Medicinal Research Reviews

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

Journal:	*Medicinal Research Reviews*
Manuscript ID	MED-17-054.R1
Wiley - Manuscript type:	Review Article
Date Submitted by the Author:	n/a
Complete List of Authors:	OJEDA MONTES, MARÍA JOSÉ; Universitat Rovira i Virgili, Dept. Biochemistry and Biotechnology Gimeno, Aleix; Universitat Rovira i Virgili, Biochemistry & Biotechnology Tomas-Hernández, Sarah; Universitat Rovira i Virgili, Biochemistry & Biotechnology Cereto-Massagué, Adrià; Universitat Rovira i Virgili, Biochemistry & Biotechnology Beltrán-Debón, Raúl; Universitat Rovira i Virgili, Biochemistry & Biotechnology Valls, Cristina; Universitat Rovira i Virgili, Biochemistry & Biotechnology Mulero, Miquel; Universitat Rovira i Virgili, Biochemistry & Biotechnology Pujadas, Gerard; Universitat Rovira i Virgili, Biochemistry & Biotechnology Garcia-Vallvé, Santiago; Universitat Rovira i Virgili, Biochemistry & Biotechnology
Keywords:	GIP, gliptins, GLP-1, hyperglycemia treatment, incretins

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[1,†], Aleix Gimeno[1,†], Sarah Tomas-Hernández[1], Adrià Cereto-Massagué[1], Raúl Beltrán-Debón[1], Cristina Valls[1], Miquel Mulero[1], Gerard Pujadas[1,2,*] and Santiago Garcia-Vallvé[1,2]

[1]Research Group in Cheminformatics & Nutrition, Departament de Bioquímica i Biotecnologia, Universitat Rovira i Virgili, Campus de Sescelades, 43007 Tarragona, Catalonia, Spain

[2]EURECAT, TECNIO, CEICS, Avinguda Universitat 1, 43204 Reus, Catalonia, Spain

[†]*Both authors contributed equally to this work*

**Correspondence to:* Gerard Pujadas, Research Group in Cheminformatics & Nutrition, tel: +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 s/n, 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)** 74 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.

Key words: GIP, gliptins, GLP-1, hyperglycemia treatment, incretins

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 2017, diabetes was responsible for 4 million deaths (which almost half of them 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 422 and 425 million people worldwide currently have diabetes (around 90% of them corresponding to T2DM and half of them undiagnosed) and that this number will increase to 629 million by 2045.[1,2] This high prevalence means that global expenditure on diabetes treatment and related complications was USD727 billion in 2017 which represents an 8% growth since the previous statistics published in 2015 and is expected to rise to USD776 billion by 2045.[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] DPP-IV inhibition has also drawn significant attention in the field of cancer treatment or diagnosis because DPP-IV regulates various biological mechanisms and it has been suggested that it can act either promoting or inhibiting tumor progression depending on cancer cell type and tumor microenvironment.[16] For instance, it has been argued whether DPP-IV must be or must not be administered to cancer patients with diabetes[17–19] whereas other studies support the oral administration of DPP-IV inhibitors in order to either enhance tumor immunotherapy[20] or to slow the growth of different tumors[21–23] or to suppress colorectal cancer lung metastases in mice.[24] Other studies have demonstrated their therapeutic effects on obesity,[25–29] neuropathy,[30] and hepatic[31–37] and renal[38–44] pathologies. Recent studies have also shown that it is possible to design DPP-IV inhibitors with dual bioactivity on other targets involved in cardiovascular diseases like ACE[45] or β-adrenergic receptors.[46,47]

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)[48] and there are many more in different stages of clinical studies.[49] 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).[50]

Virtual screening (VS) has been used over recent years to discover DPP-IV inhibitors in molecular databases,[51–59] with most of the inhibitors identified having bioactivity in the μM 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 suggested 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).[60,61] 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.[45,62–80]

In the last decade, many reviews focusing on DPP-IV inhibitors have been published.[81–91] Some of these address a wide range of topics such as the incretin system and incretin mimetics,[85] DPP-IV inhibitor selectivity and the implications of DPP-IV inhibition,[84] and even structure optimization in the search for chemical stability, selectivity and favorable pharmacokinetic properties.[90] 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.[81–83,86–89,91]

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)[92] 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, and **(b)** explain the most important selectivity cliffs for DPP-IV inhibition in comparison with other related enzymes such as DPP8 and DPP9. At this point it is interesting to remark that, although it is not clear whether selectivity on DPP8/9 can produce *in vivo* toxicity or mortality,[60,61,84,93–95] it is a common goal of published SAR projects on DPP-IV inhibition to look for selective inhibitors and, therefore, this is the reason why selectivity is also considered here. Thus, on the basis of the information deriving from the activity/selectivity cliff analysis, we 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),[96] 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). For instance, these "*minor changes*" would be related to achieve polypharmacology (like has been recently suggested for dual DPP-IV/ACE inhibitors[45] or for β-blockers/DPP-IV inhibitors[46,47]).

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.[97] 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).[88,98–100] 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.[97,101] DPP-IV is secreted as a mature monomer but requires dimerization to undergo normal proteolytic activity.[102]

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.[85,103] 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).[97,100,104] 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.[97,105,106]

On the active site of a protease there are subsites labeled according to the peptide residue that they bind.[107] 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[53,88,103,106,108–111] 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;[112] 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[113,114] and which has been termed either the S_2 extensive subsite[49,89,100,113,115] or the S_3 pocket.[116–118] Nevertheless, in line with most authors[55,75,86,88,90,91,119,120] 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)[121], 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.[121] In the case of Tyr547, two different orientations are clearly shown where the angle between the aromatic groups is around 70º.[122] 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[110], 4N8E[67] and 4N8D[67]). 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.[123]

Two water molecules have recently been identified as being common at the binding site of 92 DPP-IV crystallized structures.[121] 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.[121]

3. COMPARING THE 3D STRUCTURES FOR DPP-IV, DPP8 AND DPP9

Like DPP-IV, DPP8 and DPP9 belong to the prolyl oligopeptidase family.[124] Both are to be found as monomers in the cytoplasm of human blood lymphocytes, pulmonary leucocytes and monocytes.[125] 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.[84,87,109,125] Since two splice variants are described for DPP8 (with a length of 882 and 898 residues respectively),[126] 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),[126,127] but it has not been demonstrated whether both of these are biologically active.[126,127]

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.[116,128–132] 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).[84,124,126,129,132,133] 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)[134] and is available from the ModBase database[135] using access numbers 37577089[136] and 123983020[137] for DPP8 and DPP9 respectively. The second set of homology models was reported by Janardhan & Reddy[132] and built using the A chain of a human DPP-IV structure (PDB code 1X70) as a template.[138] 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).

The main differences between these two sets of homology models and DPP-IV involve the binding site cavity, in particular around the R-loop[129] (see Figure 4), the P_2-loop[129] (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,[87,89,116,132,139,140] Figure 7 shows that there are no major differences between them (irrespective of whether the homology models from the ModBase database[136,137] or from Janardhan & Reddy[132] are used).

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[51–59]; 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.[51] 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[141,142]; 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.[52] 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.[53] 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.[54] After screening the *Potential Drug Target Database*[143] with the TarFisDock server,[144] 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.*[55] used a VS workflow to predict new DPP-IV inhibitors from the SPECS database.[145] It included the following sequential filters: **(a)** a rigid protein-ligand docking with Glide;[141,142] **(b)** a druglikeness filter inspired by the Lipinski rule of 5;[146] and **(c)** a flexible protein-ligand docking with AutoDock v4.0.[147] 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[148] 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.[55] 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.[149] 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.[141,142] Furthermore, on the basis of 8 of these 9 compounds, a common pharmacophore hypothesis was developed using the HipHop utility of Catalyst.[150] 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.[151] 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.[56] First, the 89,165 molecules that were part of this ZINC subset were submitted to an ADME/Tox filter[152] in order to discard molecules that were either potentially toxic or exhibited poor ADME properties. Conformers were then obtained with the help of OMEGA[153,154] 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 $IC_{50} \leq 10$ 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,[155] **(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[156] 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.[157] 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[158] 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[159] (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[160]. 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.[161]

- Al-masri *et al*. used a VS with 2D and 3D filters implemented in a hierarchical cascade to identify new DPP-IV inhibitors.[57] Structure-based pharmacophore models were generated from co-crystallized ligands with potent DPP-IV inhibitory activities using Discovery Studio Visualizer software.[162] 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[163] and an in-house built database) with the help of Catalyst.[150] 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).[164] Next, conformations were built for the remaining compounds with the help of OMEGA[153,154] and the resulting conformer library was used by FRED[165,166] 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.[53]

- Xing *et al.* performed a hierarchical VS via a multistage workflow.[58] A pharmacophore was built using the HypoGen module from Discovery Studio v2.5[167] 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[168] and Glide/GScore[141,142,169]) 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.[59] 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[141,142] to 1RWQ[170] 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[171] 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[138,159,172–185] 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[186] and superposed to crystallized ligands with Cresset's Forge v10.4[187] 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[92,187] 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:[187]

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

where:

(1) if the absolute value of $\Delta activity$ is smaller than the error associated with activity, $\Delta activity$ is considered to be zero. We have considered that differences on activity lower than 0.3 log units are not meaningful (which according to Forge is considered that the errors on the activity data are "*Low*")

(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[187], 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.[51,102,103,124,188] 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.[103] 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.[189] 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[188,190] along with the fact that inhibitors of a different chemical nature establish salt bridges with Glu205/Glu206 and hydrogen bonds with Tyr662[84] have demonstrated that this interaction is essential for DPP-IV activity.[85,103,191] For instance, in the case of sitagliptin, the lack of interaction with these three residues results in a 25-fold loss of activity.[85] 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,[174] S2B[174] and S2E[138]); **(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);[177,181] **(c)** when the relative location of the amino substituent in the compound makes its protonation more difficult (see Figure 8C);[180] 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).[180]

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.[58] 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*".[58] This suggests that their pharmacophore should also contain the positive ionizable/hydrogen bond donor site that allows interaction with the N-terminal recognition region.

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[136,137] or not facing[128,132] 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[128,132] 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.[112] Moreover, SAR studies suggest that an electrostatic intermolecular interaction with the positively charged Arg125 results in an affinity gain for DPP-IV.[138,173,174,179,182] 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;[138,182] **(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),[179] introducing a methoxy substituent (see Figures S3H and

S3K),[182] or changing the configuration of a chiral center (see Figure 9C)[173]]; and **(c)** the use of a pyrazol- 1- ylmethyl substituent (see Figure S3G).[174] Thus, according to the SAR data available for these three strategies,[138,173,174,179,182] 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),[174] 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),[182] 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).[173]

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[136,137] 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[132] 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,[132] 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,[129] 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[138,173,174,179,182] 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**[138]) are selective relative to DPP8/DPP9 (data not shown). Therefore the 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[129,132,136,137] , the one that best describes the structure of the R-loop at DPP8/DPP9 according to the bioactivity data is Rummey & Metz's[129] 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,[92] 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);[183] **(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);[175] 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).[175] More modest improvements of ligand bioactivity can be achieved by using alkyl groups to interact with Phe357. This can be seen in compounds **6e** (see Figure 10C)[180] and **6g** (see Figure S4C),[180] 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);[176] 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).[172] 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).[172] 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.[89,103]

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).[172] Other examples of this are compounds **1**, **26**, **22** and **24** (see Figures S4D and S4E),[138] 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)[129] 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).[129,132,136,137] For instance, depending on the homology model used, there may either be no residue equivalent to Phe357[136,137] or there may be a His (*i.e.,* His434 for DPP8[129] and His424 for DPP9[129,132]) or a Cys (*i.e.,* Cys472 for DPP8).[132] Similarly, there may either be no residue equivalent to Arg358[136,137] or there may be an Asp (*i.e.,* Asp435[129] for DPP8 and Asp425 for DPP9[129,132]) or a Lys (*i.e.,* Lys473 for DPP8).[132] 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,[84,89,103,114,117,132,140] 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[129,132] and Arg358 by Asp in DPP8[129] and DPP9[129,132] ; see Figure 5B). Moreover, a nearby residue (*i.e.,* Ser209; see Figure 5) is replaced in both models by either Asp278 (for DPP8)[132,136] or Asp251 (for DPP9),[132,137] 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)[129] . 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,[112,124] and therefore it is essential for the catalytic activity of the enzyme.[84,112,192] Figure 3F shows that Tyr547 can adopt two different conformations but, in contrast with previous studies,[103,121] 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.[89,103] Indeed steered molecular dynamics simulations have shown that interactions with Tyr547 are important in preventing the inhibitor from leaving the active site[106] (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[193] 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.[194] 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[132,136] or Tyr653[129] for DPP8 and Tyr644[129,132,137] for DPP9), a nearby residue in DPP-IV is mutated from Cys (*i.e.,* Cys551) to Gln (Gln673[132,136] or Gln657[129] in DPP8 and Gln648[129,132,137] in DPP9; see Figure 6). However, while in Janardhan & Reddy's models[132] these residues have been modeled following the same conformation as DPP-IV (Figure 6B), in ModBase models[136,137] 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[195] 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[132,136] or Val658[129] for DPP8 and Val649[129,132,137] 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).[196] As seen above, SAR data suggest that neither the Glu dyad nor Arg125 are involved in selectivity.[138,173,174,179,182] 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

The lipophilic S_1 pocket is considered a crucial molecular anchor point for DPP-IV inhibitors,[103] and the residues that constitute this pocket are conserved among the peptidases DPP-IV, DPP8 and DPP9.[116,132] 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 and S5G);[177,181] 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 S5H).[184] 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 S5H) 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 S5H) 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 S5H). This demonstrates that full occupation of the hydrophobic S_1 pocket plays a role in the determination of DPP-IV activity.[103,185] 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 $Å^3$).[132] Nevertheless, an analysis of the homology models available for DPP8/DPP9[132,136,137] 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).[132] 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.[132] 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.[116,128–130] To our knowledge, few investigations have carried out SAR studies to discern whether or not the S_1 pocket is important for selectivity.[60,139,172,197] 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).[60,139] 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.[180,198] 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),[185] **(b)** the introduction of a methanesulfonyl group (see Figures 12B and S7A),[159] and **(c)** the introduction of a substituent that ends in a carbamoyl group (see Figures 12C and S7B).[180] 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).[159] 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 (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)[180] or between **23a** and **6e** (see Figure S7B)[180] 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[129,132,136,137] suggest that the equivalent residue in DPP8/DPP9 is a Leu residue (*i.e.,* Leu676[132,136] or Leu660[129] in DPP8 and Leu651[129,132,136] 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.[159,180,185,193,198]

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).[177] 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),[177,181] in some cases this may result in a huge decrease in activity (see Figure 13B)[177] 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).

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**[58] 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.
- Compound **MDPI-12398**[59] 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**[53]· 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**[55,149]· belongs to the SPECS database and has been identified by different VS protocols[55,149]· 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,[149]· whereas in the second it is targeted by the hydroxyl group.[55]· 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**[52] 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.*[57] 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**[54] 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),[54] whereas in a new pose that has just been suggested for a very close derivative it additionally interacts with the S_2 extensive pocket.[80] The same compound **4** derivative has also been used as a lead compound to obtain more potent DPP-IV inhibitors.[80] 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**).[80] 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.[80]

- The **NCI0294730** optimal pose is shown in Figure 7 in the original paper.[53] 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,[53] 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).[51] 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**[51] 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).[51] 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**[56] 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.[56]

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.[56,90,91] 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[100] or describing specific molecular recognition interactions from crystal structure data[103] or quantifying electrostatic and hydrophobic interactions with binding site residues[199] (but nevertheless, the approach they use is different to the one used by us). However (and not considered by the three studies cited),[100,103,199] 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.[129,132,136,137] 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).[125]

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).[96]

ACKNOWLEDGMENTS

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.

FIGURE CAPTIONS

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[159] was used to obtain this figure with the help of the Maestro program.[200]

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[136,137] 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[56,90,108,133,201] or the S_1' pocket.[89,202,203] 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.[112] Residues forming the oxyanion hole (*i.e.,* Tyr631 and Tyr547) are shown in italics.

Figure 3

Validated coordinates for the binding site for different DPP-IV chains. Validation was performed either with VHELIBS[204] or by visual comparison with the corresponding electron density map[205]. Only the Arg125 coordinates do not fit well in some electron density map (*i.e.,* 1PFQ,[123] 4KR0[206] and 4L72[207]) 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[200] 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);[159] panel **B** complex with alogliptin (PDB code 3G0B);[196] panel **C** complex with a β-substituted biarylphenylalanine amide inhibitor (PDB code 2FJP);[208] panel **D** complex with the ABT-341 inhibitor (PDB code 2I78);[209] panel **E** superposition of apo chains; and panel **F** superposition of the 64 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[112] and 1PFQ[123]; chains A, C and D for 2I78[209], 2OAG[210] and 2OQI[175]; chains A, B, C and D for 1W1I;[211] chains A and C for 4QZV[212]; chain A for 4KR0[206] and 4L72;[207] and chain B for 2OQV[175]). This figure was obtained with the help of the Maestro program.[200]

Figure 4

This figure compares the R-loop in DPP-IV with the equivalent one in the ModBase[136,137] and Janardhan & Reddy[132] homology models. Panel **A** corresponds to the superposition of the homology models downloaded from ModBase[136,137] relative to DPP-IV (PDB code: 3C45).[159] Panel **B** corresponds to the superposition of the homology models generated by Janardhan & Reddy[132] relative to DPP-IV (PDB code: 1X70).[138] 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.[200]

Figure 5

This figure compares the sequence around the P_2-loop in DPP-IV with the equivalent ones in the ModBase[136,137] and Janardhan & Reddy[132] homology models. Panel **A** corresponds to the superposition of the homology models downloaded from ModBase[136,137] relative to DPP-IV (PDB code: 3C45).[159] Panel **B** corresponds to the superposition of the homology models generated by Janardhan & Reddy[132] relative to DPP-IV (PDB code: 1X70).[138] 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.[200]

Figure 6

This figure compares the residues surrounding Tyr547 and Cys551 in DPP-IV with the equivalent ones in the ModBase[136,137] and Janardhan & Reddy[132] homology models. Panel **A** corresponds to the superposition of the homology models downloaded from ModBase[136,137] relative to DPP-IV (PDB code: 3C45).[159] Panel **B** corresponds to the superposition of the homology models generated by Janardhan & Reddy[132] relative to DPP-IV (PDB code: 1X70).[138] 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.[200]

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[136,137] and Janardhan & Reddy.[132] 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.*[103] (or the equivalent ones in DPP8/DPP9) are underlined]. Panels **C**, **E** and **G** correspond

to the 3C45[159] and ModBase models for DPP8 and DPP9, while panels **D**, **F** and **H** correspond to the 1X70[138] 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.[200]

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.[174,180,181] 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**;[174] in panel **B**, the NH_3^+ group is replaced by H and a positively charged secondary amine is introduced into the adjacent carbon;[181] in panel **C**, the substituent containing the amine group is shortened and the resulting amine is more difficult to protonate at pH = 7;[180] and in panel **D**, the positive NH_3^+ group is replaced by an alcohol group.[180] The ligand orientations are the result of their superposition with co-crystallized ligands from the same or very similar chemical series (*i.e.,* 2IIV[174] for panel **A**, 4A5S[181] for panel **B** and 3OPM[180] 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[187] and MarvinSketch programs.[213]

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.[138,173,179] 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;[179] in panel **B**, a halogen is added in the ortho position of

the phenyl ring;[138] and in panel **C**, a carboxylic acid is placed near Arg125.[173] The ligand orientations are the result of their superposition with co-crystallized ligands from the same or very similar chemical series (*i.e.,* 3KWF[179] for panel **A**, 1X70[138] for panel **B** and 2FJP[208] 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[187] and MarvinSketch programs.[213]

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.[172,175,180] 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**;[172] in panel **C**, a hydrophobic interaction is established with Phe357;[180] and in panel **D**, a π-π interaction is established with Phe357.[175] The blue dashed lines show the π-π interactions between the corresponding ligand and Phe357 and were calculated with the help of Maestro[200] 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[208] for panels **A** and **B**, 3OPM[180] for panel **C** and 2OQV[175] 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[187] and MarvinSketch programs.[213]

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.[181,184] 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[184] (this π-π interaction was calculated with the help of Maestro[200] using default options). In panel **A** a but- 2- yn- 1- yl substituent in the S_1 subsite is replaced by a prenyl substituent,[181] while in panel **B** a monobutyl group is replaced by an m- tolyl group.[184] The ligand orientations are the result of their superposition with co-crystallized ligands from the same or very similar chemical series (*i.e.,* 4A5S[181] for panel **A** and 3KWF[179] 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[187] and MarvinSketch programs.[213]

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.[159,180,185] 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;[185] in panel **B**, a methanesulfonyl replaces a trifluoromethyl group;[159] and in panel **C**, a (carbamoylmethyl)oxidanyl group replaces a chlorine group.[180] 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[196] for panel **A**, 3C45[159] for panel **B** and 3OPM[180] 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[187] and MarvinSketch programs.[213]

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.[177,181] 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;[181] in panel **B**, the extension to the S_2' pocket correlates with a worsening in DPP-IV activity.[177] The ligand orientations are the result of their superposition with co-crystallized ligands from the same or very similar chemical series (*i.e.,* 4A5S[181] for panel **A** and 2RGU[177] 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[187] and MarvinSketch programs.[213]

TABLE CAPTIONS

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.[187]

[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.*[58]

[c] Although this site is set up as a hydrogen bond acceptor by the pharmacophore authors,[58] 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.*[58] 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.*[57] 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.*[53]

[i] It is not clear from Ward *et al.*[51] 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.*,[51] it is not clear to which DPP-IV pocket this pharmacophore site belongs.

