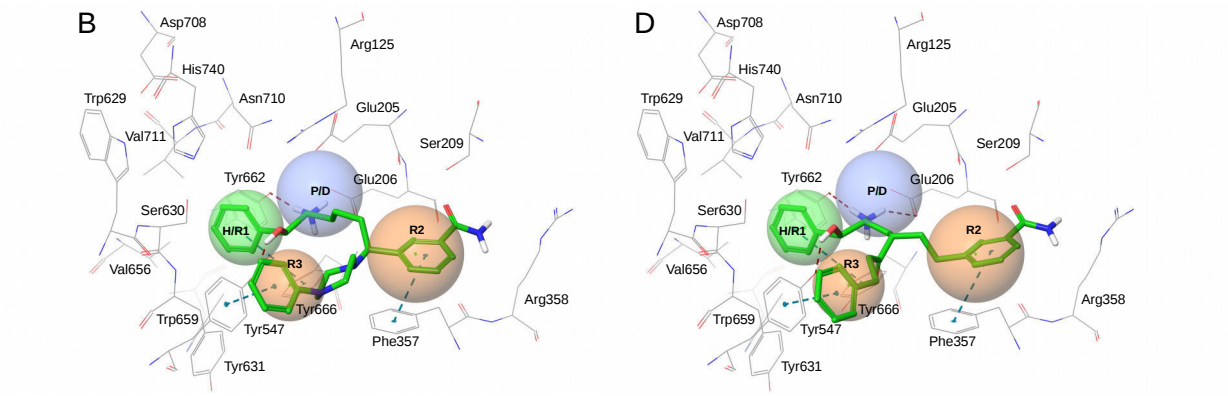
Next, a second optimization step was performed using the previous norephedrine-derivatives (*i.e.,* norphedrine-309 and norephedrine-449; see Figures 4B and 4C) with the aim of establishing additional π-π stacking interactions with Tyr547 (a residue that plays a major role stabilizing the oxyanion intermediate and is essential for the cleavage of the prolyl peptide bond of the substrate) [47,48]. This step was taken because interactions with Tyr547 have been described as improving the potency of DPP-IV inhibitors [23] and as possibly increasing the ligand's selectivity [49]. For instance, the uracil ring of alogliptin and trelagliptin and the xanthine ring of linagliptin are able to form π-π stacking interactions with Tyr547 to contribute to the inhibitory potency of these drugs (*i.e.*, 9 nM, 4.2 nM and 1 nM, respectively). In order to achieve the intermolecular interaction with Tyr547, two different attachment points were then defined for each lead (see Figures 4B and 4C). Finally, only derivatives that met all the following criteria were considered: **(a)** reaching the **R3** site of the pharmacophore; **(b)** making π-π interaction with Tyr547; **(c)** being drug-like with proper ADME properties; and **(d)** having a better XP GScore than the lead compounds. The derivatives that met all these criteria are shown in Tables 2 and 3 and the docked pose for

the best derivative at each attachment point is shown in Figure 5.



Norephedrine-309-220
-11.186 kcal/mol

Norephedrine-309-432
-11.235 kcal/mol

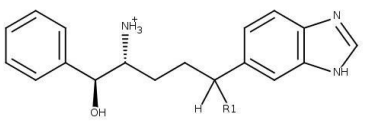


Norephedrine-449-103
-10.846 kcal/mol

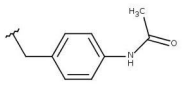
Norephedrine-449-661
-10.843 kcal/mol

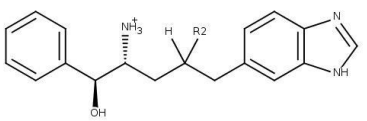
Figure 5. Lowest-energy docked poses of the derivatives obtained after second optimization step. The coordinates for DPP-IV that were used during the docking correspond to the PDB file with code 1X70. Blue dashed lines show π-π stacking intermolecular interactions whereas the red ones show either salt bridges (between the positively charged amine and the Glu dyad) or hydrogen bonds. All panels have the same orientation to allow easy comparison between them.

Table 2. Best norephedrine-309 derivatives obtained in the second optimization step. Molecules are sorted according to XP GScore. The name for each derivative was built by adding the code of the attached fragment (according to the CombiGlide Diverse Side-chain Collection) to the lead name.



Norephedrine-309

Norephedrine derivative	R1 substituent	XP GScore (Kcal/mol)
309-220		-11.186



Norephedrine-309

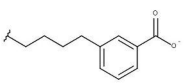
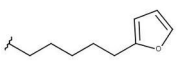
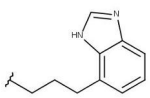
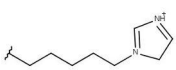
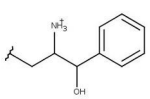
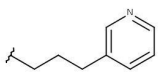
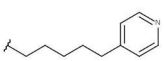
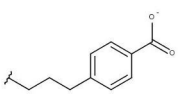
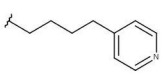
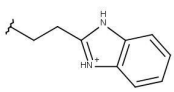
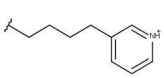
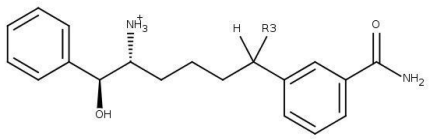
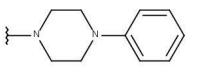
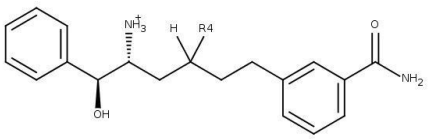
Norephedrine derivative	R2 substituent	XP GScore (Kcal/mol)	Norephedrine derivative	R2 substituent	XP GScore (Kcal/mol)
309-432		-11.235	309-330		-10.171
309-316		-10.564	309-306		-10.128
309-318		-10.550	309-280		-9.975
309-288		-10.460	309-437		-9.953
309-287		-10.340	309-321		-9.941
309-281		-10.185			

Table 3. Best norephedrine-449 derivatives obtained in the second optimization step. Molecules are sorted according to XP GScore. The name for each derivative was built by adding the code of the attached fragment (according to the CombiGlide Diverse Side-chain Collection) to the lead name.

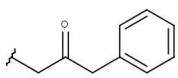
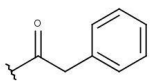
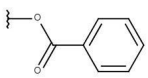


Norephedrine-449

Norephedrine derivative	R3 substituent	XP GScore (Kcal/mol)
449-103		-10.846



Norephedrine-449

Norephedrine derivative	R4 substituent	XP GScore (Kcal/mol)
449-661		-10.843
449-660		-10.730
449-734		-10.604

Remarkably, the XP GScores for these norephedrine-derivatives increased to -11.235 kcal/mol, which is even higher than the XP GScore obtained for crystallized gliptins such as alogliptin, sitagliptin, linagliptin and anagliptin (from -8.096 to -10.833 kcal/mol; see Table S2). The norephedrine-derivatives maintain the most important protein-ligand interactions found in the norephedrine core for a basal inhibitory activity. Additionally, all of them are able to reproduce most of the important intermolecular interactions described in previous SAR studies (*i.e.,* S_2 extensive subsite and Tyr547) for increasing binding affinity and selectivity for DPP-IV and they are therefore likely to exhibit nanomolar activity as inhibitors.

4. Conclusions

The World Health Organization and the International Diabetes Federation have reported that between 340 and 536 million people worldwide currently have diabetes and it is forecast that diabetes deaths will double between 2005 and 2030, thus making it the 7th leading cause of

death by 2030. Moreover, the use of several natural products (with different *Ephedra* herbs among them) has been reported as useful for reducing and preventing hyperglycemia [1,2,4]. The results of the present study experimentally demonstrated that among the antidiabetic properties previously described for *Ephedra* species [4,15,16] at least one of the mechanisms is the inhibition of DPP-IV by some of the alkaloids found in their extracts. In this regard, it has been suggested that ephedrine and pseudoephedrine account for nearly 99% of total alkaloids in some *Ephedra* species [8]. Unfortunately, several adverse effects have been described for ephedrine and ephedrine-derivatives and, as a consequence, their use has been severely restricted as food supplements by the EFSA and the FDA [14]. Nevertheless, we have shown here how their basal activity as DPP-IV inhibitors could be taken advantage of to design derivatives that, according to well-established computational approaches and SAR studies, would have proper ADME properties and improved potency and selectivity as antidiabetics.

5. Future Perspective

Recent clinical trials reported a possible cardioprotective effect of the DPP-IV inhibitors beyond the glycemic control. However, the molecular mechanism whereby these inhibitors improves cardiovascular metabolism remains unclear and further studies are needed to determine the underlying benefits. Our study reports a link between DPP-IV and β-adrenergic receptors throughout ephedrine and ephedrine-derivatives that bind to both kind of targets and, therefore, this opens the door to design potent and selective DPP-IV inhibitors that have also β-blocker activity. At this point, it is worth to remind that this is not the first time that this is suggested. Thus, Li & Yang patented a series of DPP-IV inhibitors that were claimed by the authors to show also bioactivity as β-blockers [50]. Unfortunately, Li & Yang do not explain how they find the connection between DPP-IV and β-adrenergic receptors but described the molecules they patented as useful in the treatment of neurological disorders, diabetes, inflammatory disorders such as arthritis, obesity, osteoporosis, hypertension and glaucoma. Therefore, the alkaloids of *Ephedra* are promising lead structures to develop new antidiabetic drugs with additional cardiovascular benefits as β-blockers.

Finally, preliminary studies in our lab have found that some of the β-blockers approved by the health agencies show bioactivity as DPP-IV inhibitors and we are also comparing the binding sites of both targets type in order to understand how: **(a)** this dual bioactivity is achieved; and **(b)** to optimize the structure of lead molecules to simultaneously obtain high bioactivities for a single molecule. In our opinion, the results of this research will have a great impact on the treatment of the metabolic syndrome.

6. Executive Summary

- *Ephedra* alkaloids bioactivities are evaluated *in vitro* due to their predicted antidiabetic properties as DPP-IV inhibitors.
- These compounds are experimentally confirmed to be DPP-IV inhibitors, exhibiting inhibitory activities ranged from 124 μM to 28.890 mM.
- The intermolecular interactions with DPP-IV are described by protein-ligand docking.
- These alkaloids are able to interact with the N-terminal recognition motif and the S_1 pocket, both considered to be the most important anchor points for DPP-IV inhibitor recognition.
- Norephedrine is used as a lead compound to optimize potency and selectivity.
- A link between DPP-IV and β-adrenergic receptors throughout *Ephedra* alkaloids has been established and, therefore, this opens the door to design potent and selective DPP-IV inhibitors with β-blocker activity.

Acknowledgements

This study was supported by research grants 2014PFR-URV-B2-67 and 2015PFR-URV-B2-67 from our University. AG's contract is supported by grant 2015FI_B00655 from the Catalan Government. MP and MM are Serra Húnter research fellows. We thank OpenEye Scientific Software Inc. for generously providing us with an academic license to use their programs. This manuscript has been edited in accordance with the English language usage of our University.

References

1. Patel DK, Prasad SK, Kumar R, Hemalatha S. An overview on antidiabetic medicinal plants having insulin mimetic property. *Asian Pac. J. Trop. Biomed.* 2(4), 320–30 (2012).

2. Rizvi SI, Mishra N. Traditional Indian medicines used for the management of diabetes mellitus. *J. Diabetes Res.* 2013, 712092 (2013).

3. Newman DJ, Cragg GM. Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J. Nat. Prod.* 75(3), 311–35 (2012).

4. Chauhan A, Sharma PK, Srivastava P, Kumar N, Dudhe R. Plants having potential antidiabetic activity: a review. *Der Pharm. Lett.* 2(3), 369–387 (2010).

5. Lee MR. The history of Ephedra (ma-huang). *J. R. Coll. Physicians Edinb.* 41(1), 78–84 (2011).

6. Mahdihassan S, Mehdi FS. Soma of the Rigveda and an attempt to identify it. *Am. J. Chin. Med.* 17(01n02), 1–8 (1989).

7. Abourashed EA, El-Alfy AT, Khan IA, Walker L. Ephedra in perspective – a current review. *Phytother. Res.* 17(7), 703–12 (2003).

8. Ibragic S, Sofic E. Chemical composition of various Ephedra species. *Bosn. J. Basic Med. Sci.* 15(3), 21–7 (2015).

9. Stohs SJ, Badmaev V. A review of natural stimulant and non-stimulant thermogenic agents. *Phyther. Res.* 30(5), 732–40 (2016).

10. Gurley BJ, Steelman SC, Thomas SL. Multi-ingredient, caffeine-containing dietary supplements: history, safety, and efficacy. *Clin. Ther.* 37(2), 275–301 (2015).

11. Solanki P, Yadav P, Kantharia N. Ephedrine: direct, indirect or mixed acting sympathomimetic? *Int. J. Basic Clin. Pharmacol.* 3(3), 431 (2014).

12. Stoynova V, Getov I. Review of the drug safety profile and prescription regulations of medicinal products containing ephedrine and pseudoephedrine. *J. Clin. Med.* 3(3), 41–50 (2010).

13. Fleming RM. Safety of Ephedra and related anorexic medications. *Expert Opin. Drug Saf.* 7(6),

749–59 (2008).

14. EFSA ANS Panel (EFSA Panel on Food Additives and Nutrient Sources). Scientific opinion on safety evaluation of Ephedra species for use in food. *EFSA J.* 11(11), 3467 (2013).

15. Konno C, Mizuno T, Hikino H. Isolation and hypoglycemic activity of ephedrans A, B, C, D and E, glycans of Ephedra distachya herbs. *Planta Med.* (2), 162–3 (1985).

16. Shabana MM, Mirhom YW, Genenah AA, Aboutabl EA, Amer HA. 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), 389–94 (1990).

17. Xiu LM, Miura AB, Yamamoto K, *et al.* Pancreatic islet regeneration by ephedrine in mice with streptozotocin-induced diabetes. *Am. J. Chin. Med.* 29(3-4), 493–500 (2001).
*** Experimental design showing that *Ephedra* extract promoted the regeneration of pancreas islets and restore the secretion of insulin.**

18. Guasch L, Sala E, Ojeda MJ, *et al.* Identification of novel human dipeptidyl peptidase-IV inhibitors of natural origin (part II): *in silico* prediction in antidiabetic extracts. *PLoS One*. 7(9), e44972 (2012).
**** Previous computational study that predicted ephedrine and pseudoephedrine as potential DPP-IV inhibitors.**

19. Ojeda MJ, Cereto-Massagué A, Valls C, Pujadas G. DPP-IV, an important target for antidiabetic functional food design. In: *Foodinformatics*. Springer International Publishing (2014).

20. Costante R, Stefanucci A, Carradori S, Novellino E, Mollica A. DPP-4 inhibitors: a patent review (2012 - 2014). *Expert Opin. Ther. Pat.* 25(2), 209–36 (2015).

21. Kwok AJ, Mashar M, Khavandi K, Sabir I. DPP-IV inhibitors: Beyond glycaemic control? *Trends Cardiovasc. Med.* 24(4), 157–64 (2014).

22. Opinto G, Natalicchio A, Marchetti P. Physiology of incretins and loss of incretin effect in type 2 diabetes and obesity. *Arch. Physiol. Biochem.* 119(4), 170–8 (2013).

23. Patel BD, Ghate MD. Recent approaches to medicinal chemistry and therapeutic potential of dipeptidyl peptidase-4 (DPP-4) inhibitors. *Eur. J. Med. Chem.* 74, 574–605 (2014).

24. Smelcerovic A, Miljkovic F, Kolarevic A, *et al.* An overview of recent dipeptidyl peptidase-IV inhibitors: linking their structure and physico-chemical properties with SAR, pharmacokinetics

and toxicity. *Curr. Top. Med. Chem.* 15(23), 2342–72 (2015).

*** Recent review on natural and synthetic DPP-IV inhibitors, focusing on the association between their chemical structure and mechanism of action.**

25. Shirakawa J, Amo K, Ohminami H, *et al.* Protective effects of dipeptidyl peptidase-4 (DPP-4) inhibitor against increased β cell apoptosis induced by dietary sucrose and linoleic acid in mice with diabetes. *J. Biol. Chem.* 286(29), 25467–76 (2011).

26. Mega C, Vala H, Rodrigues-Santos P, *et al.* Sitagliptin prevents aggravation of endocrine and exocrine pancreatic damage in the Zucker Diabetic Fatty rat - focus on amelioration of metabolic profile and tissue cytoprotective properties. *Diabetol. Metab. Syndr.* 6(1), 42 (2014).

27. Lankas GR, Leiting B, Roy RS, *et al.* Dipeptidyl peptidase IV inhibition for the treatment of type 2 diabetes: potential importance of selectivity over dipeptidyl peptidases 8 and 9. *Diabetes*. 54(10), 2988–94 (2005).

28. Cereto-Massagué A, Ojeda MJ, Joosten RP, *et al.* The good, the bad and the dubious: VHELIBS, a validation helper for ligands and binding sites. *J. Cheminform.* 5(1), 36 (2013).

29. Guasch L, Ojeda MJ, González-Abuín N, *et al.* Identification of novel human dipeptidyl peptidase-IV inhibitors of natural origin (part I): virtual screening and activity assays. *PLoS One*. 7(9), e44971 (2012).

30. Abbott CA, McCaughan GW, Gorrell MD. Two highly conserved glutamic acid residues in the predicted β propeller domain of dipeptidyl peptidase IV are required for its enzyme activity. *FEBS Lett.* 458(3), 278–84 (1999).

31. Pitman MR, Menz RI, Abbott CA. Hydrophilic residues surrounding the S1 and S2 pockets contribute to dimerisation and catalysis in human dipeptidyl peptidase 8 (DP8). *Biol. Chem.* 391(8), 959–72 (2010).

32. Ajami K, Abbott CA, Obradovic M, *et al.* Structural requirements for catalysis, expression, and dimerization in the CD26/DPIV gene family. *Biochemistry*. 42(3), 694–701 (2003).

33. Banno Y, Miyamoto Y, Sasaki M, *et al.* Identification of 3-aminomethyl-1,2-dihydro-4-phenyl-1-isoquinolones: a new class of potent, selective, and orally active non-peptide dipeptidyl peptidase IV inhibitors that form a unique interaction with Lys554. *Bioorg. Med. Chem.* 19(16), 4953–70 (2011).

34. Kim D, Wang L, Beconi M, *et al.*

(2R)-4-oxo-4-[3-(trifluoromethyl)-5,6-dihydro[1,2,4]triazolo[4,3-a]pyrazin-7(8H)-yl]-1-(2,4,5-triflu orophenyl)butan-2-amine: a potent, orally active dipeptidyl peptidase IV inhibitor for the treatment of type 2 diabetes. *J. Med. Chem.* 48(1), 141–51 (2005).

35. Biftu T, Feng D, Qian X, *et al.* (3R)-4-[(3R)-3-Amino-4-(2,4,5-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. *Bioorg. Med. Chem. Lett.* 17(1), 49–52 (2007).

36. Li Q, Zhou M, Han L, *et al.* Design, synthesis and biological evaluation of imidazo[1,2-a]pyridine derivatives as novel DPP-4 inhibitors. *Chem. Biol. Drug Des.* 86(4), 849–56 (2015).

37. Liu Y, Hu Y, Liu T. Recent advances in non-peptidomimetic dipeptidyl peptidase 4 inhibitors: medicinal chemistry and preclinical aspects. *Curr. Med. Chem.* 19(23), 3982–99 (2012).

38. Lübbers T, Böhringer M, Gobbi L, *et al.* 1,3-disubstituted 4-aminopiperidines as useful tools in the optimization of the 2-aminobenzo[a]quinolizine dipeptidyl peptidase IV inhibitors. *Bioorg. Med. Chem. Lett.* 17(11), 2966–70 (2007).

39. Rummey C, Nordhoff S, Thiemann M, Metz G. In silico fragment-based discovery of DPP-IV S1 pocket binders. *Bioorg. Med. Chem. Lett.* 16(5), 1405–9 (2006).

40. Ward RA, Perkins TDJ, Stafford J. Structure-based virtual screening for low molecular weight chemical starting points for dipeptidyl peptidase IV inhibitors. *J. Med. Chem.* 48(22), 6991–6 (2005).

41. Gaba M, Singh S, Gaba P. Dipeptidyl peptidase-4 inhibitors: a new approach in diabetes treatment. *Int. J. drug Dev. Res.* 1(1), 146–156 (2009).

42. Villhauer EB, Brinkman JA, Naderi GB, *et al.* 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), 2774–89 (2003).

43. Schnapp G, Klein T, Hoevels Y, Bakker RA, Nar H. Comparative analysis of binding kinetics and thermodynamics of dipeptidyl peptidase-4 inhibitors and their relationship to structure. *J. Med. Chem.* 59(16), 7466–7477 (2016).

44. Rummey C, Metz G. Homology models of dipeptidyl peptidases 8 and 9 with a focus on loop predictions near the active site. *Proteins*. 66(1), 160–71 (2007).

45. Pei Z, Li X, von Geldern TW, *et al.* Discovery of ((4R,5S)-5-amino-4-(2,4,5-trifluorophenyl)cyclohex-1-enyl)-(3-(trifluoromethyl)-5,6-dihydro-[1,2,4]triazolo[4,3-a]pyrazin-7(8H)-yl)methanone (ABT-341), a highly potent, selective, orally efficacious, and safe dipeptidyl peptidase IV inhibitor for the treatment of type 2 diabetes. *J. Med. Chem.* 49(22), 6439–42 (2006).

46. Sakashita H, Akahoshi F, Yoshida T, *et al.* Lead optimization of [*(S)*-γ-(arylamino)prolyl]thiazolidine focused on γ-substituent: Indoline compounds as potent DPP-IV inhibitors. *Bioorg. Med. Chem.* 15(2), 641–55 (2007).

47. Bjelke JR, Christensen J, Branner S, *et al.* Tyrosine 547 constitutes an essential part of the catalytic mechanism of dipeptidyl peptidase IV. *J. Biol. Chem.* 279(33), 34691–7 (2004).

48. Metzler WJ, Yanchunas J, Weigelt C, *et al.* Involvement of DPP-IV catalytic residues in enzyme-saxagliptin complex formation. *Protein Sci.* 17(2), 240–50 (2008).

49. Longenecker KL, Stewart KD, Madar DJ, *et al.* Crystal structures of DPP-IV (CD26) from rat kidney exhibit flexible accommodation of peptidase-selective inhibitors. *Biochemistry*. 45(24), 7474–82 (2006).

50. Li W, Yang C. US20130184322 A1- Novel Dipeptidyl-Peptidase-IV inhibitors. (2013).

**** The first time that some molecules are described, simultaneously, as DPP-IV inhibitors and β-blockers**

Supplementary Material

Table S1. Bioactivity data in human targets for the six Ephedra alkaloids under study. No activity data was found for N-methylpseudoephedrine in the bibliography.

Target tested	Ephedrine	Pseudoeph.	Noreph.	Norpseudoeph.	N-Methyleph.
α1A-adrenoceptor	pK_i = 4.6 [1,2] pK_i = 5.0 [3]	pK_i = 4.2 [3]	pK_i = 5.1 [3]	pK_i = 4.6 [3]	pK_i = 4.2 [3]
α1B-adrenoceptor	pK_i = 4 [1,2] pK_i < 3 [3]	pK_i < 3 [3]			
α1D-adrenoceptor	pK_i = 4.3 [1] pK_i = 4.3 [3]	pK_i < 3 [3]			
α2A-adrenoceptor	pK_i = 6.1 [1,2] pK_i = 4.8 [3]	K_i = 2011 nM [4] pK_i = 4.2 [3]	pK_i = 6.6 [8] K_i = 3024 nM [4] pK_i = 5.2 [3]	pK_i = 4.3 [3]	pK_i = 4.2 [3]
α2B-adrenoceptor	pK_i = 5.6 [1,2] K_i = 648 nM [4] pK_i = 4.8 [3]	K_i = 1870 nM [4] pK_i = 4.4 [3]	K_i = 3940 nM [4]		pK_i = 4.3 [3]
α2C-adrenoceptor	pK_i = 5.1 [1,2] K_i = 708 nM [4] pK_i = 4.8 [3]	K_i = 1220 nM [4] pK_i = 4.2 [3]	K_i = 597 nM [4] pK_i = 5.0 [3]	pK_i = 4.5 [3]	pK_i = 4.1[3]
β2-adrenoceptor	K_d = 2830 nM [5] EC_{50} = 565 nM [5]				
Solute carrier family 22	IC_{50} = 29 μM [6]	IC_{50} = 4.2 mM [7]			

1. Craig D, Forray C, Gluchowski C, Branchek T. US5610174 A1 - Use of α1C-selective adrenoceptor agonists for the treatment of urinary incontinence. (1997).

2. Craig D, Forray C, Gluchowski C, Branchek T. WO1996/38143 - The use of α1C-selective adrenoceptor agonists for the treatment of urinary incontinence. (1996).

3. Ma G, Bavadekar SA, Davis YM, *et al.* Pharmacological effects of ephedrine alkaloids on human α_1- and α_2-adrenergic receptor subtypes. *J. Pharmacol. Exp. Ther.* 322(1), 214–21 (2007).

4. Rothman RB, Vu N, Partilla JS, *et al.* In vitro characterization of ephedrine-related stereoisomers at biogenic amine transporters and the receptorome reveals selective actions as norepinephrine transporter substrates. *J. Pharmacol. Exp. Ther.* 307(1), 138–45 (2003).

5. January B, Seibold A, Whaley B, *et al.* β_2-adrenergic receptor desensitization, internalization, and phosphorylation in response to full and partial agonists. *J. Biol. Chem.* 272(38), 23871–9 (1997).

6. Suhre WM, Ekins S, Chang C, Swaan PW, Wright SH. Molecular determinants of substrate/inhibitor binding to the human and rabbit renal organic cation transporters hOCT2 and rbOCT2. *Mol. Pharmacol.* 67(4), 1067–77 (2005).

7. Rytting E, Audus KL. Novel organic cation transporter 2-mediated carnitine uptake in placental choriocarcinoma (BeWo) cells. *J. Pharmacol. Exp. Ther.* 312(1), 192–8 (2005).

8. Altenbach RJ, Khilevich A, Kolasa T, *et al.* Synthesis and structure−activity studies on N-[5-(1H-Imidazol-4-yl)-5,6,7,8-tetrahydro-1-naphthalenyl]methanesulfonamide, an imidazole-containing α_{1A} -adrenoceptor agonist. *J. Med. Chem.* 47(12), 3220–3235 (2004).

Table S2. Commercialized gliptins sorted by year of authorization by health agencies.

Compound	Year of authorization by health agencies	Developed by	XP GScore (kcal/mol)[1]
Sitagliptin	FDA, 2006 EMA, 2007	Merck & co.	-8.096 (1X70)
Vildagliptin[2]	EMA, 2007	Novartis	
Saxagliptin[2]	FDA, 2009 EMA, 2009	AstraZeneca and Bristol-Myers Squibb	
Alogliptin	Japan, 2010 FDA, 2013	Takeda Pharmaceutical Company	-9.237 (3G0B)
Linagliptin	FDA, 2011	Boehringer Ingelheim	-10.833 (2RGU)
Teneligliptin	Japan, 2012 Korea, 2014	Mitsubishi Tanabe Pharma	-9.564 (3VJK)
Gemigliptin[3]	Korea, 2012	LG Life Sciences	
Anagliptin	Japan, 2012	Sanwa Kagaku Kenkyusho Co., Ltd. and Kowa Company, Ltd	-8.096 (3WQH)
Trelagliptin	Japan, 2015	Takeda Pharmaceutical Company	-10.018 (5KBY)
Evogliptin[3]	Korea, 2015	Dong-A Pharmaceutical	
Omariglitpin	Japan, 2015	Merck & co.	-7.611 (4PNZ)[4]

[1]XP GScore for the Glide refinement of the experimental pose of the corresponding gliptin at the PDB file indicated in parenthesis. This calculation was performed only for gliptins not covalently bound to DPP-IV.

[2]Gliptin covalently bound to Ser630.

[3]No crystallized protein-ligand complex available at the PDB.

[4]Crystallized protein-ligand complex with fluoroomarigliptin and not with omarigliptin.

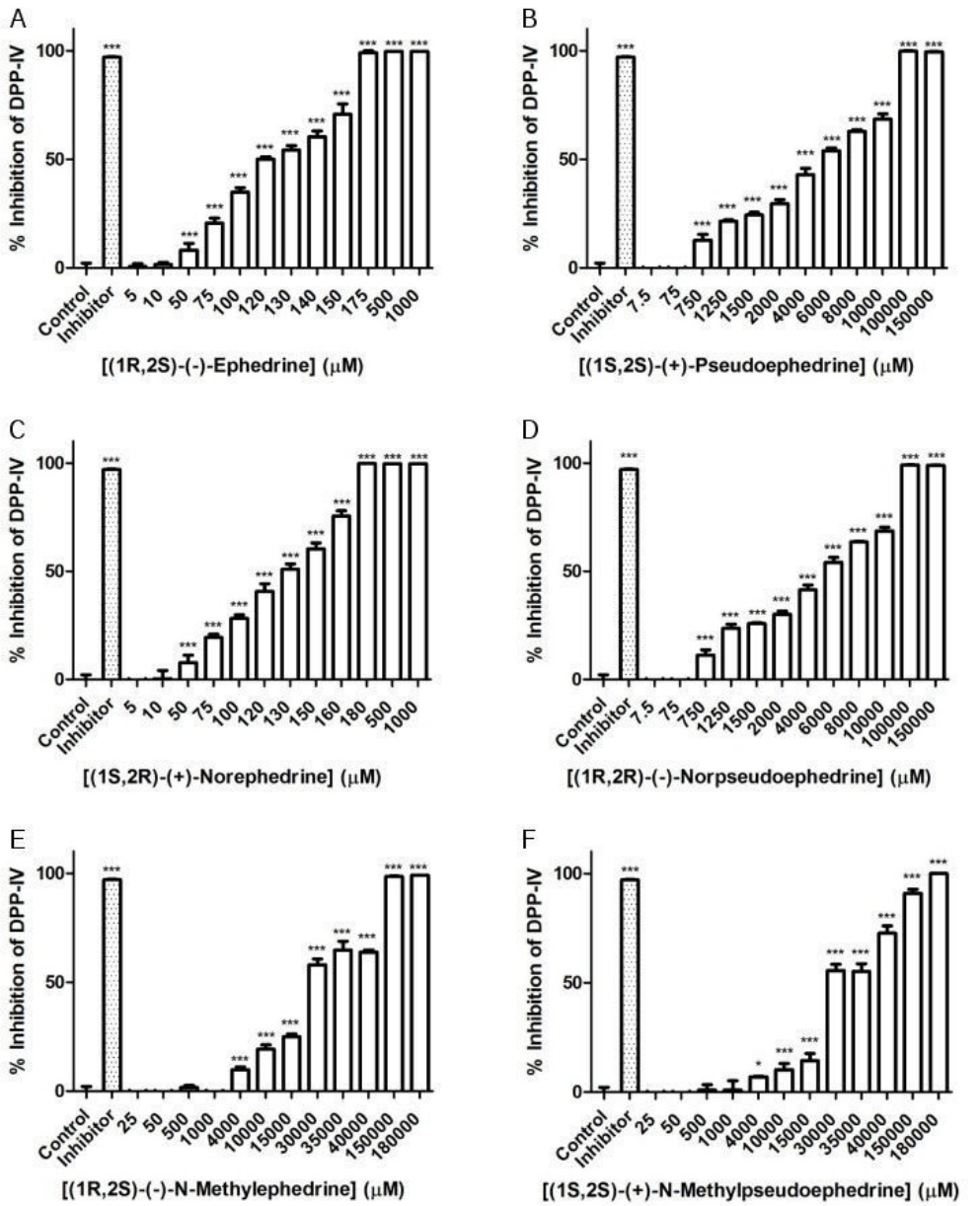


Figure S1. Dose-response of the six pure molecules tested at different concentrations to inhibit DPP-IV. *p<0.05 compared with the control. One-Way ANOVA followed by Bonferroni's Multiple Comparison Test. Data is expressed as mean ± SD of three replicates.

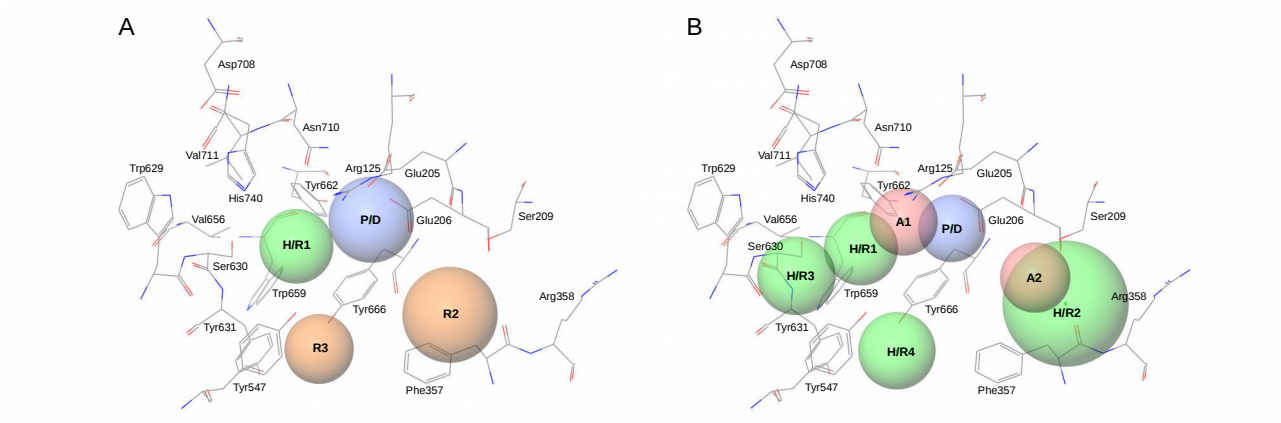


Figure S2. Comparison of the structure-based pharmacophore used in this paper (panel **A**) and the common structure-based pharmacophore used by us in previous studies (panel **B**). Main differences between them are: **(a)** optional sites **A1**, **A2** and **H/R3** are not considered now; **(b)** the optional **H/R2** and **H/R4** site have now been restricted to accept only aromatic rings, thus becoming the **R2** and the **R3** sites, respectively; and **(c)** tolerances were readjusted using, as reference, more co-crystallized DPP-IV inhibitor complexes than in the original pharmacophore.