PI: positive ionizable feature

NI: negative ionizable feature

HBA: hydrogen bond acceptor

HBD: hydrogen bond donor

HPH: hydrophobic feature

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

[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)

Table 3

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

REFERENCES

1. International Diabetes Federation. *IDF Diabetes Atlas Eighth Edition 2017*. *International Diabetes Federation* (2017).
2. World Health Organization. *Global report on diabetes*. (2016).
3. Gourgari, E., Wilhelm, E. E., Hassanzadeh, H., Aroda, V. R. & Shoulson, I. A comprehensive review of the FDA-approved labels of diabetes drugs: Indications, safety, and emerging cardiovascular safety data. *J. Diabetes Complications* **31,** 1719–1727 (2017).
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 in 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 (2017).
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. Beckenkamp, A., Davies, S., Willig, J. B. & Buffon, A. DPPIV/CD26: a tumor suppressor or a marker of malignancy? *Tumor Biol.* **37,** 7059–7073 (2016).
17. Wang, H. *et al.* NRF2 activation by antioxidant antidiabetic agents accelerates tumor metastasis. *Sci. Transl. Med.* **8,** 334ra51 (2016).
18. Decalf, J., da Silva, R. B., Werneke, S. & Albert, M. L. Comment on 'NRF2 activation by antioxidant antidiabetic agents accelerates tumor metastasis'. *Sci. Transl. Med.* **8,** 349le1-349le1 (2016).

19. Wang, H., de la Vega, M. R., Zhang, D. D., Yu, S. & Zheng, H. Response to comment on 'NRF2 activation by antioxidant antidiabetic agents accelerates tumor metastasis'. *Sci. Transl. Med.* **8,** 349lr1-349lr1 (2016).

20. Barreira da Silva, R. *et al.* Dipeptidylpeptidase 4 inhibition enhances lymphocyte trafficking, improving both naturally occurring tumor immunity and immunotherapy. *Nat. Immunol.* **16,** 850–8 (2015).

21. Adams, S. *et al.* PT-100, a small molecule dipeptidyl peptidase inhibitor, has potent antitumor effects and augments antibody-mediated cytotoxicity via a novel immune mechanism. *Cancer Res.* **64,** 5471–5480 (2004).

22. Walsh, M. P. *et al.* Val-boroPro accelerates T cell priming via modulation of dendritic cell trafficking resulting in complete regression of established murine tumors. *PLoS One* **8,** e58860 (2013).

23. Pennisi, A. *et al.* Inhibitor of DASH proteases affects expression of adhesion molecules in osteoclasts and reduces myeloma growth and bone disease. *Br. J. Haematol.* **145,** 775–787 (2009).

24. Jang, J.-H. *et al.* Suppression of lung metastases by the CD26/DPP4 inhibitor Vildagliptin in mice. *Clin. Exp. Metastasis* **32,** 677–87 (2015).

25. Fukuda-Tsuru, S., Kakimoto, T., Utsumi, H., Kiuchi, S. & Ishii, S. The novel dipeptidyl peptidase-4 inhibitor teneligliptin prevents high-fat diet-induced obesity accompanied with increased energy expenditure in mice. *Eur. J. Pharmacol.* **723**, 207–15 (2014).

26. Kos, K. *et al.* DPP-IV inhibition enhances the antilipolytic action of NPY in human adipose tissue. *Diabetes. Obes. Metab.* **11,** 285–92 (2009).

27. Rosmaninho-Salgado, J. *et al.* Dipeptidyl-peptidase-IV by cleaving neuropeptide Y induces lipid accumulation and PPAR-γ expression. *Peptides* **37,** 49–54 (2012).

28. Yamane, T. *et al.* Improvement of blood glucose levels and obesity in mice given aronia juice by inhibition of dipeptidyl peptidase IV and α-glucosidase. *J. Nutr. Biochem.* **31,** 106–112 (2016).

29. Hansen, H. H. *et al.* The DPP-IV inhibitor linagliptin and GLP-1 induce synergistic effects on body weight loss and appetite suppression in the diet-induced obese rat. *Eur. J. Pharmacol.* **741,** 254–263 (2014).

30. Bianchi, R. *et al.* Beneficial effects of PKF275-055, a novel, selective, orally bioavailable, long-acting dipeptidyl peptidase IV inhibitor in streptozotocin-induced diabetic peripheral neuropathy. *J. Pharmacol. Exp. Ther.* **340,** 64–72 (2012).

31. Balaban, Y. H. *et al.* Dipeptidyl peptidase IV (DDP IV) in NASH patients. *Ann. Hepatol.* **6,** 242–50 (2007).

32. Itou, M., Kawaguchi, T., Taniguchi, E., Oriishi, T. & Sata, M. Dipeptidyl peptidase IV inhibitor improves insulin resistance and steatosis in a refractory nonalcoholic fatty liver disease patient: A case report. *Case Rep. Gastroenterol.* **6,** 538–44 (2012).

33. Jung, Y.-A. *et al.* Sitagliptin attenuates methionine/choline-deficient diet-induced steatohepatitis. *Diabetes Res. Clin. Pract.* **105,** 47–57 (2014).

34. Ohyama, T. *et al.* MK-0626, a selective DPP-4 inhibitor, attenuates hepatic steatosis in ob/ob mice. *World J. Gastroenterol.* **20,** 16227–35 (2014).

35. Shirakawa, J. *et al.* Diet-induced adipose tissue inflammation and liver steatosis are prevented by DPP-4 inhibition in diabetic mice. *Diabetes* **60,** 1246–57 (2011).

36. Klein, T. *et al.* Linagliptin alleviates hepatic steatosis and inflammation in a mouse model of non-alcoholic

steatohepatitis. *Med. Mol. Morphol.* **47,** 137–49 (2014).

37. Shinjo, T. *et al.* DPP-IV inhibitor anagliptin exerts anti-inflammatory effects on macrophages, adipocytes, and mouse livers by suppressing NF-κB activation. *Am. J. Physiol. Endocrinol. Metab.* **309,** E214–E223 (2015).

38. Nistala, R. *et al.* DPP4 inhibition attenuates filtration barrier injury and oxidant stress in the zucker obese rat. *Obesity (Silver Spring).* **22,** 2172–9 (2014).

39. Alter, M. L. *et al.* DPP-4 inhibition on top of angiotensin receptor blockade offers a new therapeutic approach for diabetic nephropathy. *Kidney Blood Press. Res.* **36,** 119–30 (2012).

40. Groop, P.-H. *et al.* Linagliptin lowers albuminuria on top of recommended standard treatment in patients with type 2 diabetes and renal dysfunction. *Diabetes Care* **36,** 3460–8 (2013).

41. Moon, J.-Y. *et al.* The dose-dependent organ-specific effects of a dipeptidyl peptidase-4 inhibitor on cardiovascular complications in a model of type 2 diabetes. *PLoS One* **11,** e0150745 (2016).

42. Terawaki, Y. *et al.* Efficacy of dipeptidyl peptidase-4 inhibitor linagliptin in patients with type 2 diabetes undergoing hemodialysis. *Diabetol. Metab. Syndr.* **7,** 44 (2015).

43. Kim, D. R. *et al.* Ameliorating effect of gemigliptin on renal injury in murine adriamycin-induced nephropathy. *Biomed Res. Int.* **2017,** 7275109 (2017).

44. Choi, S. H., Leem, J. & Lee, I.-K. Protective effects of gemigliptin, a dipeptidyl peptidase-4 inhibitor, against cisplatin-induced nephrotoxicity in mice. *Mediators Inflamm.* **2017,** 4139439 (2017).

45. Sattigeri, J. A. *et al.* Approaches towards the development of chimeric DPP4/ACE inhibitors for treating metabolic syndrome. *Bioorg. Med. Chem. Lett.* **27,** 2313–2318 (2017).

46. Ojeda-Montes, M. J. *et al.* Ephedrine as a lead compound for the development of new DPP-IV inhibitors. *Future Med. Chem.* **9,** 2129–2146 (2017).

47. Li W, Yang C. US20130184322 A1-novel dipeptidyl-peptidase-IV inhibitors. (2013).

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

49. 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).

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

51. 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).

52. 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).

53. 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).

54. 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).

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

56. 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).

57. 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).

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

59. 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).

60. 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).

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

62. 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).

63. 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).

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

65. 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).

66. 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).

67. 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).

68. 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).

69. 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).

70. 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).

71. 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).

72. 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).

73. 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).

74. 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).

75. 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).

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

77. 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).

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

79. 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).

80. Li, S. *et al.* Discovery and rational design of natural-product-derived 2-phenyl-3,4-dihydro-2*H*-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).

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

82. 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).

83. 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).

84. 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).

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

86. 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).

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

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

89. 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).

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

91. 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).

92. Activity Miner; Forge v10.4, Cresset BioMolecular Discovery Ltd., Cambridgeshire, UK. Available at: http://www.cresset-group.com/forge/.

93. Burkey, B. F. *et al.* Adverse effects of dipeptidyl peptidases 8 and 9 inhibition in rodents revisited. *Diabetes, Obes. Metab.* **10,** 1057–1061 (2008).

94. Wu, J. J. *et al.* Biochemistry, pharmacokinetics, and toxicology of a potent and selective DPP8/9 inhibitor. *Biochem. Pharmacol.* **78,** 203–210 (2009).

95. Okondo, M. C. *et al.* DPP8 and DPP9 inhibition induces pro-caspase-1-dependent monocyte and macrophage pyroptosis. *Nat. Chem. Biol.* **13,** 46–53 (2017).

96. 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).

97. 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).

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

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

100. 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).

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

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

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

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

105. 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).

106. 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).

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

108. 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).

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

110. 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).

111. 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).

112. 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).

113. 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).

114. Yoshida, T. *et al.* Discovery and preclinical profile of teneligliptin (3-[(2*S*,4*S*)-4-[4-(3-methyl-1-phenyl-1*H*-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).

115. 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).

116. 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).

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

118. 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).

119. 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).

120. 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).

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

122. 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).

123. 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).

124. Gorrell, M. D. Dipeptidyl peptidase IV and related enzymes in cell biology and liver disorders. *Clin. Sci.*

(Lond). **108,** 277–292 (2005).

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

126. 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).

127. 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).

128. 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).

129. 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).

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

131. 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).

132. 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).

133. 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).

134. 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).

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

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

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

138. Kim, D. *et al.* (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-trifluorophenyl)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).

139. 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).

140. 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).

141. 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).

142. 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).

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

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

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

146. 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).

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

148. 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).

149. 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).

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

151. 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).

152. 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).

153. 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).

154. 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).

155. 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).

156. 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).

157. 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).

158. EON 2.0.1: OpenEye Scientific Software, Inc., Santa Fe, New Mexico, USA. Available at: http://www.eyesopen.com/eon.

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

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

161. 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).

162. Discovery Studio (DS) Visualizer; Accelrys, Inc., San Diego, CA92121, U.S.A. Available at: http://www.accelrys.com.

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

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

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

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

167. Discovery Studio (DS); Accelrys, Inc., San Diego, CA92121, U.S.A. Available at: http://www.accelrys.com.

168. 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).

169. 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).

170. 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).

171. 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).

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

173. 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).

174. 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).

175. 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).

176. 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).

177. 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).

178. 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).

179. 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).

180. 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).

181. 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).

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

183. 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).

184. 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).

185. 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).

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

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

188. 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).

189. 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).

190. 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).

191. 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).

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

193. 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).

194. 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).

195. 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).

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

197. 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).

198. 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).

199. 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).

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

201. 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).

202. 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).

203. 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).

204. 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).

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

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

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

208. 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).

209. 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 inhibitor for the treatment of type 2 diabetes. *J. Med. Chem.* **49,** 6439–6442 (2006).

210. 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).

211. 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).

212. 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).

213. MarvinSketch 14.7.21.0, ChemAxon (2014). Available at:

http://www.chemaxon.com/products/marvin/marvinsketch/.

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[1,†], Aleix Gimeno[1,†], Sarah Tomas-Hernández[1], Adrià Cereto-Massagué[1], Raúl Beltrán-Debón[1], Cristina Valls[1], Miquel Mulero[1], Gerard Pujadas[1,2,*] and Santiago Garcia-Vallvé[1,2]

[1]Research Group in Cheminformatics & Nutrition, Departament de Bioquímica i Biotecnologia, Universitat Rovira i Virgili, Campus de Sescelades, 43007 Tarragona, Catalonia, Spain

[2]EURECAT, TECNIO, CEICS, Avinguda Universitat 1, 43204 Reus, Catalonia, Spain

[†]*Both authors contributed equally to this work*

**Correspondence to:* Gerard Pujadas, Research Group in Cheminformatics & Nutrition, tel: +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 s/n, 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)** 74 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.

Key words: GIP, gliptins, GLP-1, hyperglycemia treatment, incretins

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 2017, diabetes was responsible for 4 million deaths (which almost half of them 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 422 and 425 million people worldwide currently have diabetes (around 90% of them corresponding to T2DM and half of them undiagnosed) and that this number will increase to 629 million by 2045.[1,2] This high prevalence means that global expenditure on diabetes treatment and related complications was USD727 billion in 2017 which represents an 8% growth since the previous statistics published in 2015 and is expected to rise to USD776 billion by 2045.[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] DPP-IV inhibition has also drawn significant attention in the field of cancer treatment or diagnosis because DPP-IV regulates various biological mechanisms and it has been suggested that it can act either promoting or inhibiting tumor progression depending on cancer cell type and tumor microenvironment.[16] For instance, it has been argued whether DPP-IV must be or must not be administered to cancer patients with diabetes[17–19] whereas other studies support the oral administration of DPP-IV inhibitors in order to either enhance tumor immunotherapy[20] or to slow the growth of different tumors[21–23] or to suppress colorectal cancer lung metastases in mice.[24] Other studies have demonstrated their therapeutic effects on obesity,[25–29] neuropathy,[30] and hepatic[31–37] and renal[38–44] pathologies. Recent studies have also shown that it is possible to design DPP-IV inhibitors with dual bioactivity on other targets involved in cardiovascular diseases like ACE[45] or β-adrenergic receptors.[46,47]

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)[48] and there are many more in different stages of clinical studies.[49] 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).[50]

Virtual screening (VS) has been used over recent years to discover DPP-IV inhibitors in molecular databases,[51–59] with most of the inhibitors identified having bioactivity in the μM 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 suggested 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).[60,61] 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.[45,62–80]

In the last decade, many reviews focusing on DPP-IV inhibitors have been published.[81–91] Some of these address a wide range of topics such as the incretin system and incretin mimetics,[85] DPP-IV inhibitor selectivity and the implications of DPP-IV inhibition,[84] and even structure optimization in the search for chemical stability, selectivity and favorable pharmacokinetic properties.[90] 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.[81–83,86–89,91]

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)[92] 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, and **(b)** explain the most important selectivity cliffs for DPP-IV inhibition in comparison with other related enzymes such as DPP8 and DPP9. At this point it is interesting to remark that, although it is not clear whether selectivity on DPP8/9 can produce *in vivo* toxicity or mortality,[60,61,84,93–95] it is a common goal of published SAR projects on DPP-IV inhibition to look for selective inhibitors and, therefore, this is the reason why selectivity is also considered here. Thus, on the basis of the information deriving from the activity/selectivity cliff analysis, we 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),[96] 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). For instance, these "*minor changes*" would be related to achieve polypharmacology (like has been recently suggested for dual DPP-IV/ACE inhibitors[45] or for β-blockers/DPP-IV inhibitors[46,47]).

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.[97] 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).[88,98–100] 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.[97,101] DPP-IV is secreted as a mature monomer but requires dimerization to undergo normal proteolytic activity.[102]

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.[85,103] 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).[97,100,104] 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.[97,105,106]

On the active site of a protease there are subsites labeled according to the peptide residue that they bind.[107] 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[53,88,103,106,108–111] 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;[112] 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[113,114] and which has been termed either the S_2 extensive subsite[49,89,100,113,115] or the S_3 pocket.[116–118] Nevertheless, in line with most authors[55,75,86,88,90,91,119,120] 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)[121], 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.[121] In the case of Tyr547, two different orientations are clearly shown where the angle between the aromatic groups is around 70º.[122] 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[110], 4N8E[67] and 4N8D[67]). 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.[123]

Two water molecules have recently been identified as being common at the binding site of 92 DPP-IV crystallized structures.[121] 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.[121]

3. COMPARING THE 3D STRUCTURES FOR DPP-IV, DPP8 AND DPP9

Like DPP-IV, DPP8 and DPP9 belong to the prolyl oligopeptidase family.[124] Both are to be found as monomers in the cytoplasm of human blood lymphocytes, pulmonary leucocytes and monocytes.[125] 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.[84,87,109,125] Since two splice variants are described for DPP8 (with a length of 882 and 898 residues respectively),[126] 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),[126,127] but it has not been demonstrated whether both of these are biologically active.[126,127]

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.[116,128–132] 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).[84,124,126,129,132,133] 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)[134] and is available from the ModBase database[135] using access numbers 37577089[136] and 123983020[137] for DPP8 and DPP9 respectively. The second set of homology models was reported by Janardhan & Reddy[132] and built using the A chain of a human DPP-IV structure (PDB code 1X70) as a template.[138] 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).

The main differences between these two sets of homology models and DPP-IV involve the binding site cavity, in particular around the R-loop[129] (see Figure 4), the P_2-loop[129] (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,[87,89,116,132,139,140] Figure 7 shows that there are no major differences between them (irrespective of whether the homology models from the ModBase database[136,137] or from Janardhan & Reddy[132] are used).

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[51–59]; 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.[51]; 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[141,142]; 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.[52]; 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.[53]; 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.[54] After screening the *Potential Drug Target Database*[143] with the TarFisDock server,[144] 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.*[55] used a VS workflow to predict new DPP-IV inhibitors from the SPECS database.[145] It included the following sequential filters: **(a)** a rigid protein-ligand docking with Glide;[141,142] **(b)** a druglikeness filter inspired by the Lipinski rule of 5;[146] and **(c)** a flexible protein-ligand docking with AutoDock v4.0.[147] 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[148] 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.[55] 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.[149] 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.[141,142] Furthermore, on the basis of 8 of these 9 compounds, a common pharmacophore hypothesis was developed using the HipHop utility of Catalyst.[150] 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.[151] 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.[56] First, the 89,165 molecules that were part of this ZINC subset were submitted to an ADME/Tox filter[152] in order to discard molecules that were either potentially toxic or exhibited poor ADME properties. Conformers were then obtained with the help of OMEGA[153,154] 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 $IC_{50} \leq 10$ 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,[155] **(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[156] 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.[157] 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[158] 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[159] (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[160]. 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.[161]

- Al-masri *et al*. used a VS with 2D and 3D filters implemented in a hierarchical cascade to identify new DPP-IV inhibitors.[57] Structure-based pharmacophore models were generated from co-crystallized ligands with potent DPP-IV inhibitory activities using Discovery Studio Visualizer software.[162] 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[163] and an in-house built database) with the help of Catalyst.[150] 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).[164] Next, conformations were built for the remaining compounds with the help of OMEGA[153,154] and the resulting conformer library was used by FRED[165,166] 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.[53]

- Xing *et al.* performed a hierarchical VS via a multistage workflow.[58] A pharmacophore was built using the HypoGen module from Discovery Studio v2.5[167] 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[168] and Glide/GScore[141,142,169]) 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.[59] 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[141,142] to 1RWQ[170] 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[171] 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[138,159,172–185] 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[186] and superposed to crystallized ligands with Cresset's Forge v10.4[187] 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[92,187] 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:[187]

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

where:

(1) if the absolute value of $\Delta activity$ is smaller than the error associated with activity, $\Delta activity$ is considered to be zero. We have considered that differences on activity lower than 0.3 log units are not meaningful (which according to Forge is considered that the errors on the activity data are "*Low*")

(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[187], 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.[51,102,103,124,188] 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.[103] 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.[189] 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[188,190] along with the fact that inhibitors of a different chemical nature establish salt bridges with Glu205/Glu206 and hydrogen bonds with Tyr662[84] have demonstrated that this interaction is essential for DPP-IV activity.[85,103,191] For instance, in the case of sitagliptin, the lack of interaction with these three residues results in a 25-fold loss of activity.[85] 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,[174] S2B[174] and S2E[138]); **(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);[177,181] **(c)** when the relative location of the amino substituent in the compound makes its protonation more difficult (see Figure 8C);[180] 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).[180]

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.[58] 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*".[58] This suggests that their pharmacophore should also contain the positive ionizable/hydrogen bond donor site that allows interaction with the N-terminal recognition region.

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[136,137] or not facing[128,132] 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[128,132] 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.[112] Moreover, SAR studies suggest that an electrostatic intermolecular interaction with the positively charged Arg125 results in an affinity gain for DPP-IV.[138,173,174,179,182] 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;[138,182] **(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),[179] introducing a methoxy substituent (see Figures S3H and

S3K),[182] or changing the configuration of a chiral center (see Figure 9C)[173]]; and **(c)** the use of a pyrazol- 1- ylmethyl substituent (see Figure S3G).[174] Thus, according to the SAR data available for these three strategies,[138,173,174,179,182] 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),[174] 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),[182] 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).[173]

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[136,137] 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[132] 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,[132] 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,[129] 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[138,173,174,179,182] 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**[138]) are selective relative to DPP8/DPP9 (data not shown). Therefore the 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[129,132,136,137] , the one that best describes the structure of the R-loop at DPP8/DPP9 according to the bioactivity data is Rummey & Metz's[129] 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,[92] 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);[183] **(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);[175] 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).[175] More modest improvements of ligand bioactivity can be achieved by using alkyl groups to interact with Phe357. This can be seen in compounds **6e** (see Figure 10C)[180] and **6g** (see Figure S4C),[180] 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);[176] 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).[172] 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).[172] 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.[89,103]

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).[172] Other examples of this are compounds **1**, **26**, **22** and **24** (see Figures S4D and S4E),[138] 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)[129] 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).[129,132,136,137] For instance, depending on the homology model used, there may either be no residue equivalent to Phe357[136,137] or there may be a His (*i.e.,* His434 for DPP8[129] and His424 for DPP9[129,132]) or a Cys (*i.e.,* Cys472 for DPP8).[132] Similarly, there may either be no residue equivalent to Arg358[136,137] or there may be an Asp (*i.e.,* Asp435[129] for DPP8 and Asp425 for DPP9[129,132]) or a Lys (*i.e.,* Lys473 for DPP8).[132] 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,[84,89,103,114,117,132,140] 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[129,132] and Arg358 by Asp in DPP8[129] and DPP9[129,132] ; see Figure 5B). Moreover, a nearby residue (*i.e.,* Ser209; see Figure 5) is replaced in both models by either Asp278 (for DPP8)[132,136] or Asp251 (for DPP9),[132,137] 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)[129] . 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,[112,124] and therefore it is essential for the catalytic activity of the enzyme.[84,112,192] Figure 3F shows that Tyr547 can adopt two different conformations but, in contrast with previous studies,[103,121] 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.[89,103] Indeed steered molecular dynamics simulations have shown that interactions with Tyr547 are important in preventing the inhibitor from leaving the active site[106] (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[193] 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.[194] 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[132,136] or Tyr653[129] for DPP8 and Tyr644[129,132,137] for DPP9), a nearby residue in DPP-IV is mutated from Cys (*i.e.,* Cys551) to Gln (Gln673[132,136] or Gln657[129] in DPP8 and Gln648[129,132,137] in DPP9; see Figure 6). However, while in Janardhan & Reddy's models[132] these residues have been modeled following the same conformation as DPP-IV (Figure 6B), in ModBase models[136,137] 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[195] 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[132,136] or Val658[129] for DPP8 and Val649[129,132,137] 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).[196] As seen above, SAR data suggest that neither the Glu dyad nor Arg125 are involved in selectivity.[138,173,174,179,182] 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

The lipophilic S_1 pocket is considered a crucial molecular anchor point for DPP-IV inhibitors,[103] and the residues that constitute this pocket are conserved among the peptidases DPP-IV, DPP8 and DPP9.[116,132] 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 and S5G);[177,181] 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 S5H).[184] 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 S5H) 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 S5H) 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 S5H). This demonstrates that full occupation of the hydrophobic S_1 pocket plays a role in the determination of DPP-IV activity.[103,185] 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 $Å^3$).[132] Nevertheless, an analysis of the homology models available for DPP8/DPP9[132,136,137] 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).[132] 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.[132] 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.[116,128–130] To our knowledge, few investigations have carried out SAR studies to discern whether or not the S_1 pocket is important for selectivity.[60,139,172,197] 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).[60,139] 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.[180,198] 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),[185] **(b)** the introduction of a methanesulfonyl group (see Figures 12B and S7A),[159] and **(c)** the introduction of a substituent that ends in a carbamoyl group (see Figures 12C and S7B).[180] 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).[159] 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 (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)[180] or between **23a** and **6e** (see Figure S7B)[180] 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[129,132,136,137] suggest that the equivalent residue in DPP8/DPP9 is a Leu residue (*i.e.,* Leu676[132,136] or Leu660[129] in DPP8 and Leu651[129,132,136] 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.[159,180,185,193,198]

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).[177] 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),[177,181] in some cases this may result in a huge decrease in activity (see Figure 13B)[177] 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).

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**[58] 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.
- Compound **MDPI-12398**[59] 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**[53]· 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**[55,149]· belongs to the SPECS database and has been identified by different VS protocols[55,149]· 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,[149]· whereas in the second it is targeted by the hydroxyl group.[55]· 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**[52] 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.*[57] 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**[54] 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),[54] whereas in a new pose that has just been suggested for a very close derivative it additionally interacts with the S_2 extensive pocket.[80] The same compound **4** derivative has also been used as a lead compound to obtain more potent DPP-IV inhibitors.[80] 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**).[80] 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.[80]

- The **NCI0294730** optimal pose is shown in Figure 7 in the original paper.[53] 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,[53] 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).[51] 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**[51] 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).[51] 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**[56] 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.[56]

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.[56,90,91] 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[100] or describing specific molecular recognition interactions from crystal structure data[103] or quantifying electrostatic and hydrophobic interactions with binding site residues[199] (but nevertheless, the approach they use is different to the one used by us). However (and not considered by the three studies cited),[100,103,199] 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.[129,132,136,137] 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).[125]

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).[96]

ACKNOWLEDGMENTS

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.

FIGURE CAPTIONS

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[159] was used to obtain this figure with the help of the Maestro program.[200]

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[136,137] 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[56,90,108,133,201] or the S_1' pocket.[89,202,203] 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.[112] Residues forming the oxyanion hole (*i.e.,* Tyr631 and Tyr547) are shown in italics.

Figure 3

Validated coordinates for the binding site for different DPP-IV chains. Validation was performed either with VHELIBS[204] or by visual comparison with the corresponding electron density map[205]. Only the Arg125 coordinates do not fit well in some electron density map (*i.e.,* 1PFQ,[123] 4KR0[206] and 4L72[207]) 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[200] 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);[159] panel **B** complex with alogliptin (PDB code 3G0B);[196] panel **C** complex with a β-substituted biarylphenylalanine amide inhibitor (PDB code 2FJP);[208] panel **D** complex with the ABT-341 inhibitor (PDB code 2I78);[209] panel **E** superposition of apo chains; and panel **F** superposition of the 64 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[112] and 1PFQ[123]; chains A, C and D for 2I78[209], 2OAG[210] and 2OQI[175]; chains A, B, C and D for 1W1I;[211] chains A and C for 4QZV[212]; chain A for 4KR0[206] and 4L72;[207] and chain B for 2OQV[175]). This figure was obtained with the help of the Maestro program.[200]

Figure 4

This figure compares the R-loop in DPP-IV with the equivalent one in the ModBase[136,137] and Janardhan & Reddy[132] homology models. Panel **A** corresponds to the superposition of the homology models downloaded from ModBase[136,137] relative to DPP-IV (PDB code: 3C45).[159] Panel **B** corresponds to the superposition of the homology models generated by Janardhan & Reddy[132] relative to DPP-IV (PDB code: 1X70).[138] 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.[200]

Figure 5

This figure compares the sequence around the P_2-loop in DPP-IV with the equivalent ones in the ModBase[136,137] and Janardhan & Reddy[132] homology models. Panel **A** corresponds to the superposition of the homology models downloaded from ModBase[136,137] relative to DPP-IV (PDB code: 3C45).[159] Panel **B** corresponds to the superposition of the homology models generated by Janardhan & Reddy[132] relative to DPP-IV (PDB code: 1X70).[138] 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.[200]

Figure 6

This figure compares the residues surrounding Tyr547 and Cys551 in DPP-IV with the equivalent ones in the ModBase[136,137] and Janardhan & Reddy[132] homology models. Panel **A** corresponds to the superposition of the homology models downloaded from ModBase[136,137] relative to DPP-IV (PDB code: 3C45).[159] Panel **B** corresponds to the superposition of the homology models generated by Janardhan & Reddy[132] relative to DPP-IV (PDB code: 1X70).[138] 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.[200]

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[136,137] and Janardhan & Reddy.[132] 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.*[103] (or the equivalent ones in DPP8/DPP9) are underlined]. Panels **C**, **E** and **G** correspond

to the 3C45[159] and ModBase models for DPP8 and DPP9, while panels **D**, **F** and **H** correspond to the 1X70[138] 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.[200]

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.[174,180,181] 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**;[174] in panel **B**, the NH_3^+ group is replaced by H and a positively charged secondary amine is introduced into the adjacent carbon;[181] in panel **C**, the substituent containing the amine group is shortened and the resulting amine is more difficult to protonate at pH = 7;[180] and in panel **D**, the positive NH_3^+ group is replaced by an alcohol group.[180] The ligand orientations are the result of their superposition with co-crystallized ligands from the same or very similar chemical series (*i.e.,* 2IIV[174] for panel **A**, 4A5S[181] for panel **B** and 3OPM[180] 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[187] and MarvinSketch programs.[213]

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.[138,173,179] 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;[179] in panel **B**, a halogen is added in the ortho position of

the phenyl ring;[138] and in panel **C**, a carboxylic acid is placed near Arg125.[173] The ligand orientations are the result of their superposition with co-crystallized ligands from the same or very similar chemical series (*i.e.,* 3KWF[179] for panel **A**, 1X70[138] for panel **B** and 2FJP[208] 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[187] and MarvinSketch programs.[213]

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.[172,175,180] 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**;[172] in panel **C**, a hydrophobic interaction is established with Phe357;[180] and in panel **D**, a π-π interaction is established with Phe357.[175] The blue dashed lines show the π-π interactions between the corresponding ligand and Phe357 and were calculated with the help of Maestro[200] 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[208] for panels **A** and **B**, 3OPM[180] for panel **C** and 2OQV[175] 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[187] and MarvinSketch programs.[213]

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.[181,184] 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[184] (this π-π interaction was calculated with the help of Maestro[200] using default options). In panel **A** a but- 2- yn- 1- yl substituent in the S_1 subsite is replaced by a prenyl substituent,[181] while in panel **B** a monobutyl group is replaced by an m- tolyl group.[184] The ligand orientations are the result of their superposition with co-crystallized ligands from the same or very similar chemical series (*i.e.,* 4A5S[181] for panel **A** and 3KWF[179] 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[187] and MarvinSketch programs.[213]

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.[159,180,185] 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;[185] in panel **B**, a methanesulfonyl replaces a trifluoromethyl group;[159] and in panel **C**, a (carbamoylmethyl)oxidanyl group replaces a chlorine group.[180] 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[196] for panel **A**, 3C45[159] for panel **B** and 3OPM[180] 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[187] and MarvinSketch programs.[213]

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.[177,181] 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;[181] in panel **B**, the extension to the S_2' pocket correlates with a worsening in DPP-IV activity.[177] The ligand orientations are the result of their superposition with co-crystallized ligands from the same or very similar chemical series (*i.e.,* 4A5S[181] for panel **A** and 2RGU[177] 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[187] and MarvinSketch programs.[213]

TABLE CAPTIONS

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.[187]

[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.*[58]

[c] Although this site is set up as a hydrogen bond acceptor by the pharmacophore authors,[58] 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.*[58] 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.*[57] 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.*[53]

[i] It is not clear from Ward *et al.*[51] 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.*,[51] it is not clear to which DPP-IV pocket this pharmacophore site belongs.

PI: positive ionizable feature

NI: negative ionizable feature

HBA: hydrogen bond acceptor

HBD: hydrogen bond donor

HPH: hydrophobic feature

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

[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)

Table 3

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

REFERENCES

1. International Diabetes Federation. *IDF Diabetes Atlas Eighth Edition 2017*. *International Diabetes Federation* (2017).
2. World Health Organization. *Global report on diabetes*. (2016).
3. Gourgari, E., Wilhelm, E. E., Hassanzadeh, H., Aroda, V. R. & Shoulson, I. A comprehensive review of the FDA-approved labels of diabetes drugs: Indications, safety, and emerging cardiovascular safety data. *J. Diabetes Complications* **31,** 1719–1727 (2017).
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 in 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 (2017).
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. Beckenkamp, A., Davies, S., Willig, J. B. & Buffon, A. DPPIV/CD26: a tumor suppressor or a marker of malignancy? *Tumor Biol.* **37,** 7059–7073 (2016).
17. Wang, H. *et al.* NRF2 activation by antioxidant antidiabetic agents accelerates tumor metastasis. *Sci. Transl. Med.* **8,** 334ra51 (2016).
18. Decalf, J., da Silva, R. B., Werneke, S. & Albert, M. L. Comment on 'NRF2 activation by antioxidant antidiabetic agents accelerates tumor metastasis'. *Sci. Transl. Med.* **8,** 349le1-349le1 (2016).

19. Wang, H., de la Vega, M. R., Zhang, D. D., Yu, S. & Zheng, H. Response to comment on 'NRF2 activation by antioxidant antidiabetic agents accelerates tumor metastasis'. *Sci. Transl. Med.* **8,** 349lr1-349lr1 (2016).

20. Barreira da Silva, R. *et al.* Dipeptidylpeptidase 4 inhibition enhances lymphocyte trafficking, improving both naturally occurring tumor immunity and immunotherapy. *Nat. Immunol.* **16,** 850–8 (2015).

21. Adams, S. *et al.* PT-100, a small molecule dipeptidyl peptidase inhibitor, has potent antitumor effects and augments antibody-mediated cytotoxicity via a novel immune mechanism. *Cancer Res.* **64,** 5471–5480 (2004).

22. Walsh, M. P. *et al.* Val-boroPro accelerates T cell priming via modulation of dendritic cell trafficking resulting in complete regression of established murine tumors. *PLoS One* **8,** e58860 (2013).

23. Pennisi, A. *et al.* Inhibitor of DASH proteases affects expression of adhesion molecules in osteoclasts and reduces myeloma growth and bone disease. *Br. J. Haematol.* **145,** 775–787 (2009).

24. Jang, J.-H. *et al.* Suppression of lung metastases by the CD26/DPP4 inhibitor Vildagliptin in mice. *Clin. Exp. Metastasis* **32,** 677–87 (2015).

25. Fukuda-Tsuru, S., Kakimoto, T., Utsumi, H., Kiuchi, S. & Ishii, S. The novel dipeptidyl peptidase-4 inhibitor teneligliptin prevents high-fat diet-induced obesity accompanied with increased energy expenditure in mice. *Eur. J. Pharmacol.* **723**, 207–15 (2014).