Manuscript 4

Binding site comparison of DPP-IV and β_2-adrenergic receptor: a first step for the rational design of compounds with dual bioactivity as DPP-IV inhibitors and β-blockers

María José Ojeda-Montes[a], Àngela Casanova-Martí[b], Aleix Gimeno[a], Sarah Tomás-Hernández[a], Adrià Cereto-Massagué[a], Raúl Beltrán-Debón[a], Cristina Valls[a], Miquel Mulero[a], Montserrat Pinent[b], Gerard Pujadas[a,c,*] and Santiago Garcia-Vallvé[a,c]

[a]Research group in Cheminformatics & Nutrition, Departament de Bioquímica i Biotecnologia, Universitat Rovira i Virgili, Tarragona, Catalonia, Spain

[b]MoBioFood Research Group, Departament de Bioquímica i Biotecnologia, Universitat Rovira i Virgili, Tarragona, Catalonia, Spain

[c]EURECAT, TECNIO, CEICS, Reus, Catalonia, Spain

**Correspondence to:*
Gerard Pujadas, Research group in Cheminformatics & Nutrition, phone: +34 977 55 95 65, fax: +34 977 55 82 32. Departament de Bioquímica i Biotecnologia, Facultat de Química, Universitat Rovira i Virgili, C/ Marcel·lí Domingo 1, Edifici N4, 43007 Tarragona, Catalonia, Spain. E-mail: gerard.pujadas@urv.cat

1. Introduction

Ephedrine is an alkaloid from *Ephedra* species that was commonly used for their medicinal properties which include the treatment for asthma, fever, coughs, nasal decongestion and headache, [1,2] and, more recently, it has been used as a thermogenic agent for weight loss [3,4]. Unfortunately, ephedrine presents pharmacological and toxicological effects produced by its sympathomimetic activity, the stimulant actions of which are mediated through direct and indirect activation of α- and β-adrenergic receptors (or adrenoreceptors) [2,5,6]. Thus, the mechanism of indirect action is produced by releasing noradrenaline and preventing their reuptake whereas the direct stimulation is undergone by an agonistic action of adrenergic receptors (AR) subtypes [2]. Recently, ephedrine and its derivatives have been recently identified as lead-compounds for dipeptidyl peptidase IV (DPP-IV) inhibition. These small molecules are able to interact with DPP-IV producing basal levels of the enzyme inhibition (with IC_{50} ranging from 124 μM for ephedrine to 28.890 mM for N-methylpseudoephedrine). Previously, a relationship of DPP-IV with β_1- and β_2-AR was already suggested by Li & Yang [7]. These authors patented a series of compounds which were simultaneously claimed as DPP-IV inhibitors and β-blockers (*i.e.*, drugs that suppress β-ARs by competitive inhibition to diminish the incidence of arrhythmias and hypertension) but without reporting any evidence of how they found the connection between DPP-IV with β_1- and β_2-AR [8,9].

A series of metabolic disorders referred to as the metabolic syndrome (characterized by obesity, insulin resistance, hypertension, cardiovascular risk factor and lipid abnormalities) is common associated with the risk of developing type 2 diabetes mellitus (T2DM) and cardiovascular disease (CVD) [10]. The DPP-IV inhibition is a current treatment for T2DM allowing the regulation of glucose homeostasis by the activity of the incretin hormones. Additionally, this enzyme is involved in other pathologies such as Alzheimer or heart and renal failure [11,12]. Given that diabetes is associated with an increased risk of cardiovascular abnormalities, cardiovascular safety assessment of gliptins is required for evaluating its approval as antidiabetic therapy [13,14]. In this regard, DPP-IV inhibition seems to exert a cardioprotective effect on lowering blood pressure, maintaining weight, improving postprandial lipidemia and oxidative stress [15,16]. On the other hand, the stimulation of β-AR by catecholamines (*i.e.*, epinephrine and norepinephrine) results in an

increased heart rate, vasodilatation and conduction velocity [8,9]. However, the over-stimulation of the sympathetic system may produce pathological disorders associated with cardiomyopathy and arrhythmia, among others. Consequently, the blockage of β-AR (*i.e.*, β-blockers activity) leads to decrease the stroke volume, reduce the cardiac cycle and the liberation of noradrenaline [8]. Therefore, a single mechanism that simultaneously treating hyperglycemia and hypertension will have a great impact on the treatment of the metabolic syndrome. In that sense, it has been recently published a new class of molecular framework combining the pharmacophoric features of DPP-IV inhibitors with those of angiotensin-converting enzyme (ACE) inhibitors to produce chimeric DPP-IV/ACE inhibitors that could be used to treat, at the same time, hyperglycaemia and hypertension [17].

Due to the fact that emerging evidences have established a relationship between β-AR and DPP-IV [7,17], we hypothesize that both targets share molecular features in terms of key residues and kind of interactions. In consequence, commercialized drugs for DPP-IV could achieved dual activity as β-blockers and on the contrary, commercialized β-blockers could also inhibit DPP-IV. In this regard, the main goal of this work is to analyze and determine the main molecular features that have in common both binding sites of DPP-IV and β_2-AR and establish a background for the rational design of compounds with dual bioactivity as DPP-IV inhibitors and β-blockers.

2. Results and Discussion

2.1. DPP-IV and β_2-AR binding site description

DPP-IV is a homodimeric transmembrane glycoprotein. Each subunit of the protein is anchored to the plasma membrane by a hydrophobic helix consisting of 22 amino acids [18]. Each subunit has a large globular extracellular region that contains an active site located at the interface between the β-propeller domain (from residues 56 to 497) and the α/β hydrolase domain (from residues 509 to 766) (see Figure 1A) [11,19–21]. The cleavage of the extracellular portion of DPP-IV from the transmembrane section results in a soluble circulating form of approximately 100 kDa. This soluble form is found in plasma and cerebrospinal fluid [18,22]. DPP-IV is secreted as a mature monomer but requires dimerization to undergo normal proteolytic activity [23]. The DPP-IV binding site is highly

druggable (see Figure 1B) and its more important subsites correspond to: **(a)** the N-terminal recognition region (formed by the residues Glu205, Glu206 and Tyr662); **(b)** the S_1 pocket (formed by the residues Tyr631, Val656, Trp659, Tyr662, Tyr666 and Val711); **(c)** the S_2 extensive pocket (formed by the residues Ser209, Phe357 and Arg358); and **(d)** the S_1' pocket (formed by the residue Tyr547) [21,24,25]. In this sense, hydrogen bonds and salt bridges with the N-terminal recognition region as well as hydrophobic contacts with S_1 pocket are essential interactions for basal DPP-IV bioactivity [21,24,25], whereas the interactions with the S_2 extensive pocket and with Tyr547 (mainly by π-π stacking with Phe357 and Tyr547) are suggested to be involved in DPP-IV selectivity and to further increase the bioactivity [26,27].

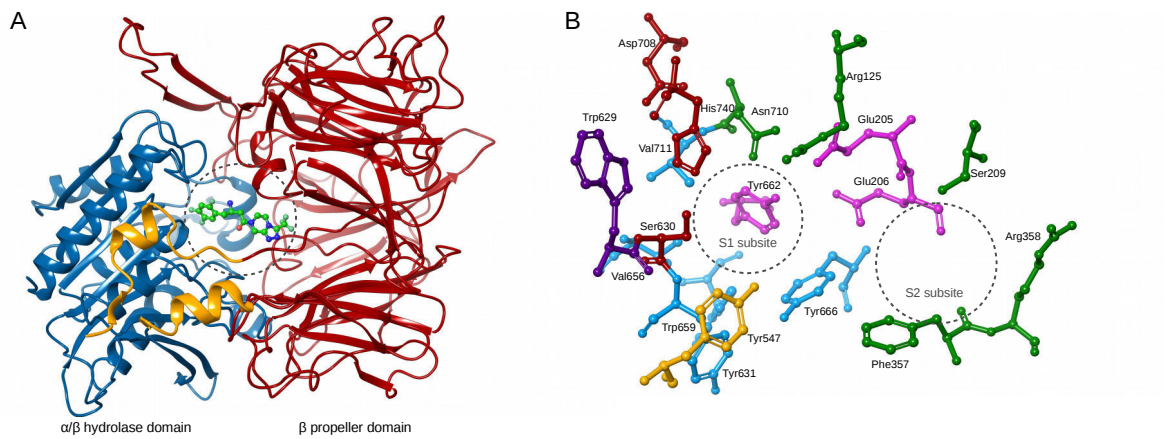


Figure 1. (A) Structure of the chain A of DPP-IV in complex with sitagliptin and **(B)** DPP-IV binding site (PDB file 1X70). Residues of DPP-IV active site are colored according to their classification, thus, those from the catalytic triad are red, those from the N-terminal recognition region are pink, those from the S_1 subsite are light blue, those from the S_2 subsite are green, those from the S_1' subsite are yellow and those from the S_2' subsite are purple.

The target β_2-AR belongs to the G-protein-coupled receptors (GPCRs) family involved in signal transduction across membranes. This receptor is constituted by a single polypeptide chain anchored in the plasma membrane organized in seven transmembrane α-helices (*i.e.*, TM) which are connected by three intracellular and three extracellular loops [28,29]. Throughout this article, the residues of β_2-AR are labeled as $ResX^{Y.Z}$ where X refers to the residue number in the β_2-AR sequence, Y to the TM segment and Z to the residue within this segment. There are three different ligand type of β_2-AR: **(a)** agonists (*i.e.*, catecholamines);

(b) antagonists and **(c)** inverse agonists; which bind the receptor in a competitive way [8]. The agonists activate the signaling pathways and increase the basal activity. The antagonists only block the signaling transductions, whereas the inverse agonists also reduce the basal activity [29]. All of them are able to bind the β_2-AR through a secondary amine group and an unsubstituted hydroxyl group, however only agonists stimulate the receptor through the hydroxyl group of the aromatic ring [8]. The binding site of β_2-AR (see Figure 2B) is described by **(a)** hydrophobic residues (Ile94$^{2.65}$, Trp109$^{3.28}$, Val117$^{3.36}$, Phe193$^{5.32}$, Trp286$^{6.48}$, Phe289$^{6.51}$, Phe290$^{6.52}$, Ile309$^{7.36}$ and Trp313$^{7.40}$); and **(b)** polar residues (His93$^{2.64}$, Asp113$^{3.32}$, Thr118$^{3.37}$, Ser203$^{5.42}$, Ser204$^{5.43}$, Ser207$^{5.46}$, Asn293$^{6.55}$, Tyr308$^{7.35}$, Asn312$^{7.39}$ and Tyr316$^{7.43}$) [29–31]. It is noteworthy that conserved hydrogen bonds are established by agonists, antagonists and inverse agonists with the receptor polar triad Asp113$^{3.32}$, Asn312$^{7.39}$, and Tyr316$^{7.43}$ [29,31]. Meanwhile, agonists frequently form polar interactions with Ser203$^{5.42}$, Ser204$^{5.43}$ and Ser207$^{5.46}$, antagonists and inverse agonists, in contrast, frequently form hydrophobic interactions with them [28,29]. Tyr308$^{7.35}$ plays an important role in the selective binding of the agonists to β_2-AR with respect toβ_1-AR [32].

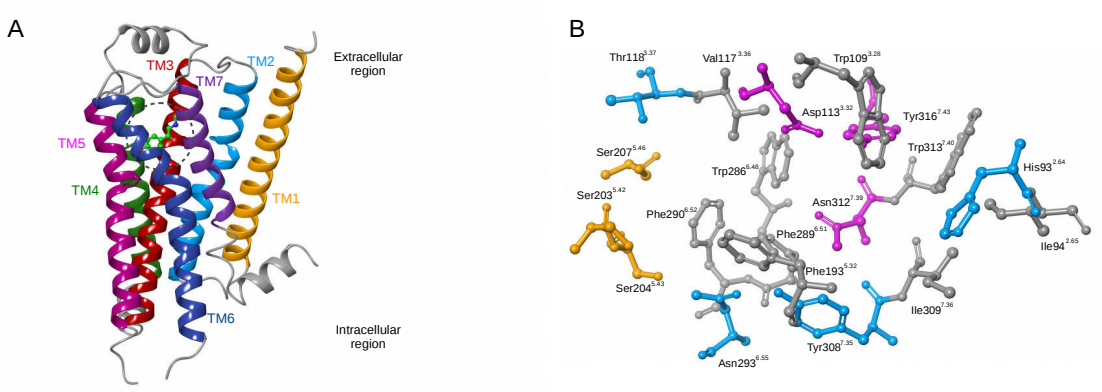


Figure 2. (A) Structure of the activated β_2-AR in complex with epinephrine and **(B)** β_2-AR binding site (PDB file 4LDO). Residues of β_2-AR colored in gray correspond to hydrophobic residues needed for stabilizing or activating the receptor, whereas the remaining residues establish polar contacts. Thus, interactions with pink residues are essential for ligand binding for agonists, antagonists and inverse agonists. The polar interaction with yellow residues are preferentially by agonists.

Although the overall structure of both proteins shown sufficient structural differences (see Figures 1A and 2A), both binding sites show quite similar volume of the cavity, thus, the

β_2-AR binding site is only 4.9% bigger than DPP-IV binding site (i.e., 2,385 and 2,507 $Å^3$ for DPP-IV and β_2-AR, respectively). Therefore, it is to be expected that ligands with similar molecular weight can be accommodated.

2.2. Search for similar binding sites to DPP-IV in web server tools

Table 1. Binding site comparison web servers used for searching the similarity of DPP-IV binding site (PDB file 1X70).

WebServer	Number of the structures at the output	PDB code for adrenergic receptor	Position found
3D-surfer [35]	1,000	-----	-----
COFACTOR [36]	10	-----	-----
GIRAF [37]	14,777	1OMW	10910
LabelHash [38]	150	-----	
PESDserv (Euclidean distance) [39]	10,000	2RH1	8417
		3D4S	8853
PoSSuM [40]	154	-----	-----
ProBiS [41]	183	-----	-----
SeSAW [42]	9	-----	-----
SPRITE [43]	400	-----	-----

A comparison of proteins on the binding sites level is a promising tool for drug repositioning, considering thus that chemically similar ligands can occupied proteins which not share a similar fold or function but share a common binding site [34]. Thus, a first approach was done by using a series of web servers for binding site comparison with the aim of disclosing any evident relationship between the different β-ARs and DPP-IV. Among the different web servers used (*i.e.*, 3D-surfer [35], COFACTOR [36], GIRAF [37], LabelHash [38], PESDserv [39], PoSSuM [40], ProBiS [41], SeSAW [42] and SPRITE [43]; see Table 1), only three crystallized structures of β-ARs have been related with the binding site of DPP-IV. Thus, a β-AR kinase (*i.e.*, PDB file 1OMW) was identified by GIRAF whereas two different entries of β_2-AR (*i.e.*, PDB files 2RH1 and 3D4S) were predicted by PESDserv for their similarity with DPP-IV. Unfortunately, these web servers do not offer any visualization of the similarity founded between both targets. Therefore, we expected that the similarity between DPP-IV

and β-ARs did not result evident based on the fact that the predicted β-ARs were found at the bottom of the corresponding sorted list in the both programs with the higher number of the structures at the output.

2.3. Critical assessment of docked poses in DPP-IV and β_2-AR binding sites

Approved drugs of both targets were used for their potency providing molecular insights into their corresponding binding modes. Therefore, self-docking and cross-docking of DPP-IV and β_2-AR co-crystallized structures were done for accommodating gliptins and β-blockers. As a result, it is remarkable to note that among the resulting XP GScores of docking to DPP-IV (see Table 2), some β-blockers such as labetalol, carvedilol, nadolol and nebivolol have obtained a value (*i.e.*, in the range -9.475 to -8.384 kcal/mol) even higher than some gliptins. The same occurs for XP GScores resulted from the docking to β_2-AR where sitagliptin, anagliptin and gemigliptin are found among the highest values. However, after visually inspected the docked poses of trelagliptin, alogliptin and linagliptin in β_2-AR and due to their low XP Gscores (*i.e.*, in the range -3.890 to -3.690 kcal/mol), it was suggested to not be suitable for blocking it. Therefore, a comparison of the binding mode of gliptins suggests that those structures which reach S2 extensive pocket of DPP-IV are more favorable than those that interact with Tyr547, it could be explained for the position of Phe193$^{5.32}$ in β_2-AR which hinder the bioactive conformations of trelagliptin, alogliptin and linagliptin into the binding site. For these reasons, a careful analyze of the binding mode of the ligands were done only focusing on these structures that have simultaneously obtained high XP GScore on both targets: **(a)** sitagliptin, anagliptin, gemigliptin and teneligliptin as DPP-IV inhibitors; and **(b)** labetalol, carvedilol, nadolol, nebivolol and bucindolol as β-blockers. Docked poses of the selected structures on both targets are shown in Figures 3 and 4 for gliptins and β-blockers, respectively.

In addition to visual inspection, the accuracy of the binding mode similarity between both docked poses on different targets was measured through **(a)** the root-mean-square deviation (RMSD) value; and **(b)** the shape and electrostatic comparison. Significant RMSD values (*i.e.*, in the 1.281 to 1.784 Å range; see Figure 3) resulted from the comparison of docked poses of gliptins in DPP-IV and β_2-AR (RMSD was calculated with the help of *rmsd.py* script of Schrödinger with the in-place option and ignoring ionization and tautomer states).

Table 2. XP GScore values for the theoretical complexes of gliptins (colored in red) and β-blockers (colored in black) against DPP-IV and β_2-AR binding sites. Columns are sorted according to increasing XP GScore values for each binding site. For carvedilol, labetalol, nebivolol and sotalol, the most potent isomer according to Ogrodowczyk *et al*. [8] was used.

DPP-IV		β2AR	
Compound name	**XP GScore (kcal/mol)**	**Compound name**	**XP GScore (kcal/mol)**
Linagliptin	-10.809	Labetalol (RR)	-12.968
Teneligliptin	-9.682	Bucindolol	-12.947
Labetalol (RR)	-9.475	Nadolol	-12.039
Carvedilol (S)	-9.298	Nebivolol (SRRR)	-11.655
Alogliptin	-9.201	Sitagliptin	-11.605
Gemigliptin	-8.976	Carvedilol (S)	-11.472
Nadolol	-8.682	Carteolol	-10.259
Nebivolol (SRRR)	-8.384	Celiprolol	-10.191
Sitagliptin	-8.365	Anagliptin	-9.834
Propranolol	-8.193	Bisoprolol	-9.745
Anagliptin	-8.177	Butaxamina	-9.633
Bucindolol	-8.174	Penbutolol	-9.576
Acebutolol	-7.922	Gemigliptin	-9.527
Celiprolol	-7.371	Timolol	-9.320
Trelagliptin	-7.010	Acebutolol	-9.273
Betaxolol	-6.820	Propranolol	-9.126
Bisoprolol	-6.463	Betaxolol	-9.098
Alprenolol	-6.379	Teneligliptin	-8.992
Pindolol	-6.200	Esmolol	-8.900
Esmolol	-6.111	Pindolol	-8.885
Butaxamina	-6.025	Alprenolol	-8.349
Atenolol	-5.652	Metoprolol	-8.172
Metoprolol	-5.591	Atenolol	-8.128
Sotalol (R)	-5.225	Sotalol (R)	-8.059
Carteolol	-4.597	Trelagliptin	-3.890
Timolol	-4.547	Alogliptin	-3.840
Penbutolol	-4.414	Linagliptin	-3.690

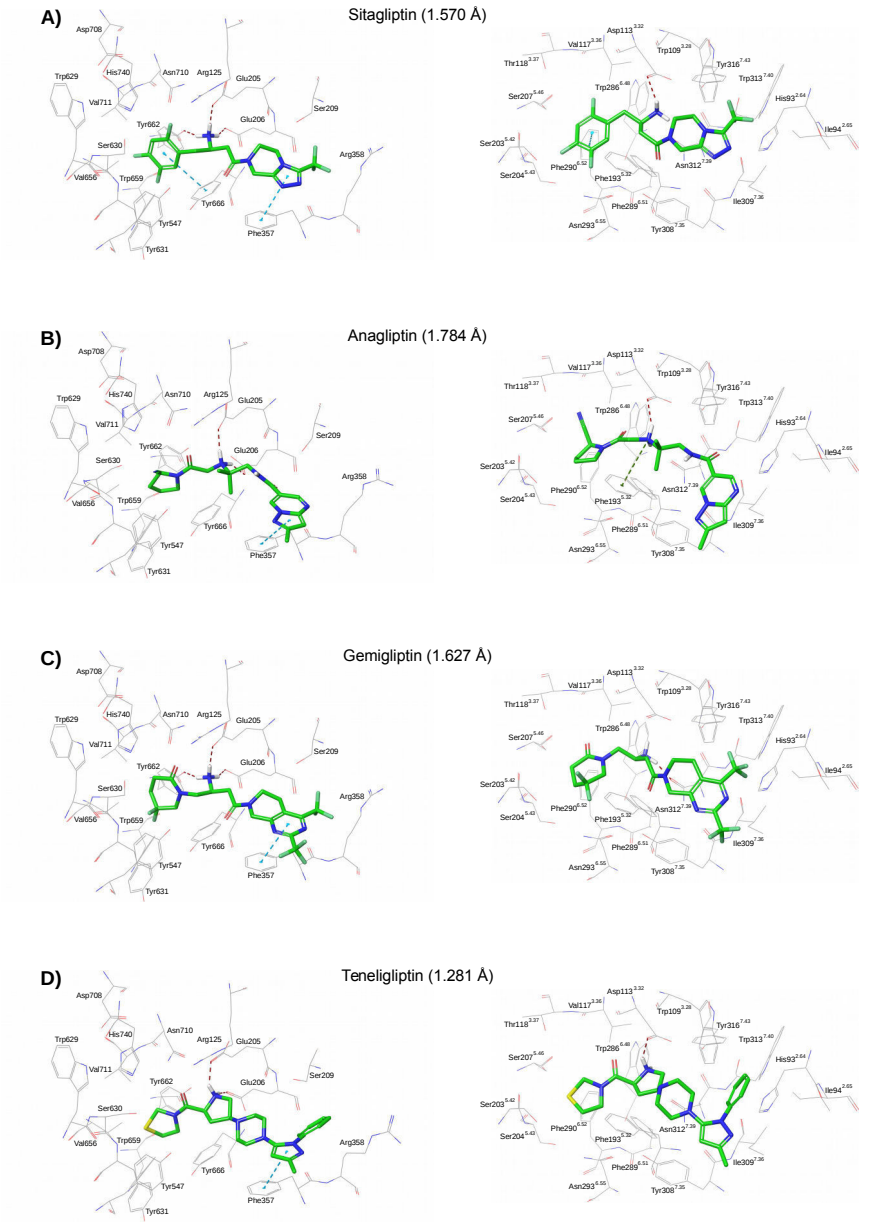
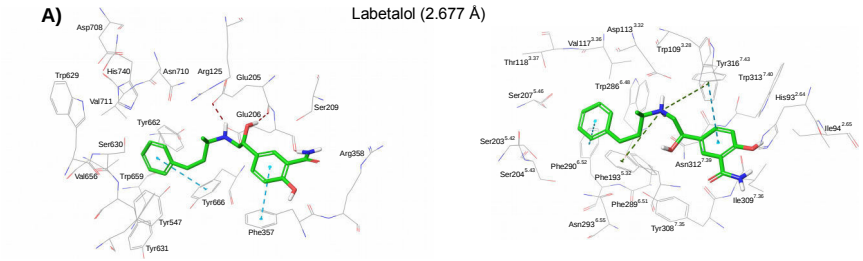
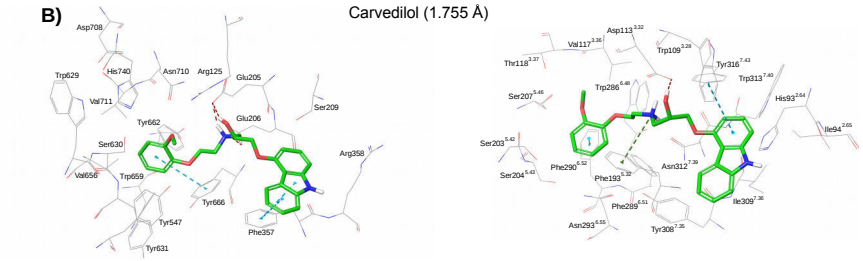


Figure 3. Comparison of the best docked poses for the gliptins at the DPP-IV (left) and β_2-AR (right; PDB file 4LDO) binding sites. When available for DPP-IV, experimental poses at the human crystallized structures have been optimized into the binding site through a docking refining process [*i.e.*, sitagliptin (PDB file 1X70), anagliptin (PDB file 3WQH) and teneligliptin (PDB file 3VJK) for DPP-IV] while gemigliptin was docked to the 1X70 file. For each theoretical complex, only the pose with the lowest XP GScore value is shown. Values in parentheses correspond to the RMSD value from the comparison of each ligand in both binding sites. Red dashed lines represent either hydrogen bonds or salt bridges whereas the blue and green ones show π-π and cation-π stacking interactions, respectively. All panels have the same orientation to allow easy comparison between them and they are sorted according decreasing XP GScore values (see Table 2) at the theoretical complexes with β_2-AR.

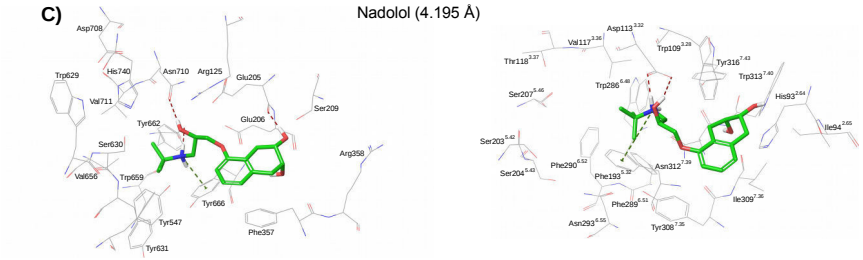
A) Labetalol (2.677 Å)



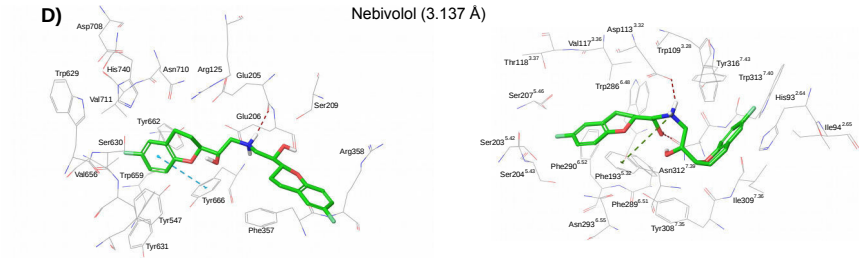
B) Carvedilol (1.755 Å)



C) Nadolol (4.195 Å)



D) Nebivolol (3.137 Å)



E) Bucindolol (9.786 Å)

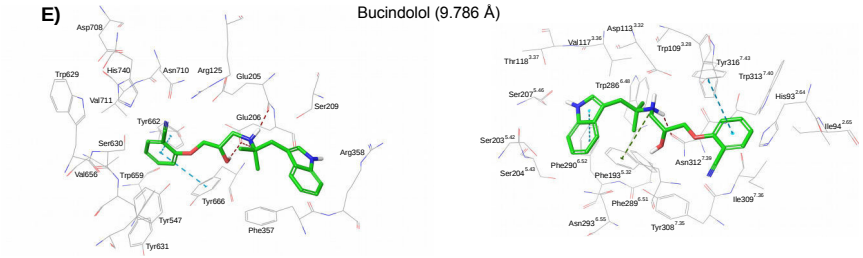


Figure 4. Comparison of the best docked poses for the β-blockers at the DPP-IV (left; PDB file 1X70) and β_2-AR (right; PDB file 4LDO) binding sites. For each theoretical complex, only the pose with the lowest XP GScore value is shown. Values in parentheses correspond to the RMSD value from the comparison of each ligand in both binding sites. Red dashed lines represent either hydrogen bonds or salt bridges whereas the blue and green ones show π-π and cation-π stacking interactions, respectively. All panels have the same orientation to allow easy comparison between them and they are sorted according decreasing XP GScore values (see Table 2) at the theoretical complexes with DPP-IV.

However, the resulted poses of β-blockers are slightly different, which thereby giving rise higher RMSD values (*i.e.*, in the 1.755 to 9.786 Å range; see Figure 4). Bucindolol becomes a particular case due to the fact that the DPP-IV pose is completely opposite to that one obtained in β_2-AR. For these reason, a modified docking protocol was done (a SP mode was initially used, followed by a XP refine) which leads the RMSD value from 9.786 to 2.006 Å (see Figure S1). It is noteworthy that the aromatic rings or the cyclohexanes of the structures which occupied the hydrophobic S_1 pocket in DPP-IV are equivalence to the hydrophobic contacts with the Ser203$^{5.42}$, Ser204$^{5.43}$, S207$^{5.46}$ characterized by the antagonists. In the same way, the amine of gliptins and β-blockers becomes relevant for interacting with the corresponding targets. In all cases, the amine is able to establish hydrogen bonds and/or salt bridges with either N-terminal recognition region of DPP-IV or Asp113$^{3.32}$ and Asn293$^{6.55}$. However, more variability is exhibited by the substituents that are accommodated in the S_2 pocket of DPP-IV respect to the pose of β_2-AR. Furthermore, the large binding mode similarity of these drugs is also reflected by the shape and electrostatic similarity values (see Figures 5 and 6 for gliptins and β-blockers, respectively). The coulombic part of the Poisson-Boltzmann electrostatics values are in the range of 0.785 for nadolol and 0.970 for teneligliptin, thus, .

2.4. Physico-chemical features shared by both DPP-IV and β$_2$-AR targets

A similar interaction pattern correlates through the docking poses of DPP-IV and β_2-AR (see Table 3). Firstly, a clear relationship is established between acidic residues of each binding site (*i.e.*, Glu205 and Glu206 for DPP-IV and Asp113$^{3.32}$ for β_2-AR) which in most cases are interacted by salt bridges with a protonated positively amine of the ligand (see Figures 3 and 4). Additionally to the significant role of these acidic residues for their corresponding inhibitory or blocking activity [24,44], they also share almost the same position on the binding site (see Figure 1). The S_1 pocket of DPP-IV has also an equivalence in β_2-AR binding site

characterized by hydrophobic features. Ser203$^{5.42}$, Ser204$^{5.43}$ and S207$^{5.46}$ are frequently interacted by hydrophobic contacts in the case of β-blockers [28,29], and are surrounded by a series of hydrophobic residues creating a cavity which is physically on S_1 pocket of the superposition with DPP-IV. In case of an aromatic ring is accommodated inside of this hydrophobic subsites, most of them are able to interact simultaneously with Tyr666 of the S_1 pocket of DPP-IV as well as the Phe290$^{6.52}$ of the β_2-AR. Consequently, the most relevant interactions for inhibit DPP-IV have an evident equivalences on β_2-AR.

Table 3. Equivalence of the residues of DPP-IV and β_2-AR binding sites and their corresponding interaction with the ligand. For each pair of residues comparison, first column corresponds to a DPP-IV residue while second column corresponds to a β_2-AR residue.