26. Kos, K. *et al.* DPP-IV inhibition enhances the antilipolytic action of NPY in human adipose tissue. *Diabetes. Obes. Metab.* **11,** 285–92 (2009).

27. Rosmaninho-Salgado, J. *et al.* Dipeptidyl-peptidase-IV by cleaving neuropeptide Y induces lipid accumulation and PPAR-γ expression. *Peptides* **37,** 49–54 (2012).

28. Yamane, T. *et al.* Improvement of blood glucose levels and obesity in mice given aronia juice by inhibition of dipeptidyl peptidase IV and α-glucosidase. *J. Nutr. Biochem.* **31,** 106–112 (2016).

29. Hansen, H. H. *et al.* The DPP-IV inhibitor linagliptin and GLP-1 induce synergistic effects on body weight loss and appetite suppression in the diet-induced obese rat. *Eur. J. Pharmacol.* **741,** 254–263 (2014).

30. Bianchi, R. *et al.* Beneficial effects of PKF275-055, a novel, selective, orally bioavailable, long-acting dipeptidyl peptidase IV inhibitor in streptozotocin-induced diabetic peripheral neuropathy. *J. Pharmacol. Exp. Ther.* **340,** 64–72 (2012).

31. Balaban, Y. H. *et al.* Dipeptidyl peptidase IV (DDP IV) in NASH patients. *Ann. Hepatol.* **6,** 242–50 (2007).

32. Itou, M., Kawaguchi, T., Taniguchi, E., Oriishi, T. & Sata, M. Dipeptidyl peptidase IV inhibitor improves insulin resistance and steatosis in a refractory nonalcoholic fatty liver disease patient: A case report. *Case Rep. Gastroenterol.* **6,** 538–44 (2012).

33. Jung, Y.-A. *et al.* Sitagliptin attenuates methionine/choline-deficient diet-induced steatohepatitis. *Diabetes Res. Clin. Pract.* **105,** 47–57 (2014).

34. Ohyama, T. *et al.* MK-0626, a selective DPP-4 inhibitor, attenuates hepatic steatosis in ob/ob mice. *World J. Gastroenterol.* **20,** 16227–35 (2014).

35. Shirakawa, J. *et al.* Diet-induced adipose tissue inflammation and liver steatosis are prevented by DPP-4 inhibition in diabetic mice. *Diabetes* **60,** 1246–57 (2011).

36. Klein, T. *et al.* Linagliptin alleviates hepatic steatosis and inflammation in a mouse model of non-alcoholic

steatohepatitis. *Med. Mol. Morphol.* **47,** 137–49 (2014).

37. Shinjo, T. *et al.* DPP-IV inhibitor anagliptin exerts anti-inflammatory effects on macrophages, adipocytes, and mouse livers by suppressing NF-κB activation. *Am. J. Physiol. Endocrinol. Metab.* **309,** E214–E223 (2015).

38. Nistala, R. *et al.* DPP4 inhibition attenuates filtration barrier injury and oxidant stress in the zucker obese rat. *Obesity (Silver Spring).* **22,** 2172–9 (2014).

39. Alter, M. L. *et al.* DPP-4 inhibition on top of angiotensin receptor blockade offers a new therapeutic approach for diabetic nephropathy. *Kidney Blood Press. Res.* **36,** 119–30 (2012).

40. Groop, P.-H. *et al.* Linagliptin lowers albuminuria on top of recommended standard treatment in patients with type 2 diabetes and renal dysfunction. *Diabetes Care* **36,** 3460–8 (2013).

41. Moon, J.-Y. *et al.* The dose-dependent organ-specific effects of a dipeptidyl peptidase-4 inhibitor on cardiovascular complications in a model of type 2 diabetes. *PLoS One* **11,** e0150745 (2016).

42. Terawaki, Y. *et al.* Efficacy of dipeptidyl peptidase-4 inhibitor linagliptin in patients with type 2 diabetes undergoing hemodialysis. *Diabetol. Metab. Syndr.* **7,** 44 (2015).

43. Kim, D. R. *et al.* Ameliorating effect of gemigliptin on renal injury in murine adriamycin-induced nephropathy. *Biomed Res. Int.* **2017,** 7275109 (2017).

44. Choi, S. H., Leem, J. & Lee, I.-K. Protective effects of gemigliptin, a dipeptidyl peptidase-4 inhibitor, against cisplatin-induced nephrotoxicity in mice. *Mediators Inflamm.* **2017,** 4139439 (2017).

45. Sattigeri, J. A. *et al.* Approaches towards the development of chimeric DPP4/ACE inhibitors for treating metabolic syndrome. *Bioorg. Med. Chem. Lett.* **27,** 2313–2318 (2017).

46. Ojeda-Montes, M. J. *et al.* Ephedrine as a lead compound for the development of new DPP-IV inhibitors. *Future Med. Chem.* **9,** 2129–2146 (2017).

47. Li W, Yang C. US20130184322 A1-novel dipeptidyl-peptidase-IV inhibitors. (2013).

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

49. 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).

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

51. 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).

52. 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).

53. 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).

54. 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).

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

56. 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).

57. 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).

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

59. 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).

60. 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).

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

62. 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).

63. 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).

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

65. 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).

66. 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).

67. 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).

68. 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).

69. 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).

70. 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).

71. 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).

72. 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).

73. 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).

74. 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).

75. 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).

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

77. 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).

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

79. 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).

80. Li, S. *et al.* Discovery and rational design of natural-product-derived 2-phenyl-3,4-dihydro-2*H*-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).

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

82. 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).

83. 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).

84. 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).

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

86. 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).

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

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

89. 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).

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

91. 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).

92. Activity Miner; Forge v10.4, Cresset BioMolecular Discovery Ltd., Cambridgeshire, UK. Available at: http://www.cresset-group.com/forge/.

93. Burkey, B. F. *et al.* Adverse effects of dipeptidyl peptidases 8 and 9 inhibition in rodents revisited. *Diabetes, Obes. Metab.* **10,** 1057–1061 (2008).

94. Wu, J. J. *et al.* Biochemistry, pharmacokinetics, and toxicology of a potent and selective DPP8/9 inhibitor. *Biochem. Pharmacol.* **78,** 203–210 (2009).

95. Okondo, M. C. *et al.* DPP8 and DPP9 inhibition induces pro-caspase-1-dependent monocyte and macrophage pyroptosis. *Nat. Chem. Biol.* **13,** 46–53 (2017).

96. 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).

97. 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).

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

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

100. 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).

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

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

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

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

105. 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).

106. 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).

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

108. 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).

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

110. 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).

111. 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).

112. 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).

113. 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).

114. Yoshida, T. *et al.* Discovery and preclinical profile of teneligliptin (3-[(2*S*,4*S*)-4-[4-(3-methyl-1-phenyl-1*H*-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).

115. 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).

116. 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).

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

118. 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).

119. 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).

120. 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).

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

122. 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).

123. 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).

124. Gorrell, M. D. Dipeptidyl peptidase IV and related enzymes in cell biology and liver disorders. *Clin. Sci.*

(Lond). **108,** 277–292 (2005).

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

126. 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).

127. 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).

128. 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).

129. 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).

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

131. 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).

132. 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).

133. 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).

134. 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).

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

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

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

138. Kim, D. *et al.* (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-trifluorophenyl)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).

139. 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).

140. 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).

141. 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).

142. 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).

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

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

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

146. 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).

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

148. 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).

149. 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).

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

151. 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).

152. 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).

153. 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).

154. 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).

155. 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).

156. 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).

157. 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).

158. EON 2.0.1: OpenEye Scientific Software, Inc., Santa Fe, New Mexico, USA. Available at: http://www.eyesopen.com/eon.

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

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

161. 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).

162. Discovery Studio (DS) Visualizer; Accelrys, Inc., San Diego, CA92121, U.S.A. Available at: http://www.accelrys.com.

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

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

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

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

167. Discovery Studio (DS); Accelrys, Inc., San Diego, CA92121, U.S.A. Available at: http://www.accelrys.com.

168. 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).

169. 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).

170. 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).

171. 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).

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

173. 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).

174. 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).

175. 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).

176. 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).

177. 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).

178. 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).

179. 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).

180. 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).

181. 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).

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

183. 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).

184. 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).

185. 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).

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

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

188. 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).

189. 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).

190. 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).

191. 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).

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

193. 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).

194. 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).

195. 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).

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

197. 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).

198. 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).

199. 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).

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

201. 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).

202. 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).

203. 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).

204. 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).

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

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

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

208. 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).

209. 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 inhibitor for the treatment of type 2 diabetes. *J. Med. Chem.* **49,** 6439–6442 (2006).

210. 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).

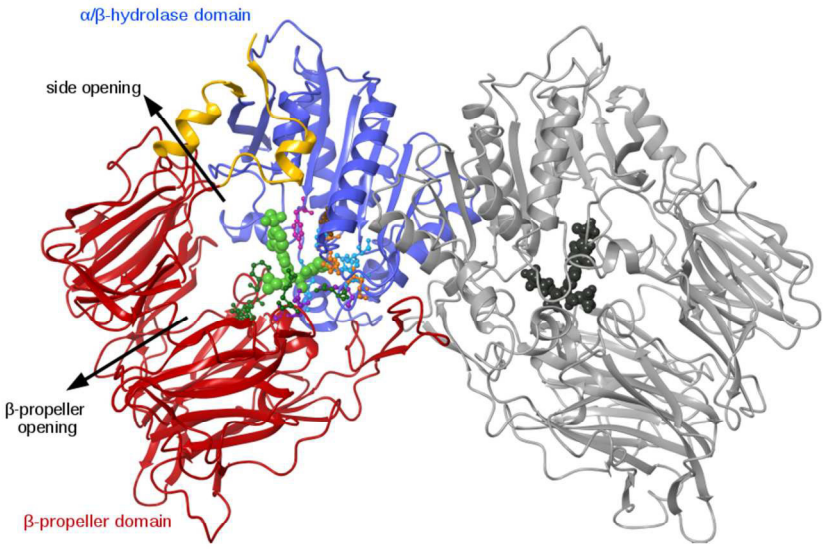
211. Weihofen, W. A., Liu, J., Reutter, W., Saenger, W. & Fan, H. Crystal structure of CD26/dipeptidyl-peptidase IV in complex with adenosine deaminase reveals a highly amphiphilic interface. *J. Biol. Chem.* **279,** 43330–43335 (2004).

212. 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).

213. MarvinSketch 14.7.21.0, ChemAxon (2014). Available at:

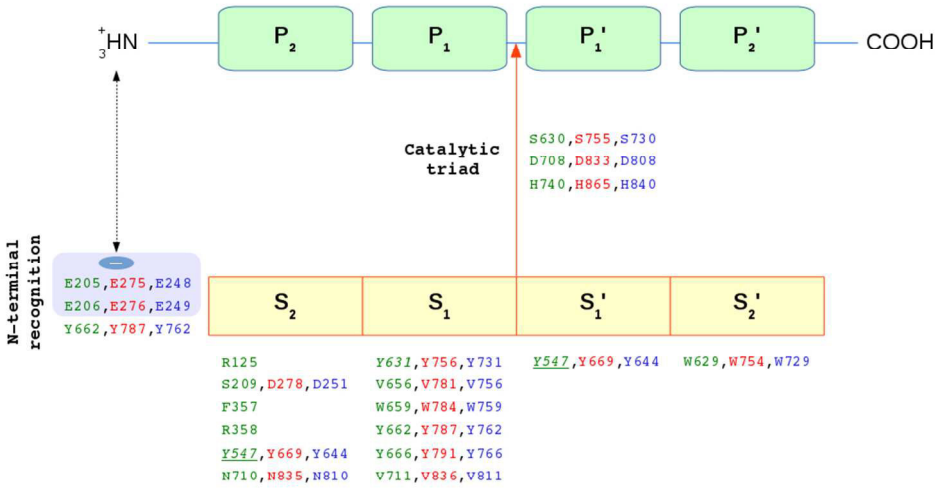
http://www.chemaxon.com/products/marvin/marvinsketch/.

Figure 1



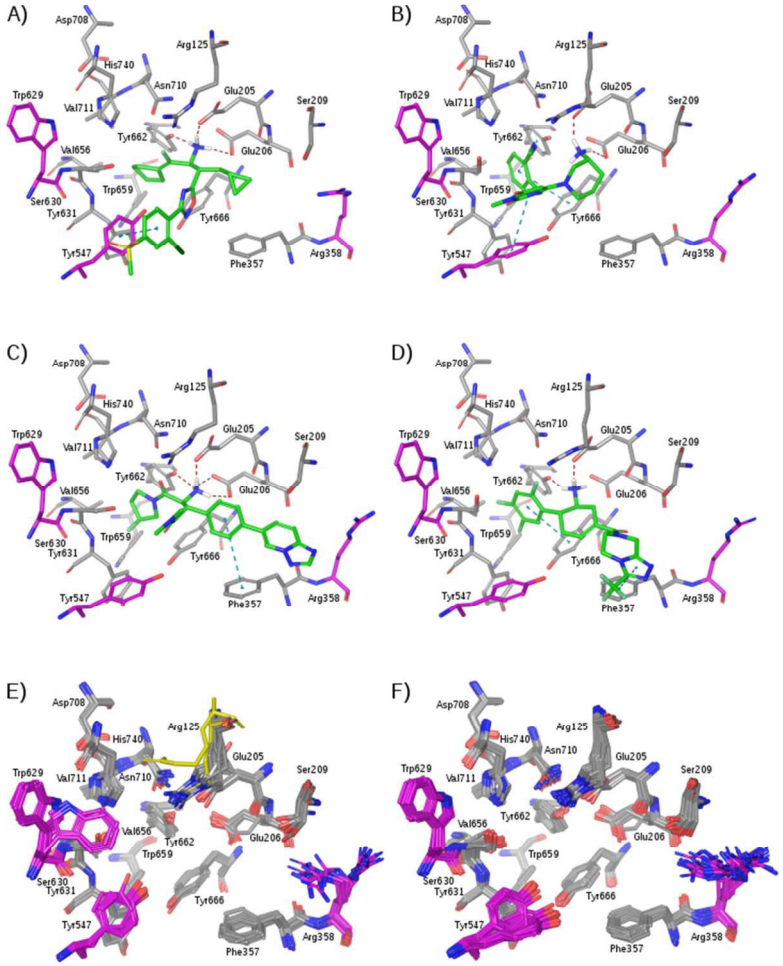
A general overview of the structure of human DPP-IV homodimer. The domain structure for one of the two subunits is also shown (with the β-propeller domain in red, the α/β hydrolase domain in blue and the interdomain region in yellow). The interface between these two domains forms a central cavity which 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 S1 subsite are light blue, those from the S2 subsite are dark green, those from the S1' subsite are pink and those from the S2' 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 3C45159• was used to obtain this figure with the help of the Maestro program [200].

Figure 2



Schematic view of subsite organization in the DPP-IV binding site. The DPP-IV subsites occupied by peptide residues P2, P1, P1', P2' are labeled S2, S1, S1', S2' respectively. The point of cleavage of the peptide substrate is between the bond binding residue P1 with residue P1'. 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 [136,137·] 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 S2 [56,90,108,133,201]· or the S1' pocket [89,202,203·]. The Arg125 and Asn710 at the S2 pocket are essential for coordinating the carbonyl of the P2 residue and, together with Glu205 and Glu206, align the substrate optimally for the nucleophilic attack by Ser630 [112]· Residues forming the oxyanion hole (i.e., Tyr631 and Tyr547) are shown in italics.

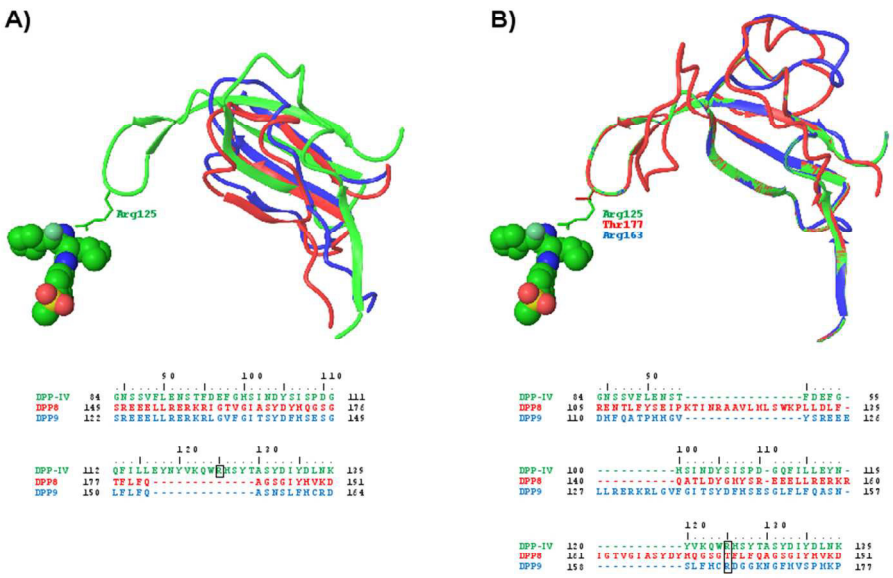
Figure 3



Validated coordinates for the binding site for different DPP-IV chains. Validation was performed either with VHELIBS [204]· or by visual comparison with the corresponding electron density map [205]· . Only the Arg125 coordinates do not fit well in some electron density map (i.e., 1PFQ [123],· 4KR0 [206]· and 4L72 [207·]) 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 [200]· 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) [159];· panel B complex with alogliptin (PDB code 3G0B) [196];· panel C complex with a β-substituted biarylphenylalanine amide inhibitor (PDB code 2FJP) [208·]; panel D complex with the ABT-341 inhibitor (PDB code 2I78) [209];· panel E superposition of apo chains; and panel F superposition of the 64 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 [112]· and 1PFQ [123]· ;

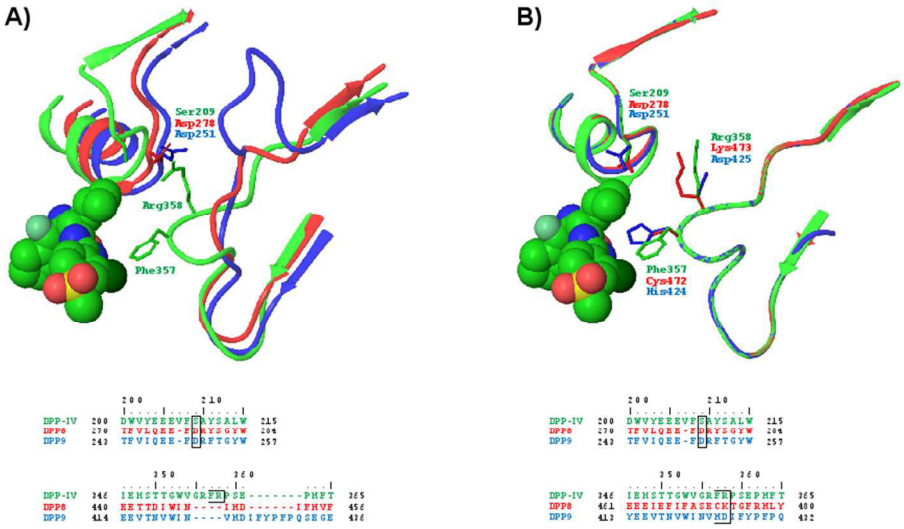
chains A, C and D for 2I78 [209]· , 2OAG [210]· and 2OQI [175]· ; chains A, B, C and D for 1W1I [211];· chains A and C for 4QZV [212]· ; chain A for 4KR0 [206]· and 4L72 [207];· and chain B for 2OQV [175]·). This figure was obtained with the help of the Maestro program [200].

Figure 4



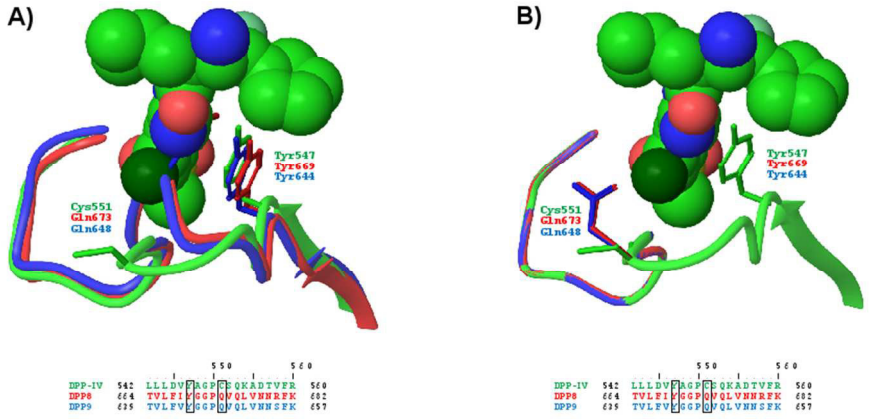
This figure compares the R-loop in DPP-IV with the equivalent one in the ModBase [136,137] and Janardhan & Reddy [132] homology models. Panel A corresponds to the superposition of the homology models downloaded from ModBase [136,137] relative to DPP-IV (PDB code: 3C45) [159]. Panel B corresponds to the superposition of the homology models generated by Janardhan & Reddy [132] relative to DPP-IV (PDB code: 1X70) [138]. 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 [200].

Figure 5



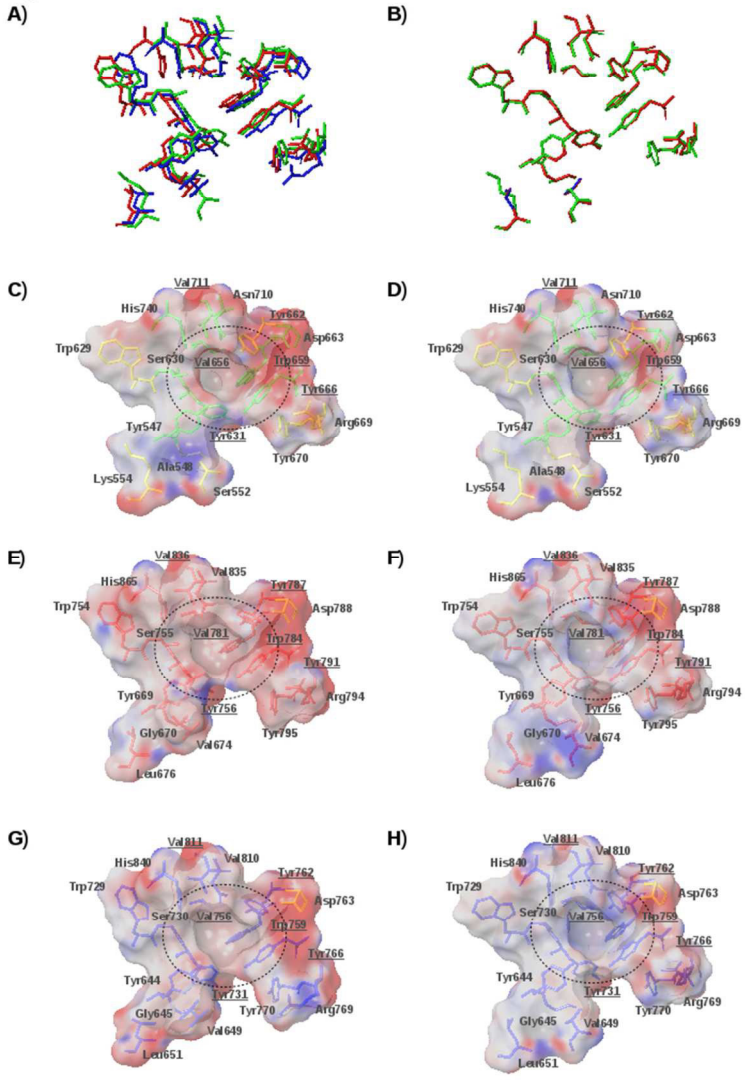
This figure compares the sequence around the P2-loop in DPP-IV with the equivalent ones in the ModBase [136,137] and Janardhan & Reddy [132] homology models. Panel A corresponds to the superposition of the homology models downloaded from ModBase [136,137] relative to DPP-IV (PDB code: 3C45) [159]. Panel B corresponds to the superposition of the homology models generated by Janardhan & Reddy132 relative to DPP-IV (PDB code: 1X70) [138]. 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 [200].

Figure 6



This figure compares the residues surrounding Tyr547 and Cys551 in DPP-IV with the equivalent ones in the ModBase [136,137]· and Janardhan & Reddy [132·] homology models. Panel A corresponds to the superposition of the homology models downloaded from ModBase [136,137]· relative to DPP-IV (PDB code: 3C45) [159].· Panel B corresponds to the superposition of the homology models generated by Janardhan & Reddy [132]· relative to DPP-IV (PDB code: 1X70) [138].· 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 [200·].

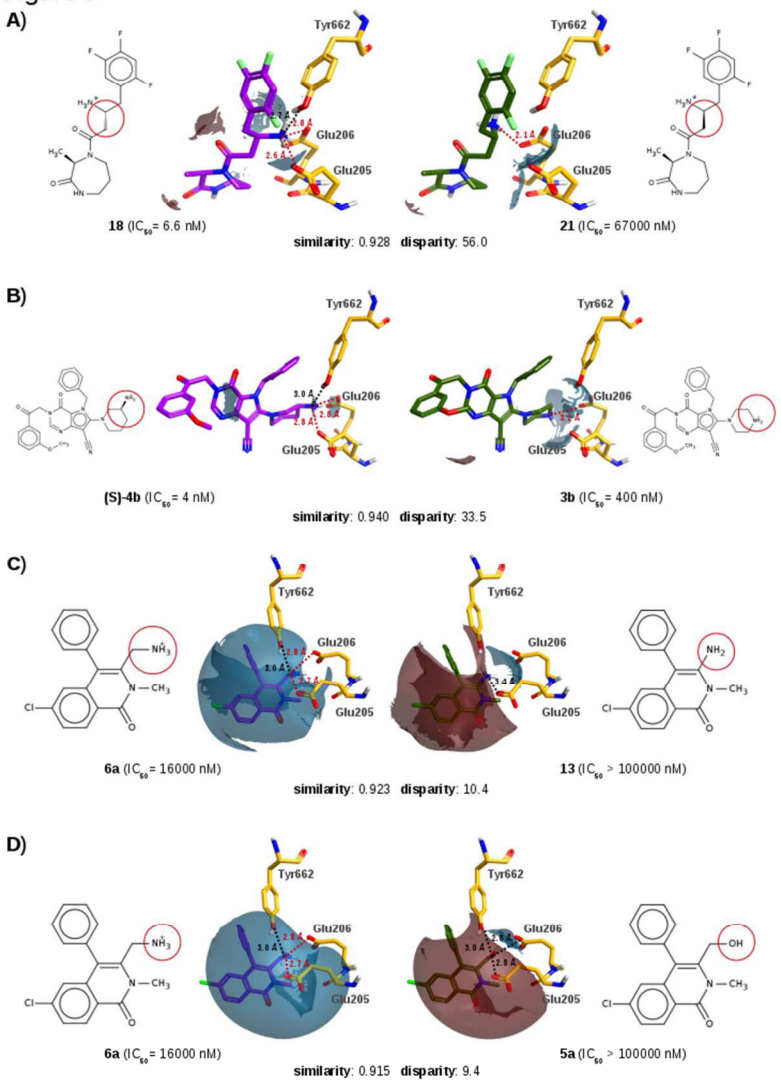
Figure 7



This figure compares the S1 pocket in human DPP-IV with the equivalent pockets in the homology models for DPP8 and DPP9 from ModBase [136,137]· and Janardhan & Reddy [132].· 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 S1 pocket circled for clarity [the DPP-IV residues that form this pocket according to Kuhn et al.103· (or the equivalent ones in DPP8/DPP9) are underlined]. Panels C, E and G correspond to the 3C45 [159]· and ModBase models for DPP8 and DPP9, while panels D, F and H correspond to the 1X70 [138]· and Janardhan & Reddy models for DPP8 and DPP9 respectively. In panels C to H the residues that according to Janardhan & Reddy form the S1 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 S1 pocket (but which are equivalent to other residues that these authors consider part of the S1 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 [200].

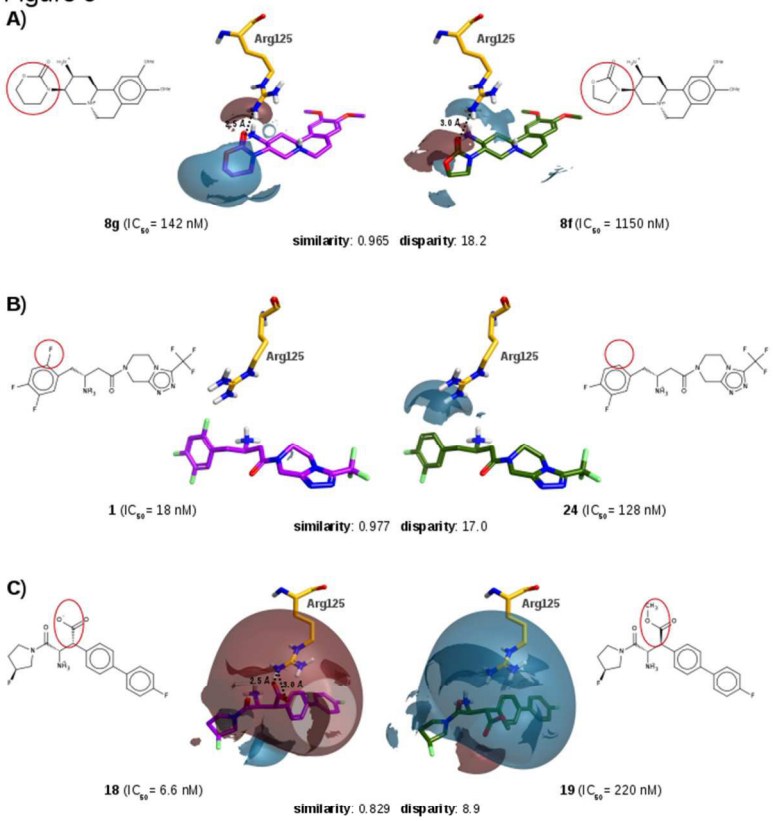
Figure 8



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 [174,180,181]. 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 NH3+ group is switched from (R) in 18 to (S) in 21 [174]; in panel B, the NH3+ group is replaced by H and a positively charged secondary amine is introduced into the adjacent carbon [181]; in panel C, the substituent containing the amine group is

shortened and the resulting amine is more difficult to protonate at pH = 7 [180]; and in panel D, the positive NH3+ group is replaced by an alcohol group.180 The ligand orientations are the result of their superposition with co-crystallized ligands from the same or very similar chemical series (i.e., 2IIV [174] for panel A, 4A5S [181] for panel B and 3OPM [180] 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 [187] and MarvinSketch programs [213].

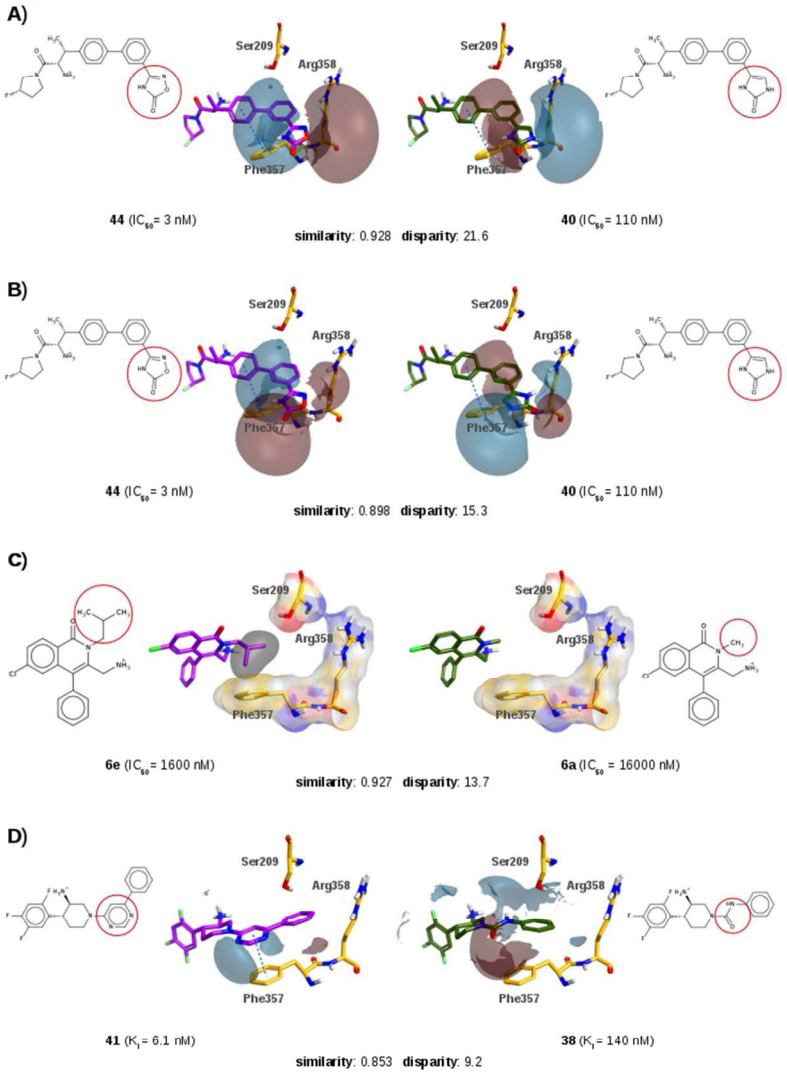
Figure 9



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 [138,173,179]. 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 [179]; in panel B, a halogen is added in the ortho position of the phenyl ring [138]; and in panel C, a carboxylic acid is placed near Arg125.173. The ligand orientations are the result of their superposition with co-crystallized ligands from the same or very similar chemical series (i.e., 3KWF [179] for panel A, 1X70 [138] for panel B and 2FJP [208] 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 [187]· and MarvinSketch programs [213].

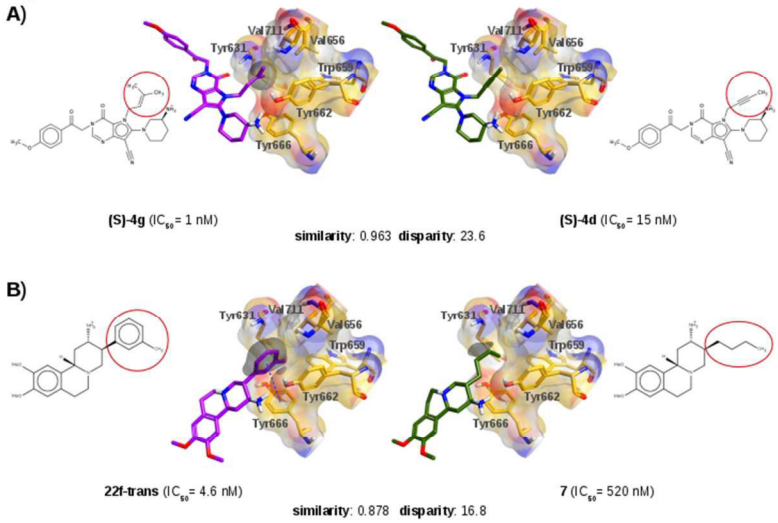
Figure 10



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 [172,175,180].· 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 [172];· in panel C, a hydrophobic interaction is established with Phe357 [180];· and in panel D, a π-π interaction is established with Phe357 [175].· The blue dashed lines show the π-π interactions between the corresponding ligand and

Phe357 and were calculated with the help of Maestro [200]· 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 [208]· for panels A and B, 3OPM [180]· for panel C and 2OQV [175]· 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 [187]· and MarvinSketch programs [213].

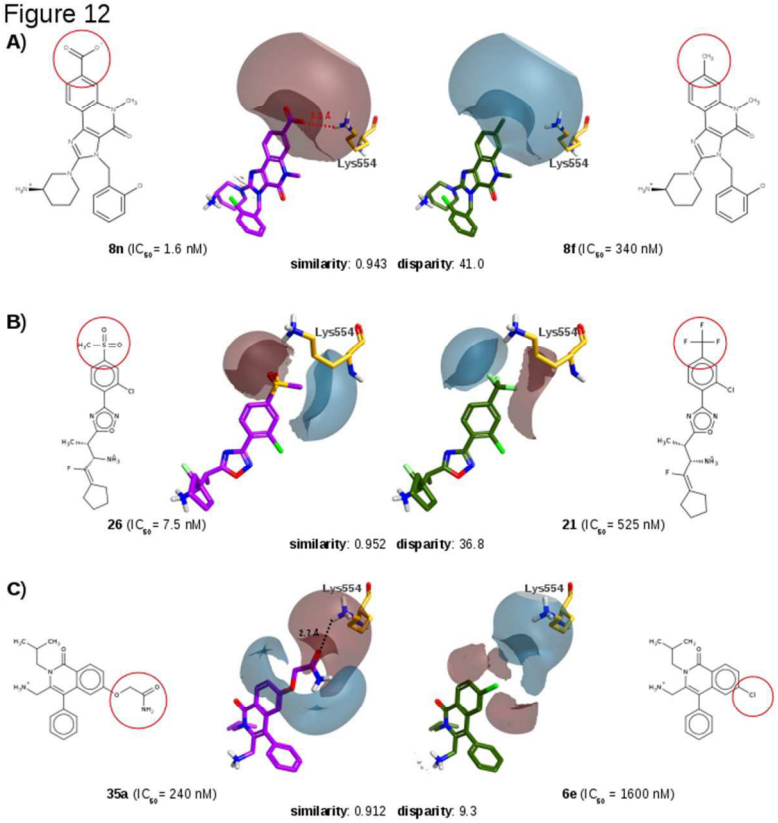
Figure 11



Comparison of the distribution of hydrophobic surfaces between pairs of compounds that differ in their interactions with the S1 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 [181,184]. 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 [184] (this π-π interaction was calculated with the help of Maestro [200] using default options). In panel A a but-2-yn-1-yl substituent in the S1 subsite is replaced by a prenyl substituent [181], while in panel B a monobutyl group is replaced by an m-tolyl group [184]. The ligand orientations are the result of their superposition with co-crystallized ligands from the same or very similar chemical series (i.e., 4A5S [181] for panel A and 3KWF [179] for panel B; residue locations in each panel are also taken from this PDB file).