Compound name	Glu dyad	Asp113$^{3.32}$	Tyr662	Asn312$^{7.39}$	Tyr666	Phe290$^{6.52}$	Phe357	Trp109$^{3.28}$
Anagliptin	Salt bridge	Salt bridge					π-π	
Gemigliptin	Salt bridge		Hbond	Hbond			π-π	
Sitagliptin	Salt bridge	Salt bridge	Hbond		π-π	π-π	π-π	
Teneligliptin	Salt bridge	Salt bridge					π-π	
Bucindolol	Salt bridge		π-π	Hbond	π-π	π-π		π-π
Carvedilol	Salt bridge	Salt bridge			π-π	π-π	π-π	π-π
Labetalol	Salt bridge				π-π	π-π	π-π	π-π
Nadolol	Hbond	Salt bridge	Hbond				cation-π	
Nebivolol	Hbond	Salt bridge		Hbond	π-π			

However, other less frequent similarities can be determined, for example Tyr662-Asn312$^{7.39}$ and Phe357-Trp109$^{3.28}$ which though not shared the same position into the binding pocket, the same type of interaction is done with the same moiety of the ligand (see Table 3 and Figure 1). In case of Tyr662-Asn312$^{7.39}$, the first residue belongs to the N-terminal recognition region of DPP-IV whereas the second one is also a key polar residue for ligand binding of agonists, antagonists and inverse agonists in β_2-AR. Both of them are hydrogen bonded with either the charged amine or the hydroxyl group of the ligand (with the exception of bucindolol which is π-π interacted; see Figure 4E). In case of S_2 pocket of DPP-IV involved in the selectivity, it become more difficult among the drugs tested to make a relationship with the residues of β_2-AR. However, it should be to note the equivalence of Phe357-Trp109$^{3.28}$ which

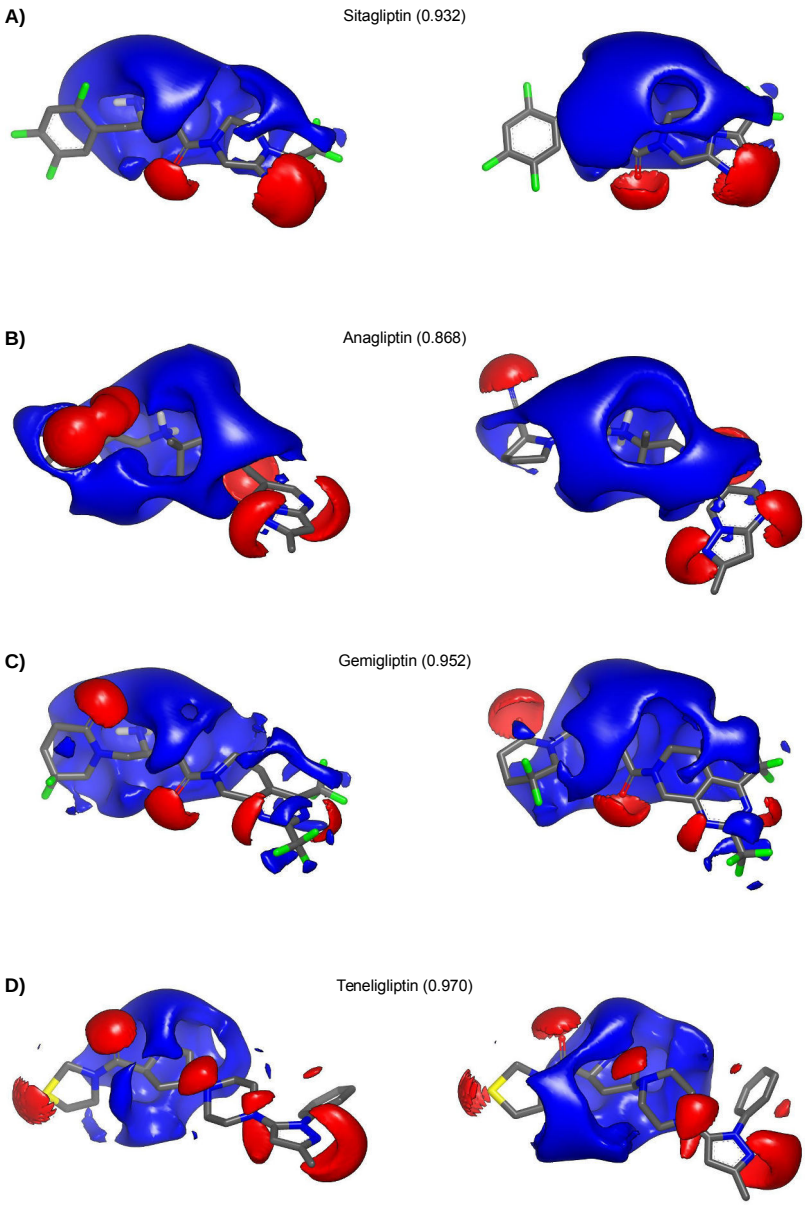


Figure 5. Electrostatic potential distribution for each gliptin docked at the DPP-IV (left) and β_2-AR (right) binding sites. Values in parentheses correspond to the coulombic part of the Poisson-Boltzmann electrostatics resulted from the comparison of both poses with EON v2.2.0.5 [58]. This Figure was drawn with VIDA v4.3.0.4 [60].

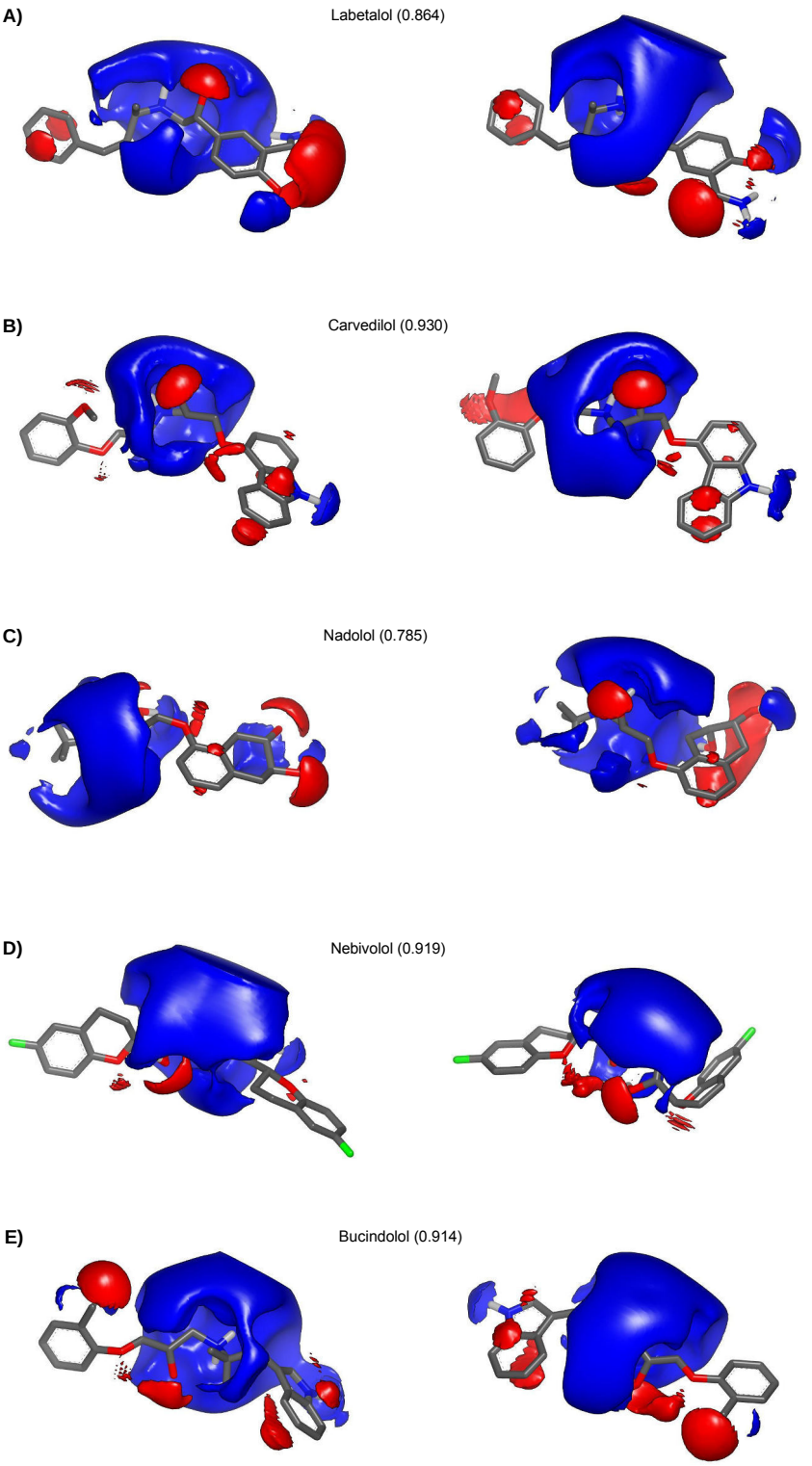

A)
Labetalol (0.864)
B)
Carvedilol (0.930)
C)
Nadolol (0.785)
D)
Nebivolol (0.919)
E)
Bucindolol (0.914)

Figure 6. Electrostatic potential distribution for each β-blocker docked at the DPP-IV (left) and β_2-AR (right) binding sites. Values in parentheses correspond to the coulombic part of the Poisson-Boltzmann electrostatics resulted from the comparison of both poses with EON v2.2.0.5 [58]. This Figure was drawn with VIDA v4.3.0.4 [60].

are both π-π stacking interacted by the same substituent of carvedilol and labetalol. As it has been mentioned before, the structures which reach Tyr547 in DPP-IV, can not optimally be accommodated in β_2-AR because of the position of Phe193$^{5.32}$ (see Figure 1). Nevertheless, this residue of β_2-AR seems to play an important role in the binding mode due to the fact that all β-blockers and anagliptin are able to interact through the charge amino by cation-π stacking.

Therefore, according to this first computational assessment some gliptins and β-blockers are suggestive of dual activity for either blocking β_2-AR and inhibiting DPP-IV.

2.5. Experimental assay of β-blockers on the DPP-IV activity

Finally, an *in vitro* evaluation has done to test if the β-blockers are able to inhibit DPP-IV enzyme (with the exception of bucindolol due to its commercial unavailability). The experimental results (see Figure 7) show that three of four β-blockers tested are able to significantly inhibit DPP-IV at 500 μM (*i.e.*, labetalol, nevibolol and carvedilol significantly inhibited 26.9%, 15.1% and 10% of the DPP-IV activity, respectively).

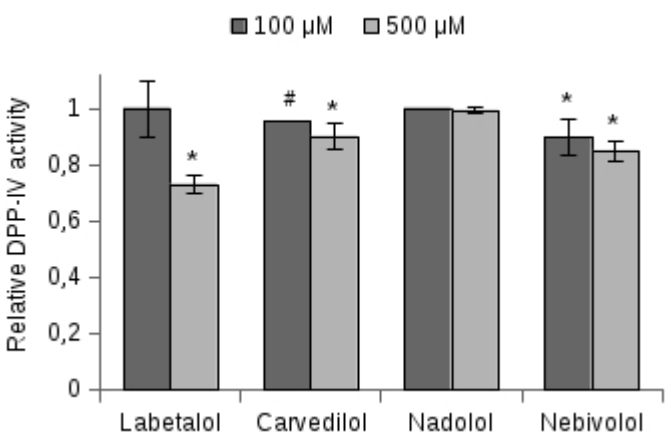


Figure 7. Dose-response results for the *in vitro* DPP-IV inhibition of the β-blockers. Each column is the average ± SEM (n=4) (vehicle, 1% DMSO). *p<0.05 # p<0.1 vs vehicle, T-student.

The fact that nadolol did not inhibit DPP-IV can be explained for its lack to fully occupy the S_1 pocket [45,46]. In addition, the secondary amine instead of establishing salt bridges with the side chains of Glu205 and Glu206, it is hydrogen bonded to Tyr662 and cation-π stacking to Tyr666. The absence of these both essential interactions can be therefore, its null activity. Regarding the actives β-blockers, their inhibitory activity can be the result of their interactions in the DPP-IV binding site. Thus, all of them contain an aromatic ring (i.e., phenyl, 6-fluoro-3,4-dihydro-2H-1-benzopyran-2-yl and 2-methoxyphenyl group for labetalol, nevibolol and carvedilol, respectively) which establishes hydrophobic contacts and π-π stacking with Tyr666. Moreover, in the case of labetalol and carvedilol, the secondary amine and the hydroxyl group are interacting by either hydrogen bond or salt bridge with the side chain of Glu205 and Glu206, and both drugs establishing π-π stacking interaction with Phe357 of the S_2 pocket.

3. Conclusions

In this work, we have established the common features in terms of residues composition and type of interactions between DPP-IV and β_2-AR. The *in vitro* evaluation of the corresponding gliptins blocking the β_2-AR is currently developing in our laboratory with the aim to experimentally demonstrate that there is a close relationship between DPP-IV and β-AR binding sites which can be targeted by chemically similar ligands achieving dual bioactivity.

Previously, recent studies reported a combined cardioprotective contribution beyond the glycemic control of the DPP-IV inhibitors. The mechanism whereby DPP-IV shows cardioprotective effect remains unclear, but different targets are suggested (*i.e.*, ACE and β_2-AR) that could be targeted by DPP-IV inhibitors. This emerging similarity between DPP-IV and β_2-AR binding sites opens the opportunity to study alternatives for a single treatment for hyperglycemia, arrhythmias and hypertension, all of them, involved in metabolic syndrome. Its relevant interest in medicinal chemistry is focused to the design of new drugs that simultaneously inhibit DPP-IV and blocking β_2-AR.

4. Experimental

4.1. Ligand references for binding site comparison

The study aims to elucidate the features shared by DPP-IV and β-AR binding sites. Thus, a first computational approach has been applied considering their corresponding approved drugs for both targets, known as gliptins and β-blockers, respectively. The fact of taking these ligands as references provides high reliability considering their potent bioactivity and the well-known binding mode. Therefore, for this purpose, seven gliptins and twenty β-blockers (antagonists and inverse agonists; see Table 2) has been used for carrying out the binding site similarity of DPP-IV and β-AR.

4.2. Ligand and protein setup

Before the protein-ligand docking with either DPP-IV or β_2-AR, the 3D structures of gliptins and β-blockers were prepared with LigPrep v3.7 [47] with the following settings: **(a)** the force field OPLS 2005 was used; **(b)** all possible ionization and tautomerization states at pH 7.0 ± 2.0 were generated with Epik; **(c)** the desalt option was activated; **(d)** the specified chiralities were retained from input geometry when generating stereoisomers; and **(e)** one low-energy ring conformation per ligand was generated. Due to the fact that the interaction of β-blockers with β-AR is highly stereoselective, only those isomers responsible of β-blocking effect were kept (see Table 2) [8].

Crystallized complexes of gliptin and β-blocker was used when available for the following docking step. Thus, in the case of β_2-AR, PDB files 3D4S [48] and 3NYA [31] were used for timolol and alprenolol, respectively, whereas 4LDO (bounded to epinephrine) [49] was used for docking the remaining of β-blockers and gliptins. On the other hand, crystallized gliptins correspond to PDB files 1X70 for sitagliptin [50], 3WQH for anagliptin [51], 3VJK for teneligliptin [52], 2RGU for linagliptin [53] and 3G0B for alogliptin [54] while PDB file 1X70 was used for docking the rest of gliptins and β-blockers. Firstly, the reliability of their coordinates fitting in the electron density map were evaluated using VHELIBS [55]. The Protein Preparation Wizard (PPW) panel [56] was used to set up these crystallized protein-ligand complexes of DPP-IV and β_2-AR. During the process and refine steps of the

PPW, all options were set to default with the exception of *remove original hydrogens*, *fill in missing side chains* and *cap termini* options that were set to on.

4.3. Volume of DPP-IV and β_2-AR binding sites

The volumes of DPP-IV and β_2-AR binding sites (PDB files 1X70 and 4LDO, respectively) were calculated by computing the molecular surface (setting the best quality) in DeepView/Swiss-PdbViewer [33]. The corresponding ligand was removed and only the significant residues described for each target were considered: **(a)** Arg125, Glu205, Glu206, Ser209, Phe357, Arg358, Tyr547, Trp629, Ser630, Tyr631, Val656, Trp659, Tyr662, Tyr666, Asp708, Asn710, Val711 and His740; and **(b)** His93$^{2.64}$, Ile94$^{2.65}$, Trp109$^{3.28}$, Asp113$^{3.32}$, Val117$^{3.36}$, Thr118$^{3.37}$, Phe193$^{5.32}$, Ser203$^{5.42}$, Ser204$^{5.43}$, Ser207$^{5.46}$, Trp286$^{6.48}$, Phe289$^{6.51}$, Phe290$^{6.52}$, Asn293$^{6.55}$, Tyr308$^{7.35}$, Ile309$^{7.36}$, Asn312$^{7.39}$, Trp313$^{7.40}$ and Tyr316$^{7.43}$; for DPP-IV and β_2-AR, respectively.

4.4. Protein-ligand docking of DPP-IV and β_2-AR

Protein-ligand docking studies were carried out with Glide v7.0 [57] with the following settings: **(a)** the binding sites for both proteins were defined by using the *Schrödinger's Grid Generation* panel with default options by using as reference the crystallized ligand; **(b)** the standard precision mode (*i.e.*, XP) was used; **(c)** *Refine* option was used for those available crystallized drug-complex and *Flexible* for the remaining gliptins and β-blockers; and **(d)** the rest of values for the docking parameters were by default. Only those structures that obtained a high XP GScore for both targets were selected for further analysis the binding site similarity.

4.5. Electrostatic and shape similarity of docked poses

The software EON v2.2.0.5 [58] compares the poses for two different compounds by calculating Tanimoto coefficients associated either to their the electrostatic potentials (*i.e.*, the Poisson-Boltzmann electrostatics and the coulombic part of the Poisson-Boltzmann electrostatics) or to their shape or to the combination of the Poisson-Boltzmann electrostatics and their shape. Thus, for the Poisson-Boltzmann electrostatic and for the coulombic part of the Poisson-Boltzmann electrostatic, the Tanimoto score can be in the -⅓ to 1 range (where

a value of 1 corresponds to identical electrostatic potential overlap whereas negative values correspond to the overlap of opposite charges between the two poses). For the shape, the Tanimoto score is a quantitative measure of three-dimensional overlap between the two poses where 1 corresponds to a perfect overlap (*i.e.*, the same shape) and 0.