This figure was obtained with the help of the Forge [187]· and MarvinSketch programs [213].

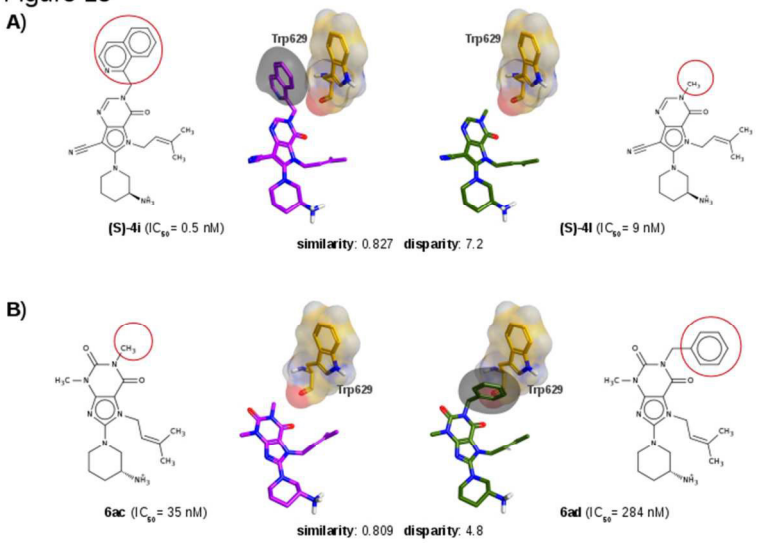
Figure 12



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 [159,180,185].· 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 [185];· in panel B, a methanesulfonyl replaces a trifluoromethyl group [159];· and in panel C, a (carbamoylmethyl)oxidanyl group replaces a chlorine group.180· 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 [196]· for panel A, 3C45 [159]· for panel B and 3OPM [180]· 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 [187]· and MarvinSketch programs [213].

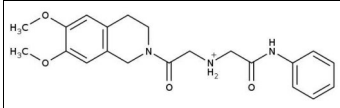
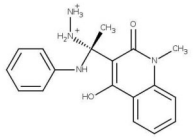
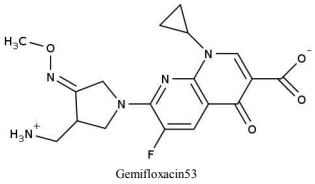
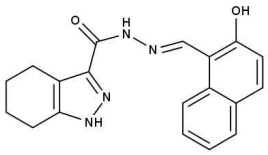
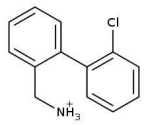
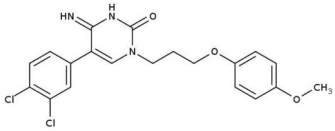
Figure 13



Comparison of the distribution of hydrophobic surfaces between pairs of compounds that differ in their interactions with the S2' 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 [177,181].· 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 S2' pocket correlates with an improvement in DPP-IV activity [181];· in panel B, the extension to the S2' pocket correlates with a worsening in DPP-IV activity [177]. The ligand orientations are the result of their superposition with co-crystallized ligands from the same or very similar chemical series (i.e., 4A5S [181]· for panel A and 2RGU [177]· 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 [187·] and MarvinSketch

programs [213].·

Table 1

Structure and activity data for the most active VS hits	Main VS filter[a]							
	N-terminal recognition region	S_2 pocket					S_1 pocket	Others
		Arg125	Ser209	Phe357	Arg358	Tyr547		
 HWL-89258, b IC_{50} for DPP-IV = 0.148 μM	**HBA**[c]			**HPH**	**HBA**		**HPH**	**HPH**[d]
 **MDPI-12398**[59] 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)							
 Gemifloxacin53 IC_{50} for DPP-IV = 1.12 μM	**PI**	**HBA**	**HBA**					**HBA** (Gln553)
 **Compound 1**[55,149] IC_{50} for DPP-IV = 2.12 μM and 5.77 μM	**HBD**			**HPH**			**HPH**	**HBA**[e] (Arg669, His740)
 Compound **7a**[52] 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**[57,f] 40% of DPP-IV inhibition at 10 μM	**PI**					**HPH/ aromatic**	**HPH**	**HBA** (Gln553)

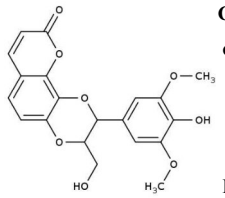
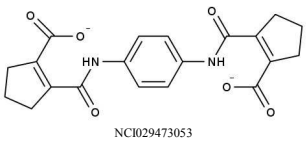
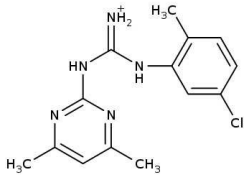
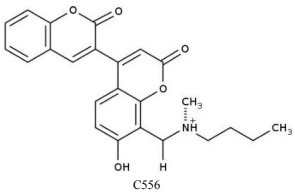
 **Compound 4**[54] IC_{50} for DPP-IV = 14.13 μM	Reverse docking using the TarFisDock. Ligand interactions: HBA (Arg125), HPH (Tyr547), HBA (Tyr547 and Ser630), HPH (S_1 pocket)						
 NCI029473053 34% of DPP-IV inhibition at 10 μM		**NI, HBA**[g]	**HPH**		**HBA, HPH**		**HBA**[g] (Trp629)[h]
 **Compound 14**[51,i] 81.9% of DPP-IV inhibition at 30μM	**PI**	**HBA**				**HPH**	
	PI					**HPH**	**HPH**[j]
 C556 IC_{50} for DPP-IV = 61.55 μM	**PI/HBD**	HBA	HPH/ aromatic	HBA	HPH/ aromatic	**HPH/ aromatic**	HPH/ aromatic (Trp629/ Ser630)

Table 2

Bibliographic reference or ModBase database entry	DPP8[a]	DPP9[a]
Pitman *et al.* (2006)[128]	1N1M (hDPP-IV; 51%) 1XFD (hDPP6; 52%)	Not modelled
Rummey & Metz, (2007)[129]	2BGR (chain A; hDPP-IV; 37%)	2BGR (chain A; hDPP-IV; 37%)
Kang *et al.* (2007)[116]	Data not available	Data not available
Yazbeck *et al.* (2009)[131]	1R9N (hDPP-IV; 55%) 1Z68 (hFAP) 1XFD (hDPP6)	Not modelled
Park *et al.* (2008)[130]	1N1M (hDPP-IV) 1Z68 (hFAP) 1XFD (hDPP6)	1N1M (hDPP-IV) 1Z68 (hFAP) 1XFD (hDPP6)
Janardhan & Reddy (2011)[132]	1X70 (chain A; hDPP-IV; 43%)	1X70 (chain A; hDPP-IV; 43%)
37577089[135, 136]	2ECF (chain A; sDPP-IV; 33%[b])	
123983020[135, 137]		2ECF (chain A; sDPP-IV; 34%[b])

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.

Supplementary material for

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

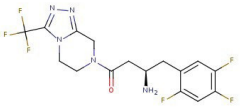
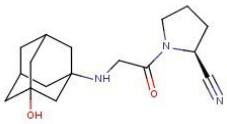
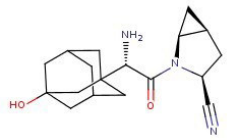
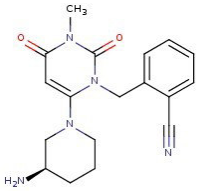
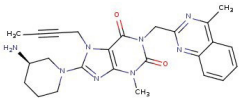
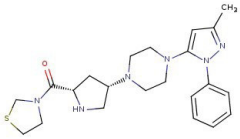
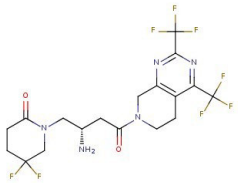
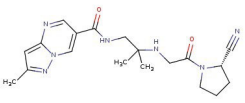
María José Ojeda-Montes[1,†], Aleix Gimeno[1,†], Sarah Tomas-Hernández[1], Adrià Cereto-Massagué[1], Raúl Beltrán-Debón[1], Cristina Valls[1], Miquel Mulero[1], Gerard Pujadas[1,2,*] and Santiago Garcia-Vallvé[1,2]

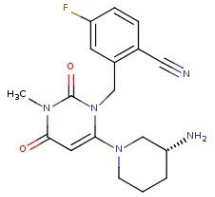
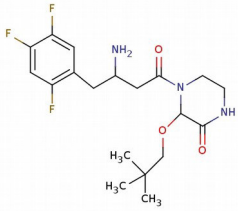
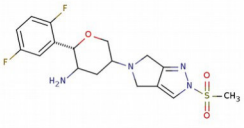
[1]Research Group in Cheminformatics & Nutrition, Departament de Bioquímica i Biotecnologia, Universitat Rovira i Virgili, Campus de Sescelades, 43007 Tarragona, Catalonia, Spain
[2]EURECAT, TECNIO, CEICS, Avinguda Universitat 1, 43204 Reus, Catalonia, Spain

[†]*Both authors contributed equally to this work*

**Correspondence to:* Gerard Pujadas, Research Group in Cheminformatics & Nutrition, tel: +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 s/n, Edifici N4, 43007 Tarragona, Catalonia, Spain. E-mail: gerard.pujadas@urv.cat

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, saxagliptin and anagliptin covalently bind to DPP-IV through Ser630.

Table S2. Summary of the PDB entries for human DPP-IV (updated on 23.1.2018)

Apo form (10)

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

DPP-IV/inhibitor complexes (90)

Complex with a ligand of oligopeptide nature (5)

1NU8†, 1R9N†, 1WCY†,2BGN, 2BGR

Complex with a ligand of drug-like nature (85)

Covalently bounded to DPP-IV through Ser630 (11)

1TKR, 2AJL, 2G5P, 2G5T, 2G63, 2I03, 2QKY†, 3BJM[1], 3W2T[2], 6B1E[2], 6B1O

Non-covalently bounded to DPP-IV (74)

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 (64)

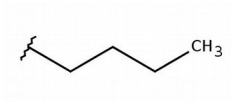
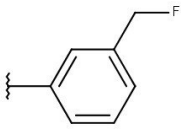
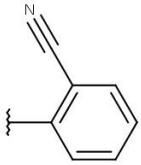
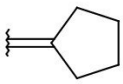
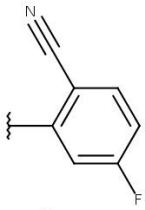
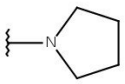
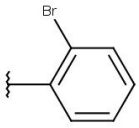
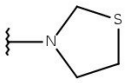
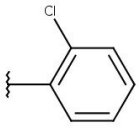
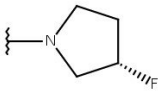
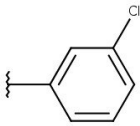
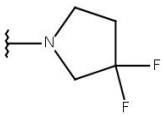
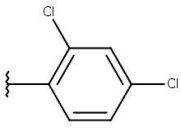
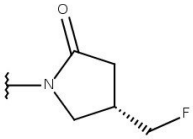
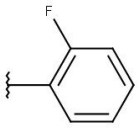
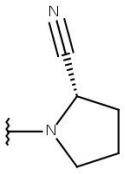
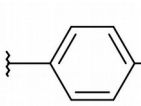
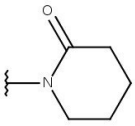
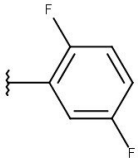
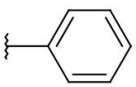
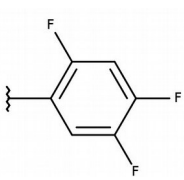
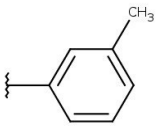
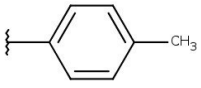
1N1M, **1X70**[4], **2FJP**, 2HHA, 2I78, **2IIT**, **2IIV**, 2OAG, 2OGZ, 2OLE, 2ONC, 2OPH, 2OQI, 2OQV, 2P8S, **2QJR**, 2QOE, **2QT9**, **2QTB**, **2RGU**[5], 2RIP, 3C43, **3C45**, 3CCB, **3CCC**, 3D4L, 3EIO, 3F8S, **3G0B**[6], 3G0C, **3G0D**, **3G0G**, **3H0C**, **3HAB**, **3HAC**, **3KWF**[7], **3KWJ**, 3NOX, **3O95**, 3O9V, 3OPM, 3Q0T, 3QBJ, 3SWW, 3SX4, **3VJL**, **3VJM**, **3WQH**[8], 4A5S, 4DTC, 4G1F, 4JH0, 4LKO, 4N8D, 4N8E, **4PNZ**[9], **5I7U**, **5ISM**, **5KBY**[10], **5J3J**, **5T4B**, 5T4E, **5T4F**, **5T4H**

Complexes with drug-like inhibitors with $IC_{50} \leq 10$ 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. Although in 3WQH, anagliptin is not covalently bounded to Ser630, the dipole interaction of the cyano group with the S_1 subsite is considered to be a transition state before the formation of a covalent imidate (like the one found in vildagliptin and saxagliptin) that explains anagliptin potency and why a major metabolite of this gliptin is a hydrolyzed compound of the cyano group.
[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 2OQV 3Q8W 2I78 2P8S 4DTC 2IIT 2QJR 4DSA 2IIV 2QOE 4DSZ 2OAG 3EIO 4J3J 2OLE 3HAB 4PNZ 2OQI 3HAC 5J3J
	2JID		
	3O9V 3O95		

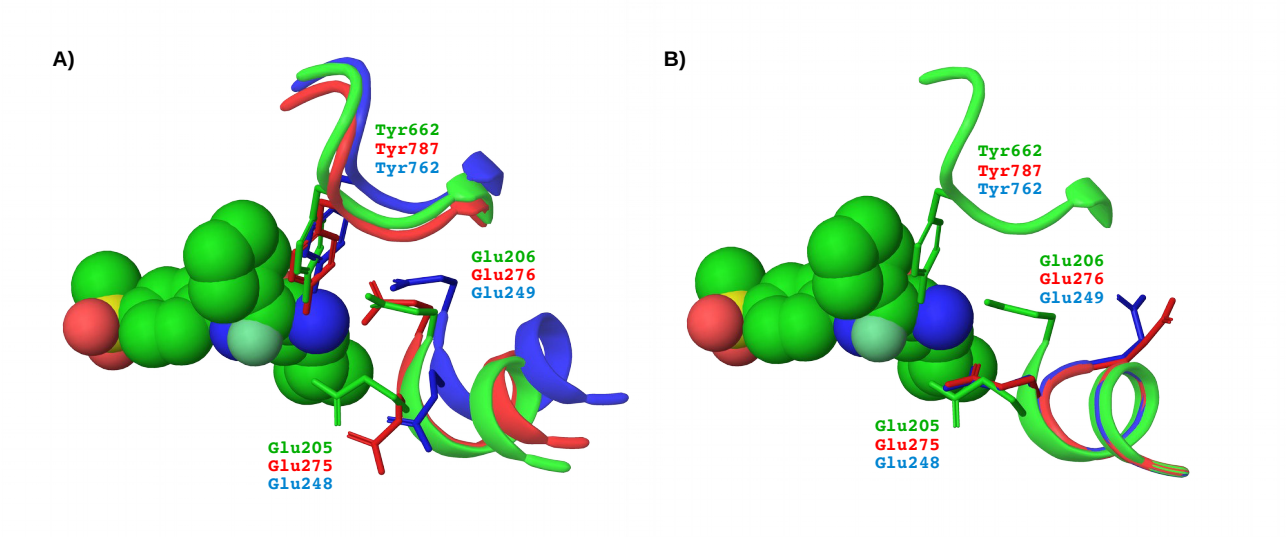
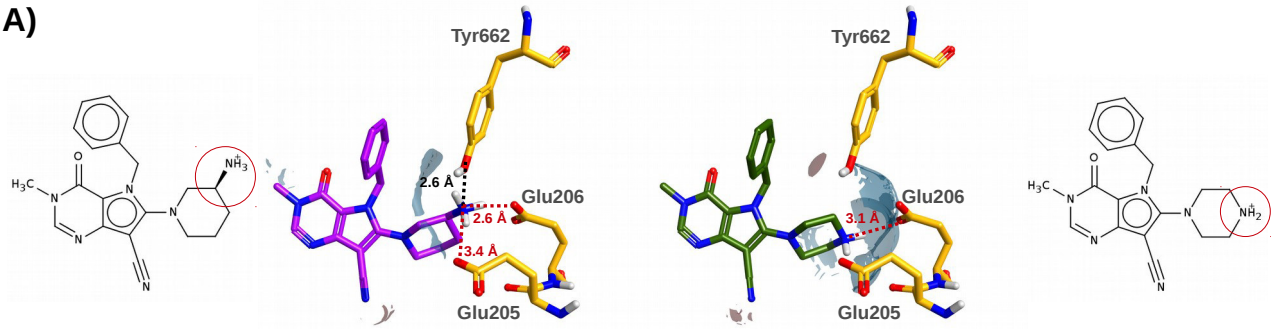


Figure S1. This figure compares the N-terminal recognition region in DPP-IV with the equivalent region in the ModBase[136,137] and Janardhan & Reddy[132] homology models for DPP8 and DPP9. Panel **A** corresponds to the superposition of the homology models downloaded from ModBase[136,137] relative to DPP-IV (PDB code: 3C45).[159] Panel **B** corresponds to the superposition of the homology models generated by Janardhan & Reddy[132] relative to DPP-IV (PDB code: 1X70).[138] 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[159] 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.[200]

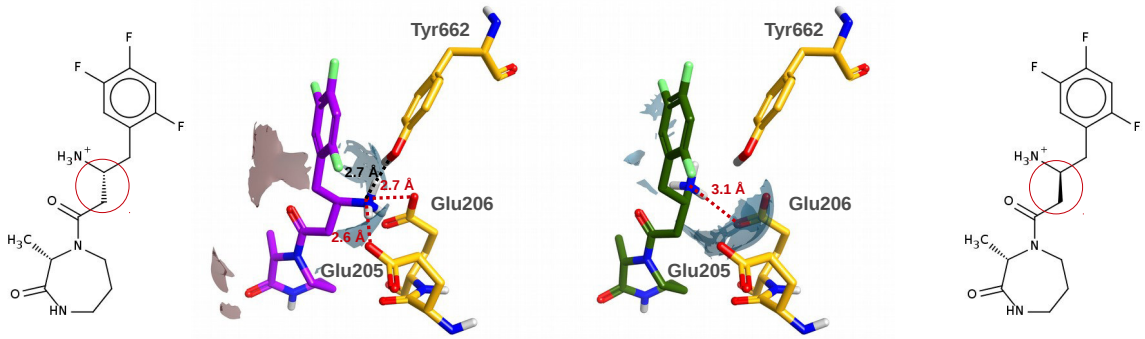
A)



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

similarity: 0.965 **disparity**: 40.4

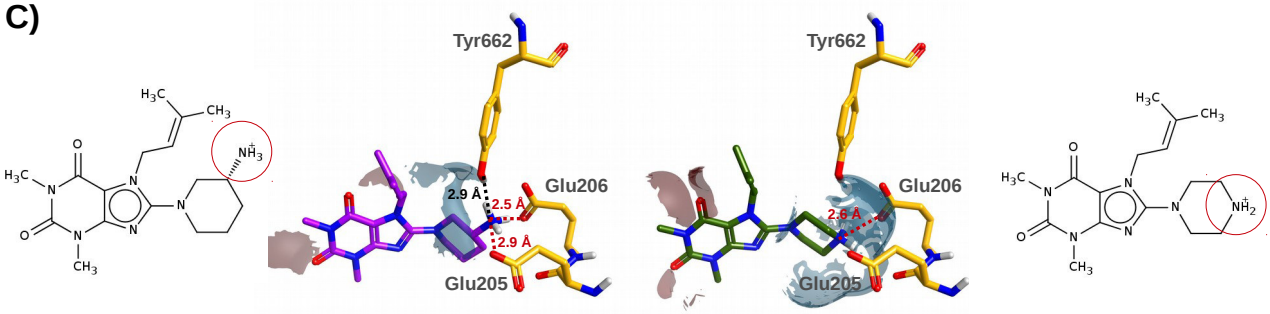
B)



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

similarity: 0.928 **disparity**: 16.3

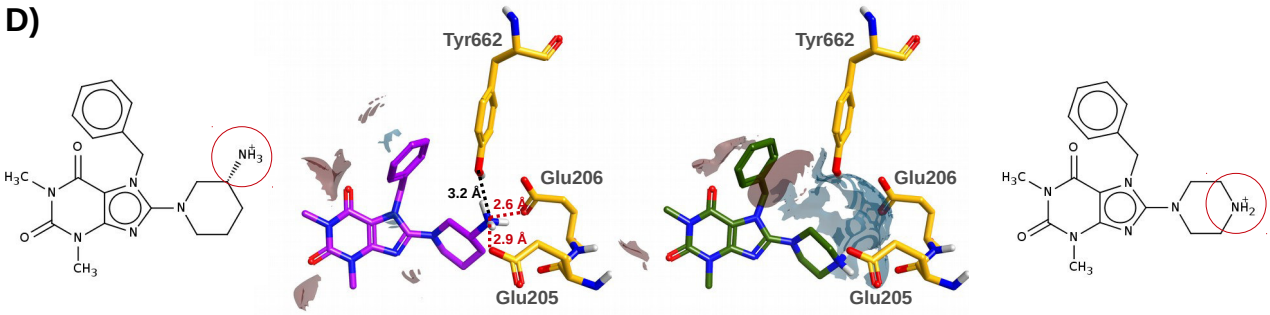
C)



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

similarity: 0.920 **disparity**: 15.2

D)



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

similarity: 0.898 **disparity**: 15.1

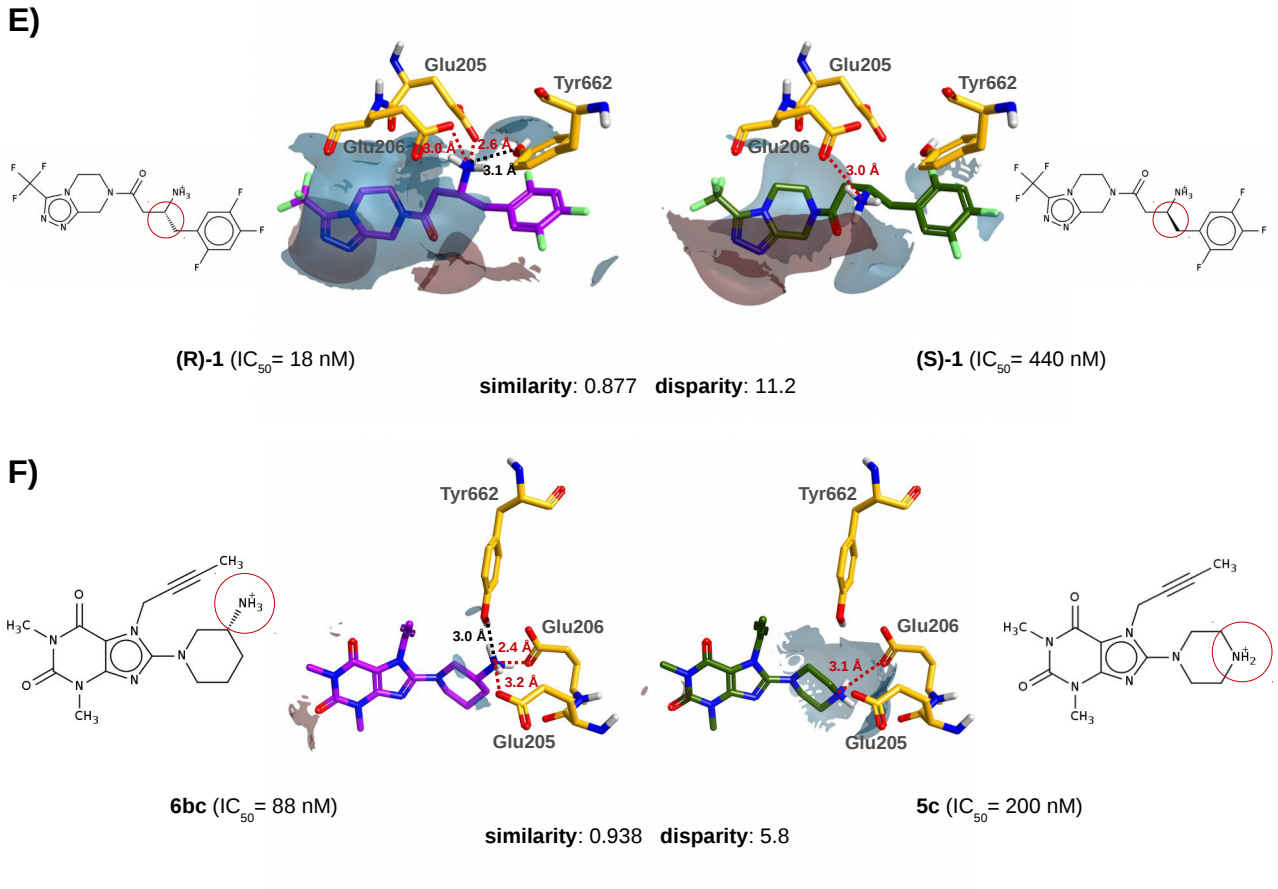
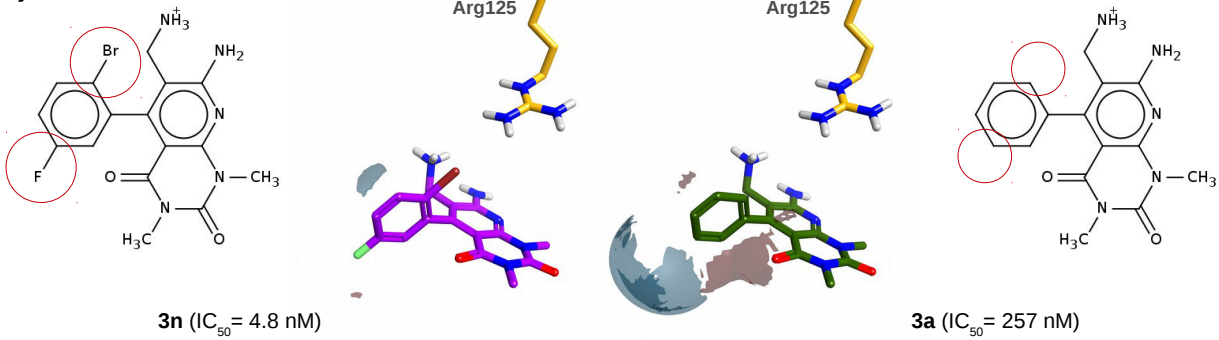


Figure S2. Comparison of the distribution of electrostatic surfaces between pairs of compounds that differ in their interactions with residues Glu205, Glu206 and Tyr662. For each panel the compound with the highest activity is shown in purple on the left and the compound with the lowest activity is shown in green on the right. Molecules are labeled with the same names that identify them in the corresponding paper.[138,174,177,181] 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;[177,181] 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).[138,174] The ligand orientations are the result of their superposition with co-crystallized ligands from the same or very similar chemical series (*i.e.*, 4A5S[181] for panel **A**, 2IIV[174] for panel **B**, 1X70[138] for panel **E** and 2RGU[177] 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[187] and MarvinSketch programs.[213]

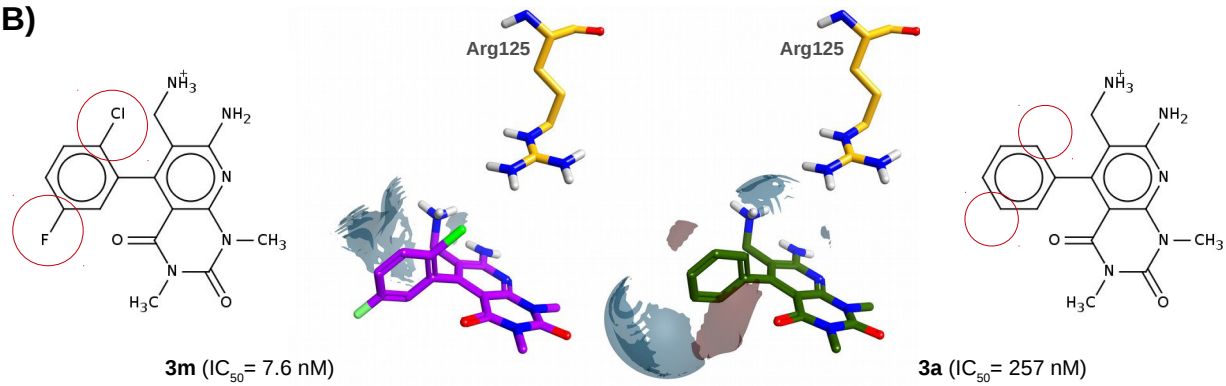
A)



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

similarity: 0.925 **disparity**: 23.1

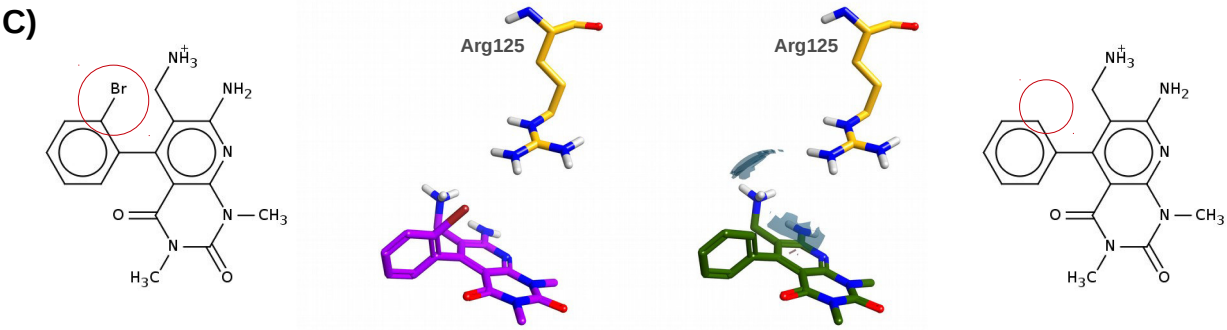
B)



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

similarity: 0.931 **disparity**: 22.1

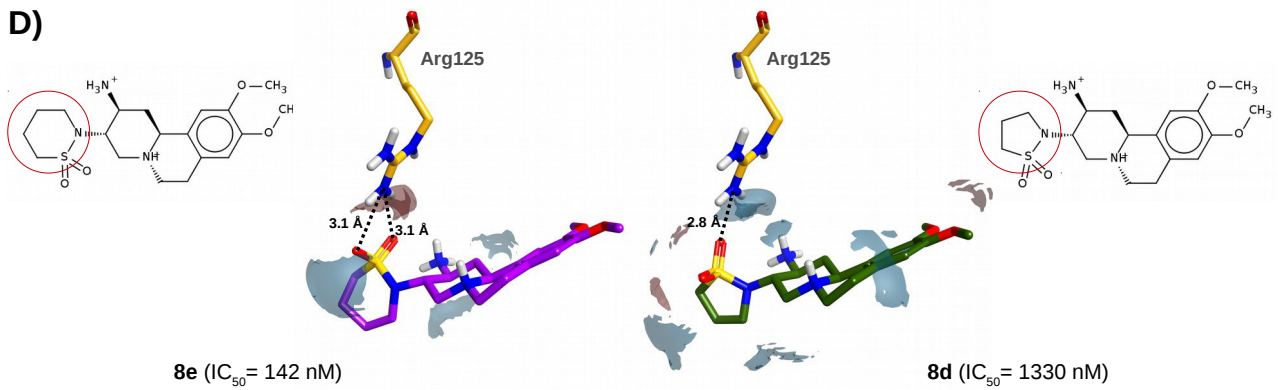
C)



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

similarity: 0.943 **disparity**: 20.8

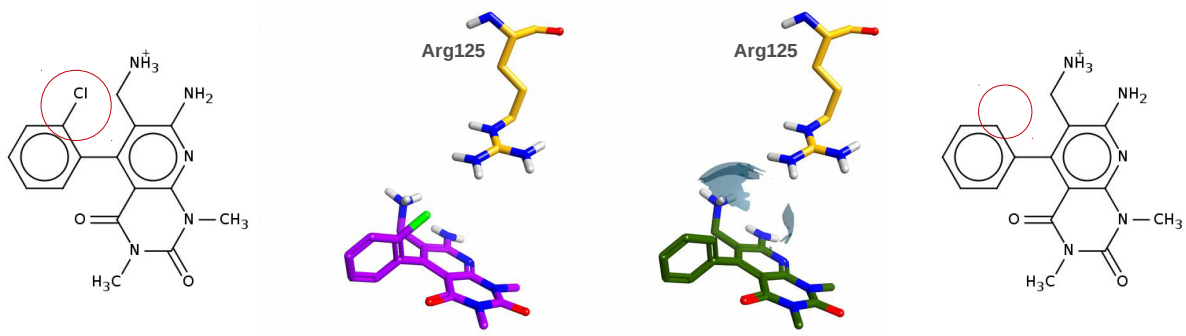
D)



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

similarity: 0.948 **disparity**: 18.8

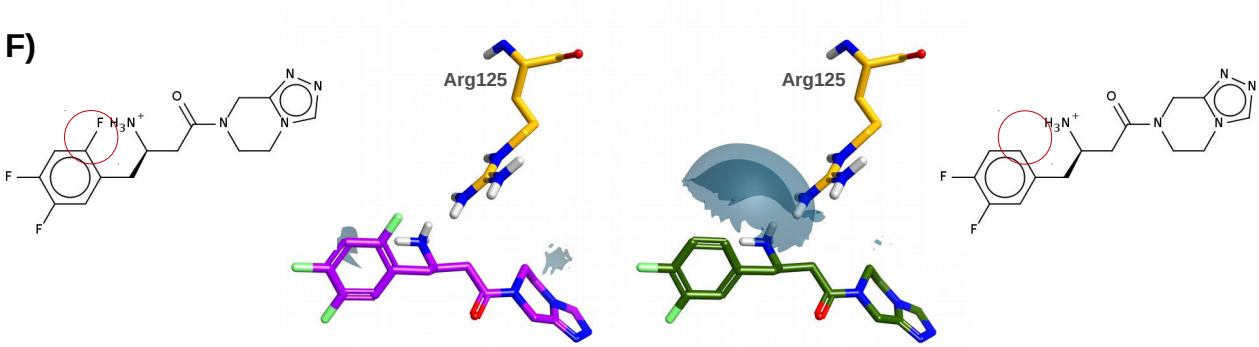
E)



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

similarity: 0.952 **disparity**: 17.8

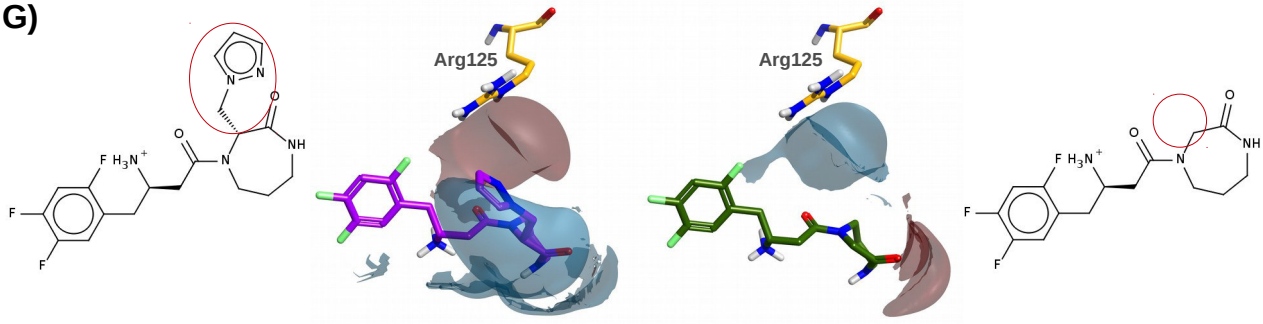
F)



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

similarity: 0.977 **disparity**: 16.6

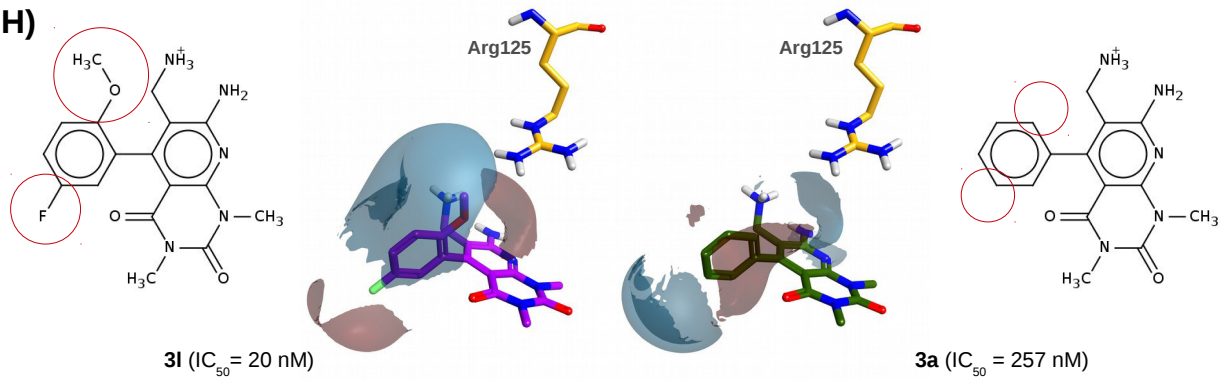
G)