In order to obtain the Tanimoto scores for the electrostatic and shape similarity by EON, the resulted DPP-IV docked poses were compared with their corresponding β_2-AR docked poses for the gliptins and β-blockers selected. All the settings were used as default with the exception that any optimization for the the terminal torsion conformer was allowed and it just scored each input conformation as-is.

4.6. Structural alignment of DPP-IV and β_2-AR

Structural alignment of DPP-IV and β_2-AR was achieved with the help of binding poses in both targets for the gemigliptin. Thus, an optimal superposition of both binding poses was firstly obtained using ROCS v3.2.1.4 [59]. Then, the residues of the β_2-AR binding site were manually displaced in the DeepView/Swiss-PdbViewer [33] using as reference the previous superposition of the ligands. As a result, DPP-IV and β_2-AR binding sites are structurally aligned based on their corresponding docked ligands.

4.7. In vitro assay of selected compounds on the inhibition of DPP-IV

Racemic mixture of the β-blockers selected (*i.e.*, labetalol [Amb3989463], nebivolol [Amb8389506], carvedilol [Amb535738] and nadolol [Amb4316155]) were purchased from Ambinter c/o Greenpharma (Orléans, France). The effect of these compounds on DPP-IV bioactivity was evaluated with the DPP-IV enzyme purified from porcine kidney (product number 317640, Merck Millipore Corporation). Stock solutions of the assayed compounds were made in DMSO and diluted in buffer (50 mM Tris-HCl) to final concentration of 500 μM and 100 μM in the assay (for a final DMSO concentration of 1%). The DPP-IV enzyme (diluted with 100 mM Tris HCl buffer pH 8.0 to 0.26 mU per well) and 10 μL of different concentration of test sample were pre-incubated for 10 min at 37°C using 96-well microplates to allow compound/enzyme interaction. Finally, the enzymatic assay was started by the addiction of 50 μL of the fluorimetric substrate H-Gly-Pro-AMC [product number I-1225, purchased from Bachem (Bubendorf, Switzerland)] at a final concentration of

0.01mM. Fluorescence was measured in a Biotek FLx800 Fluorescence Microplate Reader at Ex:380nm/Em:460nm and 37°C for 30 min. Sitagliptin was used as reference inhibitor and positive control. Two independent assays were performed, each with two technical replicates. DPP-IV inhibition is expressed as a percentage, which is the difference of the activity in presence of test compounds versus total activity of enzyme. Significant results showed $p<0.05$ with Student T test (SPSS software (SPSS, Chicago,USA)).

Acknowledgements

This study was supported by research grants 2014PFR-URV-B2-67 and 2015PFR-URV-B2-67 from our University. AG's contract is supported by grant 2015FI_B00655 from the Catalan Government. MP and MM are Serra Húnter research fellows. We thank OpenEye Scientific Software Inc. for generously providing us with an academic license to use their programs. This manuscript has been edited in accordance with the English language usage of our University.

References

[1] M.G. Soni, I.G. Carabin, J.C. Griffiths, G.A. Burdock, Safety of ephedra: lessons learned, Toxicol. Lett. 150 (2004) 97–110. doi:10.1016/j.toxlet.2003.07.006.

[2] P. Solanki, P. Yadav, N. Kantharia, Ephedrine: direct, indirect or mixed acting sympathomimetic?, Int. J. Basic Clin. Pharmacol. 3 (2014) 431. doi:10.5455/2319-2003.ijbcp20140603.

[3] S.J. Stohs, V. Badmaev, A review of natural stimulant and non-stimulant thermogenic agents, Phyther. Res. 30 (2016) 732–40. doi:10.1002/ptr.5583.

[4] M.A. Inchiosa, Experience (mostly negative) with the use of sympathomimetic agents for weight loss, J. Obes. 2011 (2011). doi:10.1155/2011/764584.

[5] S.S. Vansal, D.R. Feller, Direct effects of ephedrine isomers on human beta-adrenergic receptor subtypes, Biochem. Pharmacol. 58 (1999) 807–10.

[6] S. Kobayashi, M. Endou, F. Sakuraya, N. Matsuda, X.-H. Zhang, M. Azuma, N. Echigo, O. Kemmotsu, Y. Hattori, S. Gando, The sympathomimetic actions of l-ephedrine and d-pseudoephedrine: direct receptor activation or norepinephrine release?, Anesth. Analg. 97 (2003) 1239–45. doi:10.1213/01.ANE.0000092917.96558.3C.

[7] W. Li, C. Yang, US20130184322 A1- Novel Dipeptidyl-Peptidase-IV inhibitors, US20130184322 A1, 2013.

[8] M. Ogrodowczyk, K. Dettlaff, A. Jelinska, Beta-blockers: Current state of knowledge and perspectives, Mini Rev. Med. Chem. 16 (2016) 40–54.

[9] A.J. Szentmiklosi, N. Szentandrássy, B. Hegyi, B. Horvath, J. Magyar, T. Bányász, P.P. Nanasi, Chemistry, physiology, and pharmacology of β-adrenergic mechanisms in the heart. Why are β-blocker antiarrhythmics superior?, Curr. Pharm. Des. 21 (2015) 1030–41. doi:10.2174/1381612820666141029111240.

[10] M. Safavi, A. Foroumadi, M. Abdollahi, The importance of synthetic drugs for type 2 diabetes drug discovery, Expert Opin. Drug Discov. 8 (2013) 1339–63. doi:10.1517/17460441.2013.837883.

[11] L. Juillerat-Jeanneret, Dipeptidyl peptidase IV and its inhibitors: therapeutics for type 2 diabetes and what else?, J. Med. Chem. 57 (2014) 2197–212. doi:10.1021/jm400658e.

[12] B.D. Patel, M.D. Ghate, Recent approaches to medicinal chemistry and therapeutic potential of dipeptidyl peptidase-4 (DPP-4) inhibitors, Eur. J. Med. Chem. 74 (2014) 574–605. doi:10.1016/j.ejmech.2013.12.038.

[13] P. Gupta, W.B. White, Cardiovascular safety of therapies for type 2 diabetes, Expert Opin. Drug Saf. 16 (2017) 13–25. doi:10.1080/14740338.2017.1239707.

[14] A.B. Goldfine, Assessing the cardiovascular safety of diabetes therapies, N. Engl. J. Med. 359 (2008) 1092–5. doi:10.1056/NEJMp0805758.

[15] T.A. Salles, L. dos Santos, V.G. Barauna, A.C.C. Girardi, Potential role of dipeptidyl peptidase IV in the pathophysiology of heart failure, Int. J. Mol. Sci. 16 (2015) 4226–49. doi:10.3390/ijms16024226.

[16] A.J. Scheen, Cardiovascular effects of dipeptidyl peptidase-4 inhibitors: from risk factors to clinical outcomes, Postgrad. Med. 125 (2013) 7–20. doi:10.3810/pgm.2013.05.2659.

[17] J.A. Sattigeri, S. Sethi, J.A. Davis, S. Ahmed, G. V. Rayasam, B.G. Jadhav, S.M. Chilla, D. Datta, A. Gadhave, V.K. Tulasi, T. Jain, S. Voleti, B. Benjamin, S. Udupa, G. Jain, Y. Singh, K. Srinivas, V.S. Bansal, A. Ray, P.K. Bhatnagar, I.A. Cliffe, Approaches towards the development of chimeric DPP4/ACE inhibitors for treating metabolic syndrome, Bioorg. Med. Chem. Lett. 27 (2017) 2313–2318. doi:10.1016/j.bmcl.2017.04.036.

[18] O. Power, a B. Nongonierma, P. Jakeman, R.J. Fitzgerald, Food protein hydrolysates as a source of dipeptidyl peptidase IV inhibitory peptides for the management of type 2 diabetes., Proc. Nutr. Soc. 73 (2014) 34–46. doi:10.1017/S0029665113003601.

[19] R. Mentlein, Dipeptidyl-peptidase IV (CD26)-role in the inactivation of regulatory peptides, Regul. Pept. 85 (1999) 9–24. doi:10.1016/S0167-0115(99)00089-0.

[20] R. Thoma, B. Löffler, M. Stihle, W. Huber, A. Ruf, M. Hennig, Structural basis of proline-specific exopeptidase activity as observed in human dipeptidyl peptidase-IV., Structure. 11 (2003) 947–59. doi:10.1016/S0969-2126(03)00160-6.

[21] M. Nabeno, F. Akahoshi, H. Kishida, I. Miyaguchi, Y. Tanaka, S. Ishii, T. Kadowaki, A comparative study of the binding modes of recently launched dipeptidyl peptidase IV inhibitors

in the active site., Biochem. Biophys. Res. Commun. 434 (2013) 191–6. doi:10.1016/j.bbrc.2013.03.010.

[22] S.L. Gwaltney, J.A. Stafford, Inhibitors of Dipeptidyl Peptidase 4, in: Annu. Rep. Med. Chem., 2005: pp. 149–165. doi:10.1016/S0065-7743(05)40010-X.

[23] C.-H.H. Chien, L.-H.H. Huang, C.-Y.Y. Chou, Y.-S.S. Chen, Y.-S.S. Han, G.-G.G. Chang, P.-H.H. Liang, X. Chen, One site mutation disrupts dimer formation in human DPP-IV proteins., J. Biol. Chem. 279 (2004) 52338–45. doi:10.1074/jbc.M406185200.

[24] H. Nojima, K. Kanou, G. Terashi, M. Takeda-Shitaka, G. Inoue, K. Atsuda, C. Itoh, C. Iguchi, H. Matsubara, Comprehensive analysis of the Co-structures of dipeptidyl peptidase IV and its inhibitor, BMC Struct. Biol. 16 (2016) 11. doi:10.1186/s12900-016-0062-8.

[25] S. Arulmozhiraja, N. Matsuo, E. Ishitsubo, S. Okazaki, H. Shimano, H. Tokiwa, Comparative Binding Analysis of Dipeptidyl Peptidase IV (DPP-4) with Antidiabetic Drugs - An Ab Initio Fragment Molecular Orbital Study., PLoS One. 11 (2016) e0166275. doi:10.1371/journal.pone.0166275.

[26] G. Scapin, Structural Chemistry and Molecular Modeling in the Design of DPP4 Inhibitors, in: Multifaceted Roles Crystallogr. Mod. Drug Discov., 2015: pp. 53–67. doi:10.1007/978-94-017-9719-1_5.

[27] A. Smelcerovic, F. Miljkovic, A. Kolarevic, J. Lazarevic, A. Djordjevic, G. Kocic, M. Anderluh, An overview of recent dipeptidyl peptidase-IV inhibitors: linking their structure and physico-chemical properties with sar, pharmacokinetics and toxicity., Curr. Top. Med. Chem. 15 (2015) 2342–72.

[28] S. Costanzi, S. Vilar, In Silico screening for agonists and blockers of the β2 adrenergic receptor: Implications of inactive and activated state structures, J. Comput. Chem. 33 (2012) 561–572. doi:10.1002/jcc.22893.

[29] H.C.S. Chan, S. Filipek, S. Yuan, The principles of ligand specificity on beta-2-adrenergic receptor, Sci. Rep. 6 (2016) 34736. doi:10.1038/srep34736.

[30] C.J. Dickson, V. Hornak, C. Velez-Vega, D.J.J. McKay, J. Reilly, D.A. Sandham, D. Shaw, R.A. Fairhurst, S.J. Charlton, D.A. Sykes, R.A. Pearlstein, J.S. Duca, Uncoupling the Structure-Activity Relationships of β2 Adrenergic Receptor Ligands from Membrane Binding, J. Med. Chem. 59 (2016) 5780–9. doi:10.1021/acs.jmedchem.6b00358.

[31] D. Wacker, G. Fenalti, M.A. Brown, V. Katritch, R. Abagyan, V. Cherezov, R.C. Stevens, Conserved binding mode of human beta2 adrenergic receptor inverse agonists and antagonist revealed by X-ray crystallography, J. Am. Chem. Soc. 132 (2010) 11443–5. doi:10.1021/ja105108q.

[32] M. Isogaya, Y. Sugimoto, R. Tanimura, R. Tanaka, H. Kikkawa, T. Nagao, H. Kurose, Binding pockets of the beta(1)- and beta(2)-adrenergic receptors for subtype-selective agonists, Mol. Pharmacol. 56 (1999) 875–85.

[33] N. Guex, M.C. Peitsch, SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling, Electrophoresis. 18 (1997) 2714–23. doi:10.1002/elps.1150181505.

[34] C. Ehrt, T. Brinkjost, O. Koch, Impact of binding site comparisons on medicinal chemistry and rational molecular design, J. Med. Chem. 59 (2016) 4121–4151. doi:10.1021/acs.jmedchem.6b00078.

[35] Y. Xiong, J. Esquivel-Rodriguez, L. Sael, D. Kihara, 3D-SURFER 2.0: web platform for real-time search and characterization of protein surfaces, Methods Mol. Biol. 1137 (2014) 105–17. doi:10.1007/978-1-4939-0366-5_8.

[36] C. Zhang, P.L. Freddolino, Y. Zhang, COFACTOR: improved protein function prediction by combining structure, sequence and protein–protein interaction information, Nucleic Acids Res. (2017). doi:10.1093/nar/gkx366.

[37] A.R. Kinjo, H. Nakamura, Geometric similarities of protein-protein interfaces at atomic resolution are only observed within homologous families: an exhaustive structural classification study, J. Mol. Biol. 399 (2010) 526–40. doi:10.1016/j.jmb.2010.04.028.

[38] M. Moll, D.H. Bryant, L.E. Kavraki, The LabelHash server and tools for substructure-based functional annotation, Bioinformatics. 27 (2011) 2161–2. doi:10.1093/bioinformatics/btr343.

[39] S. Das, M.P. Krein, C.M. Breneman, PESDserv: a server for high-throughput comparison of protein binding site surfaces, Bioinformatics. 26 (2010) 1913–4. doi:10.1093/bioinformatics/btq288.

[40] J. Ito, K. Ikeda, K. Yamada, K. Mizuguchi, K. Tomii, PoSSuM v.2.0: data update and a new function for investigating ligand analogs and target proteins of small-molecule drugs, Nucleic

Acids Res. 43 (2015) D392–8. doi:10.1093/nar/gku1144.

[41] J. Konc, D. Janežič, ProBiS tools (algorithm, database, and web servers) for predicting and modeling of biologically interesting proteins, Prog. Biophys. Mol. Biol. (2017). doi:10.1016/j.pbiomolbio.2017.02.005.

[42] Osaka University, SeSAW, (n.d.). https://sysimm.ifrec.osaka-u.ac.jp/sesaw.2.0/index.php.

[43] N. Nadzirin, E.J. Gardiner, P. Willett, P.J. Artymiuk, M. Firdaus-Raih, SPRITE and ASSAM: web servers for side chain 3D-motif searching in protein structures, Nucleic Acids Res. 40 (2012) W380–6. doi:10.1093/nar/gks401.

[44] A. Plazinska, M. Kolinski, I.W. Wainer, K. Jozwiak, Molecular interactions between fenoterol stereoisomers and derivatives and the β_2-adrenergic receptor binding site studied by docking and molecular dynamics simulations, J. Mol. Model. 19 (2013) 4919–30. doi:10.1007/s00894-013-1981-y.

[45] R. Costante, A. Stefanucci, S. Carradori, E. Novellino, A. Mollica, DPP-4 inhibitors: a patent review (2012 - 2014), Expert Opin. Ther. Pat. 25 (2015) 209–36. doi:10.1517/13543776.2014.991309.

[46] Q. Li, M. Zhou, L. Han, Q. Cao, X. Wang, L. Zhao, J. Zhou, H. Zhang, Design, synthesis and biological evaluation of imidazo[1,2-a]pyridine derivatives as novel DPP-4 inhibitors, Chem. Biol. Drug Des. 86 (2015) 849–56. doi:10.1111/cbdd.12560.

[47] LigPrep, version 3.7, Schrödinger, LLC, New York, NY, (2016).

[48] M.A. Hanson, V. Cherezov, M.T. Griffith, C.B. Roth, V.-P. Jaakola, E.Y.T. Chien, J. Velasquez, P. Kuhn, R.C. Stevens, A specific cholesterol binding site is established by the 2.8 A structure of the human beta2-adrenergic receptor, Structure. 16 (2008) 897–905. doi:10.1016/j.str.2008.05.001.

[49] A.M. Ring, A. Manglik, A.C. Kruse, M.D. Enos, W.I. Weis, K.C. Garcia, B.K. Kobilka, Adrenaline-activated structure of β2-adrenoceptor stabilized by an engineered nanobody, Nature. 502 (2013) 575–9. doi:10.1038/nature12572.

[50] D. Kim, L. Wang, M. Beconi, G.J. Eiermann, M.H. Fisher, H. He, G.J. Hickey, J.E. Kowalchick, B. Leiting, K. Lyons, F. Marsilio, M.E. McCann, R.A. Patel, A. Petrov, G. Scapin, S.B. Patel, R.S. Roy, J.K. Wu, M.J. Wyvratt, B.B. Zhang, L. Zhu, N.A. Thornberry, A.E. Weber,

(2R)-4-oxo-4-[3-(trifluoromethyl)-5,6-dihydro[1,2,4]triazolo[4,3-a]pyrazin-7(8H)-yl]-1-(2,4,5-triflu orophenyl)butan-2-amine: a potent, orally active dipeptidyl peptidase IV inhibitor for the treatment of type 2 diabetes, J. Med. Chem. 48 (2005) 141–51. doi:10.1021/jm0493156.

[51] Y.S. Watanabe, Y. Yasuda, Y. Kojima, S. Okada, T. Motoyama, R. Takahashi, M. Oka, Anagliptin, a potent dipeptidyl peptidase IV inhibitor: its single-crystal structure and enzyme interactions, J. Enzyme Inhib. Med. Chem. 30 (2015) 981–8. doi:10.3109/14756366.2014.1002402.

[52] T. Yoshida, F. Akahoshi, H. Sakashita, H. Kitajima, M. Nakamura, S. Sonda, M. Takeuchi, Y. Tanaka, N. Ueda, S. Sekiguchi, T. Ishige, K. Shima, M. Nabeno, Y. Abe, J. Anabuki, A. Soejima, K. Yoshida, Y. Takashina, S. Ishii, S. Kiuchi, S. Fukuda, R. Tsutsumiuchi, K. Kosaka, T. Murozono, Y. Nakamaru, H. Utsumi, N. Masutomi, H. Kishida, I. Miyaguchi, Y. Hayashi, Discovery and preclinical profile of teneligliptin (3-[(2S,4S)-4-[4-(3-methyl-1-phenyl-1H-pyrazol-5-yl)piperazin-1-yl]pyrrolidin-2-ylcarbonyl]thiazo lidine): a highly potent, selective, long-lasting and orally active dipeptidyl peptidase IV inhibitor for t, Bioorg. Med. Chem. 20 (2012) 5705–19. doi:10.1016/j.bmc.2012.08.012.

[53] M. Eckhardt, E. Langkopf, M. Mark, M. Tadayyon, L. Thomas, H. Nar, W. Pfrengle, B. Guth, R. Lotz, P. Sieger, H. Fuchs, F. Himmelsbach, 8-(3-(R)-aminopiperidin-1-yl)-7-but-2-ynyl-3-methyl-1-(4-methyl-quinazolin-2-ylmethyl)-3,7-dihy dropurine-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 (2007) 6450–3. doi:10.1021/jm701280z.

[54] Z. Zhang, M.B. Wallace, J. Feng, J. a Stafford, R.J. Skene, L. Shi, B. Lee, K. Aertgeerts, A. Jennings, R. Xu, D.B. Kassel, S.W. Kaldor, M. Navre, D.R. Webb, S.L. Gwaltney, Design and synthesis of pyrimidinone and pyrimidinedione inhibitors of dipeptidyl peptidase IV, J. Med. Chem. 54 (2011) 510–24. doi:10.1021/jm101016w.

[55] A. Cereto-Massagué, M.J. Ojeda, R.P. Joosten, C. Valls, M. Mulero, M.J. Salvado, A. Arola-Arnal, L. Arola, S. Garcia-Vallvé, G. Pujadas, The good, the bad and the dubious: VHELIBS, a validation helper for ligands and binding sites, J. Cheminform. 5 (2013) 36. doi:10.1186/1758-2946-5-36.

[56] Schrödinger Suite 2016-1 Protein Preparation Wizard; Epik, Schrödinger, LLC, New York, NY, 2016; Impact, Schrödinger, LLC, New York, NY, 2016; Prime, Schrödinger, LLC, New York, NY, 2016, (2016).

[57] Glide, version 7.0, Schrödinger, LLC, New York, NY, (2016).

[58] EON 2.2.0.5: OpenEye Scientific Software, Santa Fe, NM., (n.d.). http://www.eyesopen.com.

[59] ROCS 3.2.1.4: OpenEye Scientific Software, Santa Fe, NM., (n.d.).

[60] VIDA 4.3.0.4: OpenEye Scientific Software, Santa Fe, NM., (n.d.). http://www.eyesopen.com.

Supplementary Material

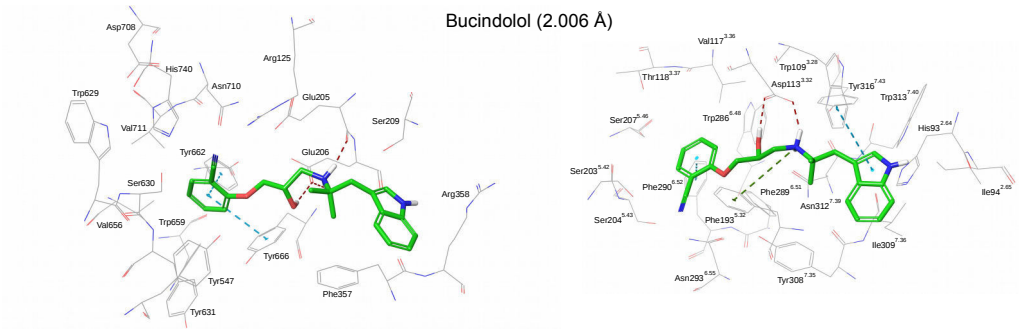


Figure S1. Comparison of the best docked poses for bucindolol at the DPP-IV (left; PDB file 1X70) and β_2-AR (right; PDB file 4LDO) binding sites. Docked pose into DPP-IV binding site was obtained by XP glide docking. Docked pose into β_2-AR binding site was obtained by SP glide docking followed by XP refined docking. As result, XP GScore values were -8.174 and -10.233, respectively. Value in parentheses correspond to the RMSD value from the comparison of bucindolol in both binding sites. Red dashed lines represent either hydrogen bonds or salt bridges whereas the blue and green ones show π-π and cation-π stacking interactions, respectively. Both panels have the same orientation to allow easy comparison between them.

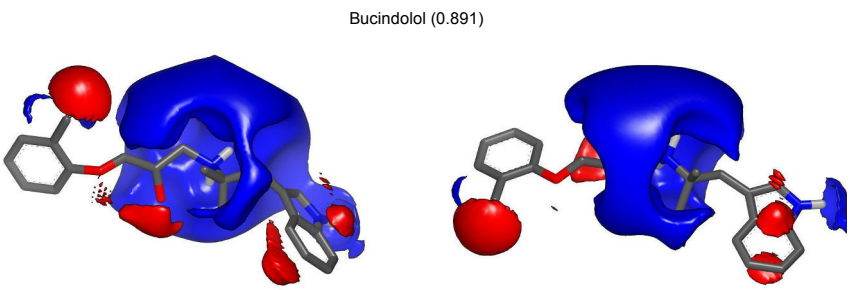
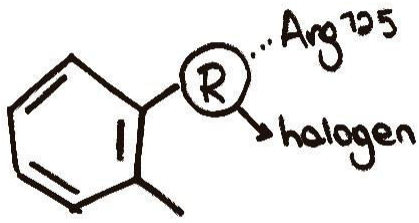


Figure S2. Electrostatic potential distribution for bucindolol at the DPP-IV (left) and β_2-AR (right) binding sites. Values in parentheses correspond to the coulombic part of the Poisson-Boltzmann electrostatics resulted from the comparison of both poses with EON v2.2.0.5 [58]. This Figure was drawn with VIDA v4.3.0.4 [60].

Summarizing Discussion



The inhibition of dipeptidyl peptidase-IV (DPP-IV) has emerged over the last decade as one of the most effective treatments for type II diabetes mellitus with low risk of hypoglycemia and weight gain. The success of this target is reflected on the launch of eleven gliptins on the market in different countries since 2006. DPP-IV inhibitors avoid the degradation of glucagon-like peptide-1 (GLP-1) and gastric inhibitory polypeptide (GIP), thus amplifying the incretin effect (*i.e.*, increasing insulin release from the β-cells and inhibiting glucagon secretion). Moreover, beyond glycemic control, the inhibition of this enzyme has an effect on cardiovascular and renal outcomes. Consequently, DPP-IV has been extensively studied and there is a considerable amount of information related to the protein structure and its inhibition, including a large number of crystal structures of the human DPP-IV and several structure-activity relationship (SAR) studies. In recent years, different virtual screening (VS) protocols have used this accurate information available for explaining how ligands interact with the DPP-IV binding site and to mine large databases of small molecules searching for new DPP-IV inhibitors. Nevertheless, most of the DPP-IV inhibitors identified have inhibitory bioactivities in the μM range and no measurement of their selectivity was performed over related enzymes like DPP8 and DPP9.

The present doctoral thesis has been therefore focused on: **(a)** the characterization of DPP-IV inhibition in order to suggest how virtual screening protocols may be improved either to favor the identification of potent and selective DPP-IV inhibitors or to look for new lead molecules; **(b)** the design of a computational strategy suitable for identifying new lead compounds with very low (or no) similarity to known actives in purchasable databases; **(c)** the demonstration that, at least partly, the described antidiabetic effect of different *Ephedra* species extracts is the result of the DPP-IV inhibitory bioactivity by ephedrine and the ephedrine-derivatives found in these extracts and **(d)** the analysis of the physico-chemical features shared by the DPP-IV and β_2-adrenergic receptors (β_2-AR) binding sites and their comparison in order to evaluate if small molecules with dual bioactivity as DPP-IV inhibitors and β-blockers are possible.

First, critical insights into the DPP-IV inhibition were established from congeneric series of SAR studies for highlighting key activity changes. Thus, we disclosed the essential features that a ligand requires to interact with DPP-IV ***(Manuscript 1)***. As a result we were able to draw eight simple rules when searching for potent and selective DPP-IV inhibitors through

VS: **(1)** a positively charged donor group (preferably a primary or secondary amine) that can establish salt bridges and/or hydrogen bonds with the N-terminal recognition region; **(2)** a group (preferably an aromatic ring) that hydrophobically contacts with the S_1 pocket; **(3)** a negatively charge group (with a partial or net charge), an acceptor group or a phenyl ring with a halogen substituent in ortho position that can establish an electrostatic interaction with Arg125; **(4)** an aromatic ring near Phe357 to form additional π-π interactions to further increase activity and selectivity; **(5)** a negatively charge group (with a partial or net charge) that can establish an electrostatic interaction with Arg358 to further increase activity and selectivity; **(6)** an aromatic ring near Tyr547 to form additional π-π interactions to further increase activity and selectivity; **(7)** a negatively charge group (with a partial or net charge) that can establish an electrostatic interaction with Lys554; and **(8)** do not reach the S_2' pocket since this may result in a decrease activity. Among these interactions, the Glu205, Glu206 and Tyr662 residues form the N-terminal recognition region and the S_1 pocket are considered to be the most important anchor points for getting basal DPP-IV inhibition.

Then, we have applied a VS workflow following protocols similar to those used by pharmaceutical industries in drug discovery, in order to focus on the most promising candidates from small molecules databases and remove those compounds that do not possess the required features. Thus, we have successfully developed a VS workflow identifying new scaffolds for DPP-IV inhibition with basal bioactivity and no similarity to known actives which therefore allow that the synthetic effort to focus solely on improving the core structure ***(Manuscript 2)***. The different sequential filters used included ADME, protein-ligand docking, pharmacophore screening and electrostatic and shape similarity analyses. It is noteworthy that a fingerprint similarity analysis was also performed at the beginning of the VS in order to find those structures that could contribute to new scaffolds and that were significantly different from co-crystallized inhibitors. The VS protocol for identifying these lead DPP-IV inhibitors was validated by an *in silico* analysis with the help of a set of actives and decoys, and by an *in vitro* analysis using an enzymatic assay. From the five compounds tested, ZINC02751967 and ZINC03823281, at 500 μM, reduced a 25.4% and a 7.6% the DPP-IV enzymatic activity, respectively. Although the bioactivity value is significantly lower than most DPP-IV inhibitors, these results showed that the VS workflow is able to identify novel scaffolds not structurally related to any known molecule that inhibits DPP-IV in purchasable databases of small molecules. Since the experimental measure of

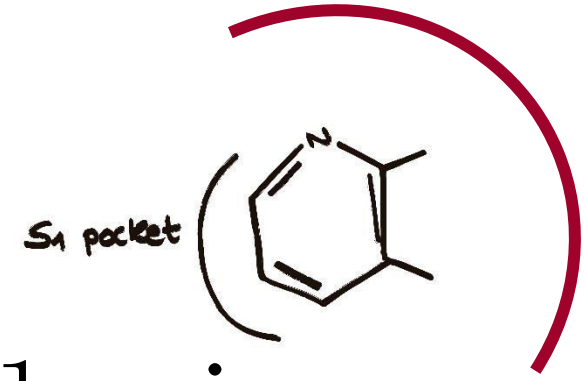
the IC_{50} value for ZINC02751967 showed that this compound only presents basal activity as DPP-IV inhibitor, different substituents were attached to ZINC02751967 with the purpose of improving the binding affinity and its selectivity according to the most important interactions described for DPP-IV. Thus, lead optimization of ZINC02751967 was addressed in order to improve the occupation of the hydrophobic S_1 pocket and to reach the S_2 extensive subsite. The resulting derivatives with proper ADME properties were finally selected.

A previous *in silico* study of our group (Guasch *et al.*, 2012) predicted that ephedrine and five ephedrine-derivatives alkaloids inhibit DPP-IV, which could explain, at least partially, the described antidiabetic effect of *Ephedra* extracts. Our experimental results confirmed that all six molecules are DPP-IV inhibitors, with an IC_{50} ranging from 124 μM for ephedrine to 28.890 mM for N-methylpseudoephedrine ***(Manuscript 3)***. However, more important than their bioactivity values is the fact that these alkaloids are able to simultaneously inhibit DPP-IV and activate the adrenergic receptors. Unfortunately, several cardiovascular adverse effects are the result of the adrenergic receptor activation by ephedrine and ephedrine-derivatives and, as a consequence, their use has been severely restricted as food supplements by the EFSA and the FDA. Therefore, a second goal of this work was to use norephedrine as a lead compound for improving affinity and selectivity for DPP-IV while avoiding the adverse effects associated to norephedrine. Thus, norephedrine-derivatives reached the S_2 extensive subsite and interacted with Tyr54. The resulting derivatives with proper ADME properties were finally selected.

Li & Yang (US20130184322, 2013) patented a series of DPP-IV inhibitors that were claimed by the authors to show also bioactivity as β-blockers. These evidences awoke our interest further research the relationship between the binding sites of DPP-IV and β_2-AR to design molecules which treat simultaneously hyperglycemia and hypertension. For this reason, the second part of this thesis is focused on analyzing the binding site similarity between both targets in terms of residues composition and type of interactions with the ligand ***(Manuscript 4)***. Gliptins and β-blockers, the drugs approved by the health agencies for DPP-IV and β_2-AR, respectively, were used as references to establish the similarities by computational and experimental techniques. As a result, a series of common molecular features shared by both targets came from the *in silico* analysis: **(a)** the acidic residues Glu205 and Glu206 of DPP-IV and $Asp113^{3.32}$ of β_2-AR interact in most cases by making salt bridges with a

positively protonated amine of the ligand; **(b)** the hydrophobic S_1 pocket of DPP-IV is spatially equivalent to the hydrophobic residues of the β_2-AR binding site, such as Ser203$^{5.42}$, Ser204$^{5.43}$, S207$^{5.46}$ and Phe290$^{6.52}$; **(c)** residues Tyr662-Asn312$^{7.39}$ and Phe357-Trp109$^{3.28}$ establish the same type of interaction with the same moiety of the ligand; and **(d)** the ligands which interact with Tyr547 in DPP-IV cannot optimally be accommodated in β_2-AR because of the steric hindrance with the residue Phe193$^{5.32}$. Regarding the experimental results, three of four β-blockers tested were able to significantly inhibit DPP-IV at 500 μM (*i.e.*, labetalol, nevibolol and carvedilol significantly inhibited DPP-IV at 26.9%, 15.1% and 10%, respectively).

More research is still needed to reinforce the binding site similarity of DPP-IV and β_2-AR, such as the *in vitro* evaluation of the corresponding gliptins blocking the β_2-AR, which is currently being developed in our laboratory. However, it is remarkable to note that our work provided new hypothesis about the cardioprotective effect associated to the inhibition of DPP-IV. The binding site similarity between DPP-IV and β_2-AR opens the door to a single treatment focused toward type II diabetes mellitus and cardiovascular diseases involved in the metabolic syndrome.



Conclusions

1. Different activity cliffs have been identified from a receptor point of view for DPP-IV inhibition. Thus, the main interactions for improving DPP-IV activity and to favor early recognition of potent DPP-IV inhibitors during VS are: Glu dyad, the S_1 pocket, the S_2 extensive subsite, Arg125 (S_2 pocket), Tyr547 (S_2/S_1' pockets and oxyanion hole), Lys554 and Trp629 (S_2' pocket).

2. The main interactions suggested for improving DPP-IV selectivity correspond to S_2 extensive subsite and Tyr547.

3. A fingerprint similarity analysis at the beginning of a virtual screening workflow can be used to enrich the sample with scaffolds with no structural similarities to any known active for the target of interest. Thus, this sample can be used to find completely new lead molecules and, therefore, reducing the cost and time of *the novo* drug design relative to other experimental or computational approaches. This computational strategy was used to identify ZINC02751967 as a new lead molecule to develop potent and selective DPP-IV inhibitors.

4. We have demonstrated that the antidiabetic properties previously associated to the *Ephedra* extract are the result, at least in part, of DPP-IV inhibition by ephedrine and ephedrine derivatives.

5. We have established a series of equivalences between the binding site of DPP-IV and β_2-AR in terms of residues composition and type of interactions with gliptins and β-blockers: **(a)** the acidic residues Glu205 and Glu206 of DPP-IV and Asp113$^{3.32}$ of β_2-AR interact in most cases by making salt bridges with a positively protonated amine of the ligand; **(b)** the hydrophobic S_1 pocket of DPP-IV is spatially equivalent to the hydrophobic residues of the β_2-AR binding site, such as Ser203$^{5.42}$, Ser204$^{5.43}$, S207$^{5.46}$ and Phe290$^{6.52}$; **(c)** residues Tyr662-Asn312$^{7.39}$ and Phe357-Trp109$^{3.28}$

establish the same type of interaction with the same moiety of the ligand; and **(d)** the ligands which interact with Tyr547 in DPP-IV cannot optimally be accommodated in β_2-AR because of the steric hindrance with the residue Phe193$^{5.32}$.

6. Three of the four β-blockers tested have been demonstrated *in vitro* to have significant DPP-IV inhibitory activity (*i.e.*, labetalol, nevibolol and carvedilol, at 500 μM, significantly reduce DPP-IV activity by 26.9%, 15.1% and 10%, respectively), and we have therefore established a first step for the rational design of compounds with dual bioactivity as DPP-IV inhibitors and β-blockers.

1. Diferents interaccions que produeixen un salt d'activitat han estat identificades per a la inhibició de DPP-IV des del punt de vista de la proteïna. De forma que s'ha establert que les interaccions més rellevants per a afavorir la cerca d'inhibidors potents de DPP-IV durant un cribratge virtual són: Glu dyad, el S_1 pocket, el S_2 extensive subsite, Arg125 (S_2 pocket), Tyr547 (S_2/S_1'pockets i oxyanion hole), Lys554 i Trp629 (S_2' pocket).

2. Les interaccions més rellevants que s'han proposat per a la millora de la selectivitat per DPP-IV corresponent al S_2 extensive subsite i al residu Tyr547.

3. Una anàlisi de similitud per *fingerprints* realitzat al principi del cribratge virtual pot ser utilitzat per a enriquir la base de dades inicial en molècules de partida que no presenten cap similitud estructural amb els actius coneguts per a una diana d'interès. De manera que la mostra de compostos obtinguda pot servir per a trobar estructures completament noves. En conseqüència, aquest procediment resulta una bona alternativa en relació a altres estratègies experimentals o computacionals atès que redueix el cost econòmic i temporal del disseny *de novo*. Com a resultat d'aquesta estratègia computacional s'ha identificat el compost ZINC02751967 com una nova molècula de partida per a desenvolupar inhibidors potents i selectius de DPP-IV.

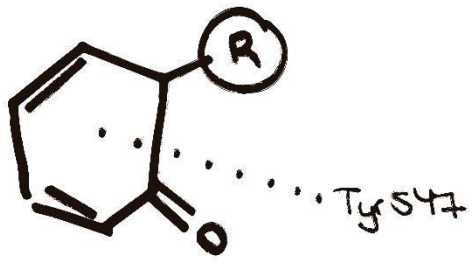
4. Hem demostrat que les propietats antidiabètiques que prèviament s'havien associat a extractes d'*Ephedra* són el resultat en certa mesura de la inhibició de DPP-IV per part de les molècules d'efedrina i dels seus derivats

5. Hem establert una sèrie d'equivalències entre entre els llocs d'unió de DPP-IV i β_2-AR a nivell de composició de residus i tipus d'interaccions amb les gliptines i els β-bloquejants: **(a)** els residus àcids Glu205 i Glu206 de DPP-IV i el residu Asp113$^{3.32}$

de β_2-AR interaccionen en molts casos fent ponts salins amb l'amina carregada positivament del lligand; **(b)** el pocket hidrofòbic S_1 de DPP-IV és equivalent a l'espai respecte als residus hidrofòbics del lloc d'unió de β_2-AR, com ara Ser203$^{5.42}$, Ser204$^{5.43}$, S207$^{5.46}$ i Phe290$^{6.52}$; **(c)** els residus Tyr662-Asn312$^{7.39}$ i Phe357-Trp109$^{3.28}$ estableixen el mateix tipus d'interaccions amb el mateix fragment del lligand; i **(d)** els lligands que interaccionen amb la Tyr547 de DPP-IV no són acomodats de forma òptima en β_2-AR a causa d'impediments estèrics amb el residu Phe193$^{5.32}$.

6. Tres dels quatre β-bloquejants testats experimentalment han demostrat que poden inhibir de forma significativa DPP-IV (labetalol, nevibolol i carvedilol han reduït l'activitat de DPP-IV en un 26.9%, 15.1% and 10% a una concentració de 500 μM, respectivament). Per consegüent, hem establert un primer pas per al disseny racional de compostos amb activitat dual com a inhibidors de DPP-IV i β-bloquejants.

List of Publications



FULL PAPERS

1. **M.J. Ojeda-Montes**, À. Casanova-Martí, A. Gimeno, S. Tomás-Hernández, A. Cereto-Massagué, G. Wolber, R. Beltrán-Debón, C. Valls, M. Mulero, M. Pinent, G. Pujadas, S. Garcia-Vallvé, *A virtual screening strategy to mine large molecular databases to find new leads with low similarity to known actives: application to find new DPP-IV inhibitors* [Submitted to European Journal of Medicinal Chemistry]

2. **M.J. Ojeda-Montes**, A. Gimeno, S. Tomás-Hernández, A. Cereto-Massagué, R. Beltrán-Debón, C. Valls, M. Mulero, G. Pujadas, S. Garcia-Vallvé, *Activity and selectivity cliffs for DPP-IV inhibitors: lessons we can learn from SAR studies and their application to virtual screening* [Submitted to Medicinal Research Reviews]

3. **M.J. Ojeda-Montes**, A. Ardid-Ruiz, S. Tomás-Hernández, A. Gimeno, A. Cereto-Massagué, R. Beltrán-Debón, M. Mulero, S. Garcia-Vallvé, G. Pujadas, C. Valls, *Ephedrine as a lead compound for the development of new DPP-IV inhibitors* [Submitted to Future Medicinal Chemistry]

4. A. Cereto-Massagué, **M.J. Ojeda-Montes,** A. Gimeno, S. Tomás-Hernández, R. Beltrán-Debón, J.M. Mateo-Sanz, C. Valls, M. Mulero, G. Pujadas, S. Garcia-Vallvé, *Molecular weight-based decoys: A simple decoy set finding alternative for fingerprint similarity approaches* [Submitted]

5. A. Cereto-Massagué, **M.J. Ojeda-Montes,** A. Gimeno, S. Tomás-Hernández, R. Beltrán-Debón, J.M. Mateo-Sanz, C. Valls, M. Mulero, S. Garcia-Vallvé, G. Pujadas, *Anglerfish: A webserver for quantitative prediction of ligand bioactivity* [Submitted]

6. S. Tomás-Hernández, S. Garcia-Vallvé, G. Pujadas, R. Beltrán-Debón, C. Valls, **M.J. Ojeda-Montes**, A. Gimeno, A. Cereto-Massagué, J. Roca-Martinez, M. Suárez, L. Arola, M. Mulero, *Anti-inflammatory and pro-apoptotic properties of the natural compound o-orsellinaldehyde through IKK-2 inhibition* [Submitted]

7. S. Tomás-Hernández, S. Garcia-Vallvé, G. Pujadas, **M.J. Ojeda-Montes**, A. Gimeno, A. Cereto-Massagué, L. Arola, L. Minghetti, M. Mulero, R. Beltrán-Debón, *Neuroprotective effects of the natural compound o-orsellinaldehyde in murine primary glial cells* [Submitted]

8. S. Tomás-Hernández, J. Blanco, C. Rojas, J. Roca-Martinez, **M.J. Ojeda-Montes**, R. Beltrán-Debón., S. Garcia-Vallvé, G. Pujadas, L. Arola, M. Mulero, *Resveratrol potently counteracts quercetin starvation-induced autophagy and sensitizes HepG2 cancer cells to apoptosis* [Submitted]

9. A. Cereto-Massagué, **M.J. Ojeda,** C. Valls, M. Mulero, G. Pujadas, S. Garcia-Vallvé, *Tools for in silico target fishing*, Methods. 71 (2015) 98–103.
DOI:10.1016/j.ymeth.2014.09.006, PMID: 25277948
Impact factor: 3.503; Category: Biochemical Research Methods; Quartile 1

10. A. Cereto-Massagué, **M.J. Ojeda**, C. Valls, M. Mulero, S. Garcia-Vallvé, G. Pujadas, *Molecular fingerprint similarity search in virtual screening*, Methods. 71 (2015) 58–63. DOI:10.1016/j.ymeth.2014.08.005, PMID: 25132639
Impact factor: 3.503; Category: Biochemical Research Methods; Quartile 1

11. A. Cereto-Massagué, **M.J. Ojeda**, R.P. Joosten, C. Valls, M. Mulero, M.J. Salvado, A. Arola-Arnal, L. Arola, S. Garcia-Vallvé, G. Pujadas, *The good, the bad and the dubious: VHELIBS, a validation helper for ligands and binding sites*, J. Cheminform. 5 (2013) 36.
DOI:10.1186/1758-2946-5-36, PMID: 23895374
Impact factor: 4.540; Category: Chemistry, multidisciplinary; Quartile 1

12. L. Guasch, **M.J. Ojeda**, N. González-Abuín, E. Sala, A. Cereto-Massagué, M. Mulero, C. Valls, M. Pinent, A. Ardévol, S. Garcia-Vallvé, G. Pujadas, *Identification of novel human dipeptidyl peptidase-IV inhibitors of natural origin (part I): virtual screening and activity assays,* PLoS One. 7 (2012) e44971.
DOI:10.1371/journal.pone.0044971, PMID: 22984596
Impact factor: 3.730; Category: Multidisciplinary science; Quartile 1

13. L. Guasch, E. Sala, **M.J. Ojeda**, C. Valls, C. Bladé, M. Mulero, M. Blay, A. Ardévol, S. Garcia-Vallvé, G. Pujadas, *Identification of novel human dipeptidyl peptidase-IV inhibitors of natural origin (part II): in silico prediction in antidiabetic extracts*, PLoS One. 7 (2012) e44972. DOI:10.1371/journal.pone.0044972, PMID: 23028712
Impact factor: 3.730; Category: Multidisciplinary science; Quartile 1

BOOK CHAPTERS

1. **M.J. Ojeda**, A. Cereto-Massagué, C. Valls, G. Pujadas, *DPP-IV, an important target for antidiabetic functional food design*, In: Foodinformatics, pp 177-212, Springer International Publishing, 2014. ISBN: 978-3-319-10225-2; DOI:10.1007/978-3-319-10226-9_7.

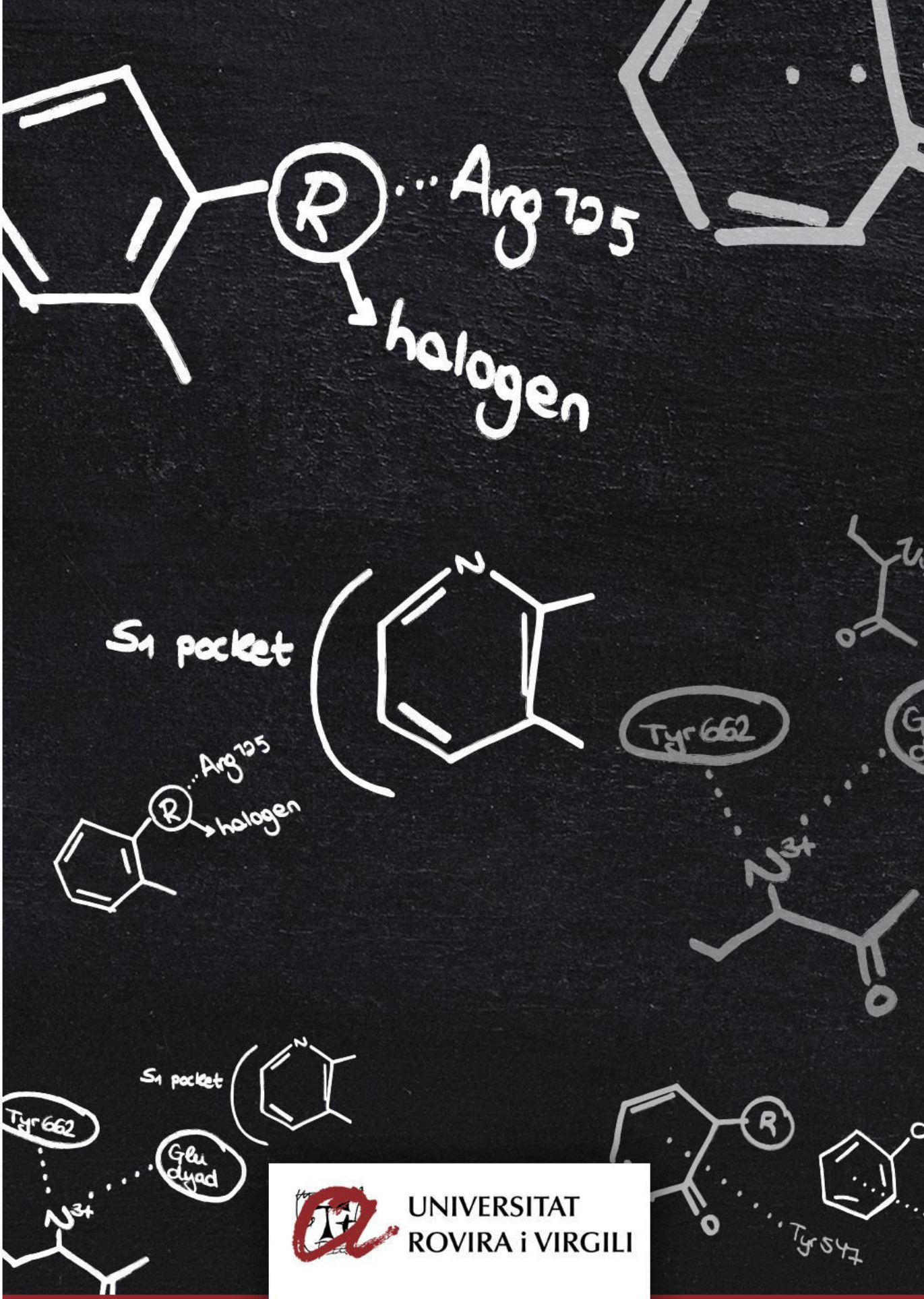
WORKSHOPS

1. Oral presentation: **"Evaluation of conformer generation tools to reproduce bioactive conformer"**
   31 de Octubre 2013 – A Practical Introduction to Using Cresset's Tools for Imaginative Molecule Design
   Centre Supercomputació de Catalunya (Barcelona)

2. Oral presentation and poster: **"Evaluation of conformer generation tools to reproduce bioactive conformer"**
   20-21 de Junio 2013 – Field Based Chemistry Europe 2013 Cresset Group
   Cambridge (UK)

3. Oral presentation: **"Identification of novel Human Dipeptidyl Peptidase-IV Inhibitors of natural origin (Part I): Virtual screening and activity assays"**
   4-5th June 2013 – Computational Chemistry and Protein Modelling, Schrödinger Workshop
   Centre Supercomputació de Catalunya (Barcelona)

4. Poster: **"Evaluation of conformer generation tools to reproduce bioactive conformer"**
   19-25th May 2013 – 9th European workshop in drug design
   Certosa di Pontignano (Siena, Italia)



UNIVERSITAT ROVIRA i VIRGILI