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

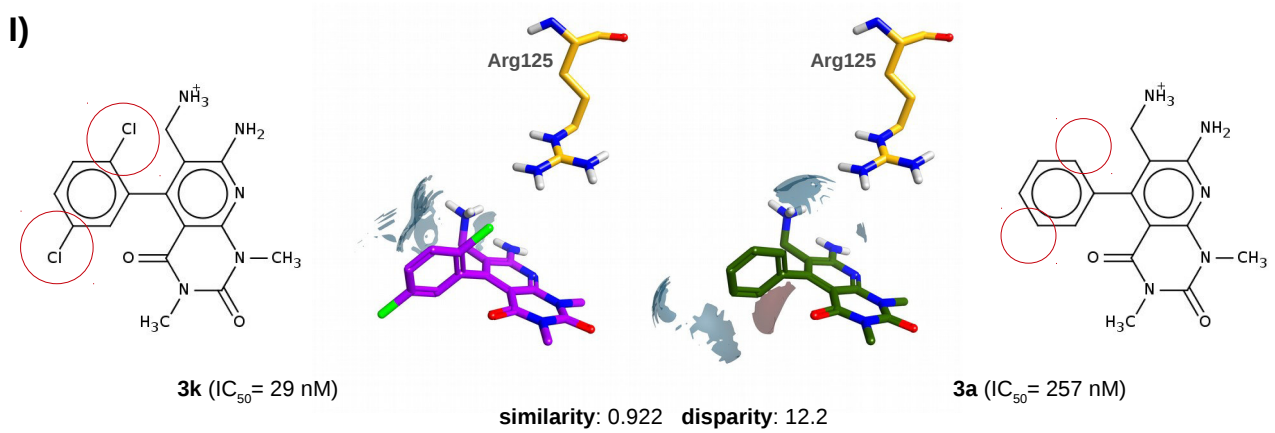
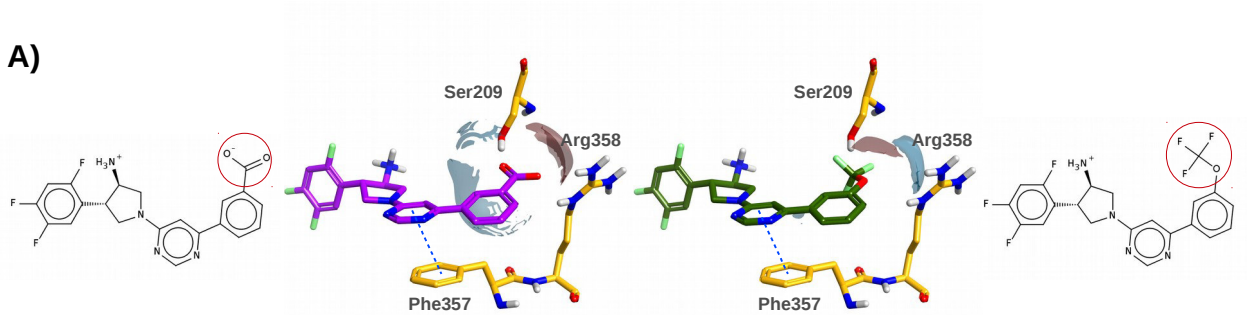


Figure S3. Comparison of the distribution of electrostatic surfaces between pairs of compounds that differ in their interactions with residue Arg125. For each panel the compound with the highest activity is shown in purple on the left and the compound with the lowest activity is shown in green on the right. Molecules are labeled with the same names that identify them in the corresponding paper.[138,174,179,182] 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;[138,182] in panels **G, H** and **K** a negative environment is placed around Arg125;[174,182] in panel **D** the ring size is increased.[179] The ligand orientations are the result of their superposition with co-crystallized ligands from the same or very similar chemical series (*i.e.*, 4G1F[182] for panels **A**, **B**, **C**, **E**, **H**, **I**, **J** and **K**, 3KWF[179] for panel **D**, 1X70[138] for panel **F** and 2IIV[174] 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[187] and MarvinSketch programs.[213]

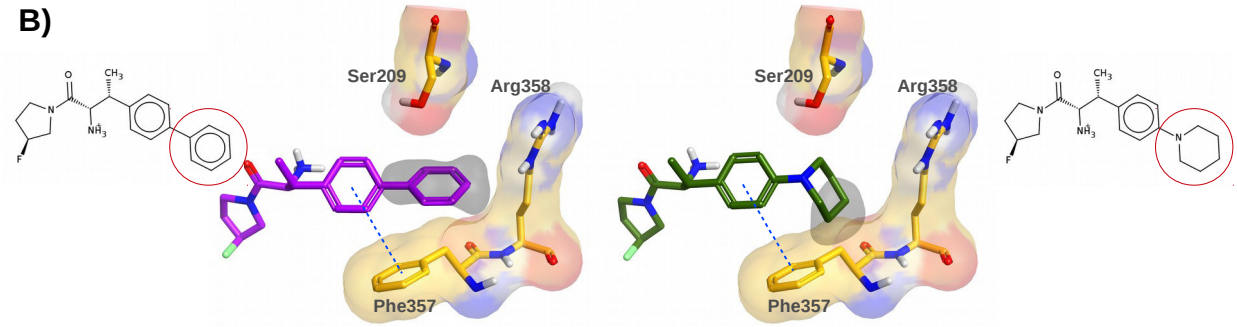
A)



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

similarity: 0.936 **disparity**: 37.6

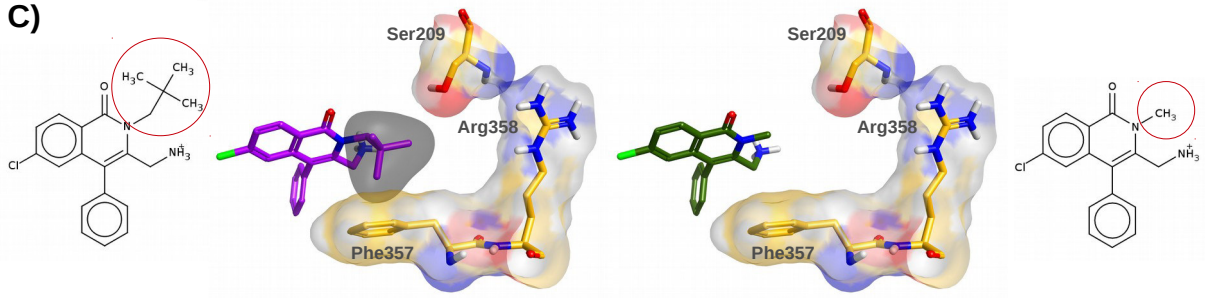
B)



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

similarity: 0.883 **disparity**: 15.5

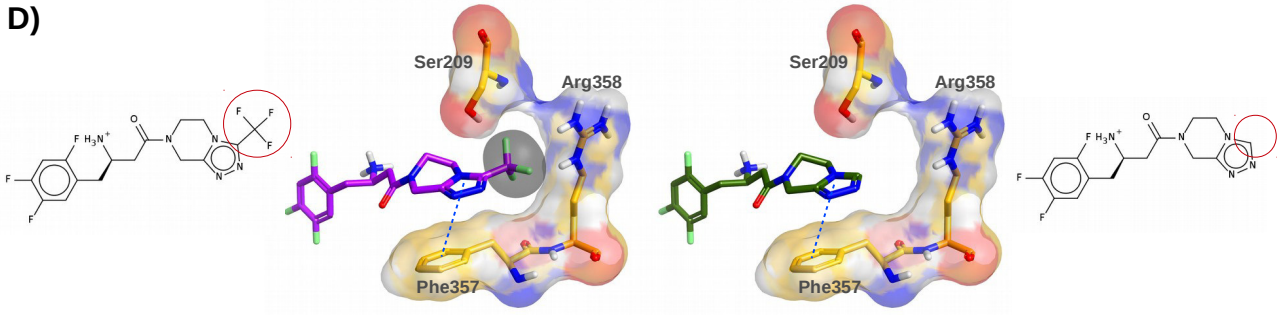
C)



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

similarity: 0.904 **disparity**: 9.6

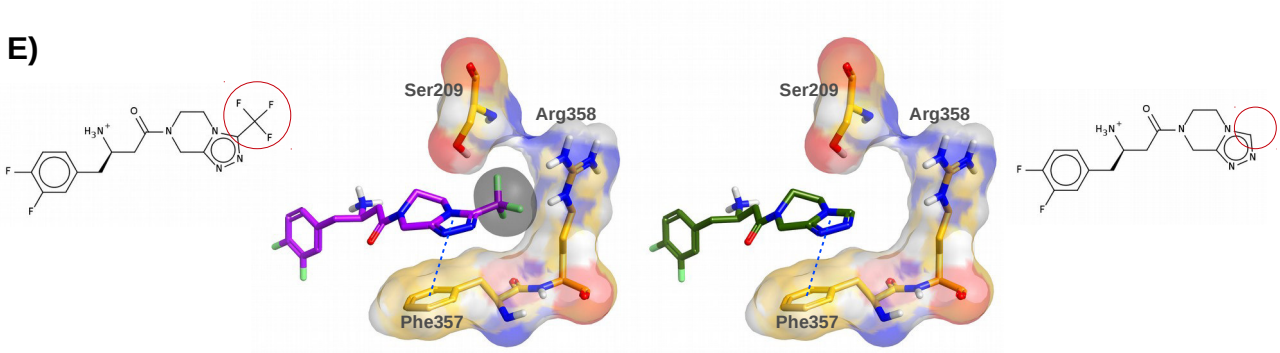
D)



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

similarity: 0.924 **disparity**: 7.5

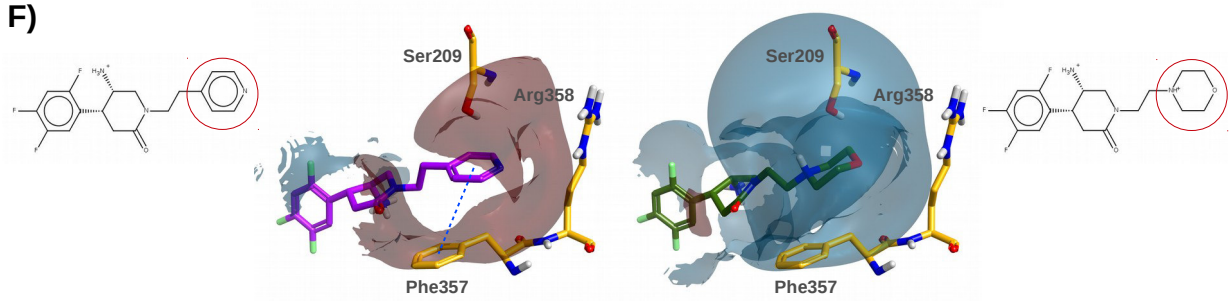
E)



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

similarity: 0.925 **disparity**: 7.3

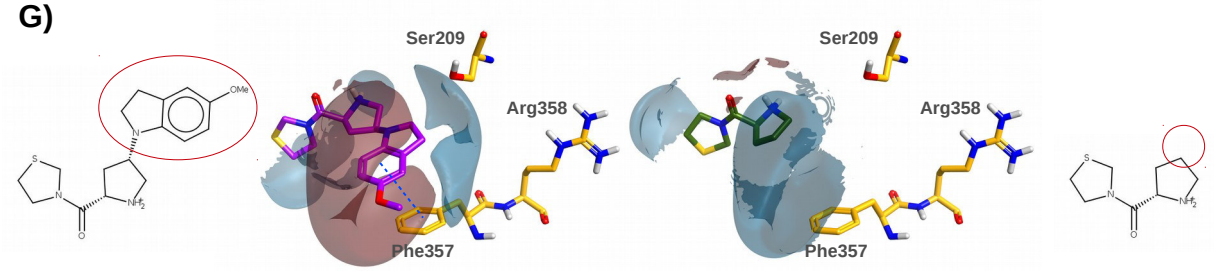
F)



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

similarity: 0.777 **disparity**: 6.2

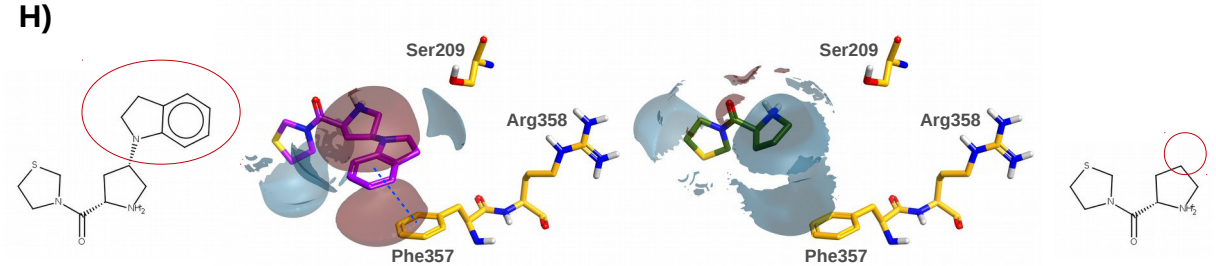
G)



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

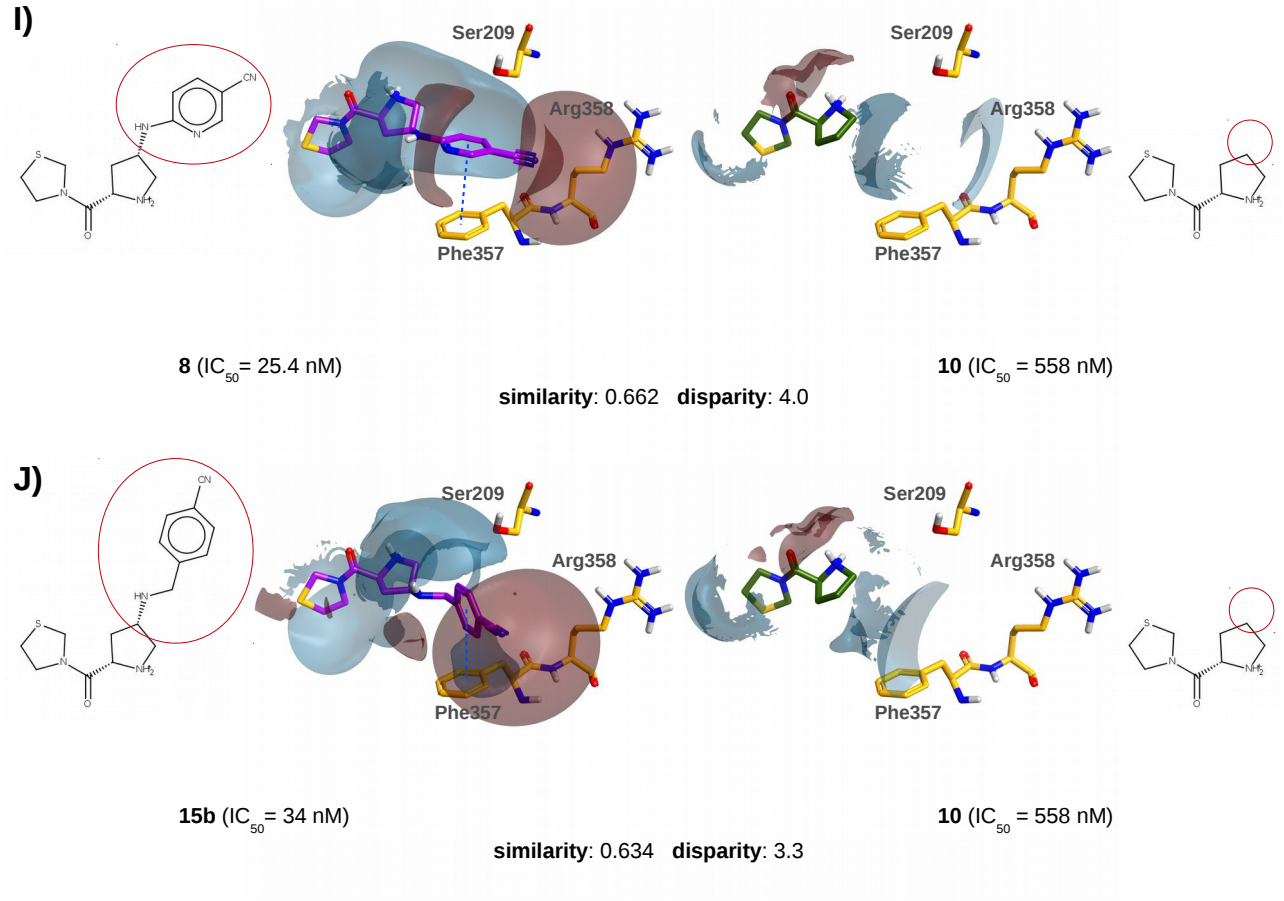
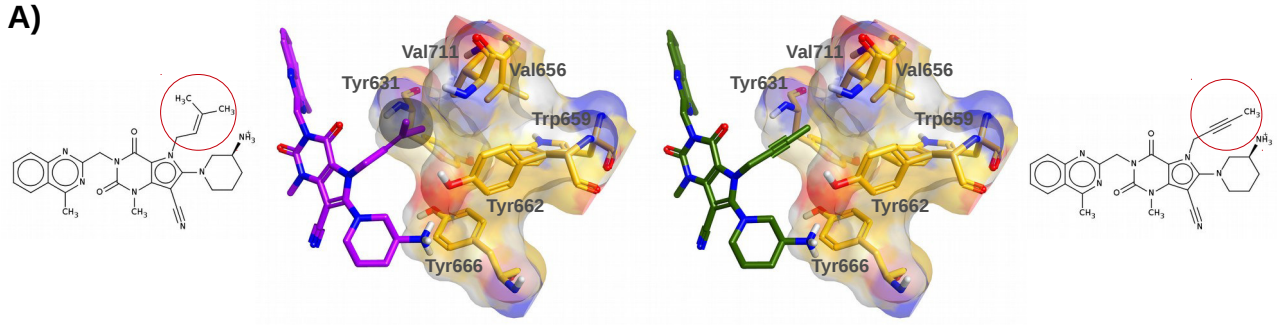


Figure S4. Comparison of the distribution of electrostatic and hydrophobic surfaces between pairs of compounds that differ in their interactions with residues Ser209, Phe357, Arg358. For each panel the compound with the highest activity is shown in purple on the left and the compound with the lowest activity is shown in green on the right. Molecules are labeled with the same names that identify them in the corresponding paper.[138,172,175,176,180,183] 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;[176] in panels **B**, **D** and **E** a hydrophobic interaction is established with the S_2 pocket;[138,172] in panel **C**, a new hydrophobic interaction is established with Phe357;[180] and in panels **F**, **G, H, I** and **J** a new π-π interaction is established with Phe357.[175,183] The blue dashed lines show the π-π interactions between the corresponding ligand and Phe357 and were calculated with the help of Maestro[200] 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[208] for panel **B**, 3OPM[180] for panel **C**, 1X70[138] for panels **D** and **E**, 2OQV[175] for panel **F** and 3VJK[114] 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[187] and MarvinSketch programs.[213]

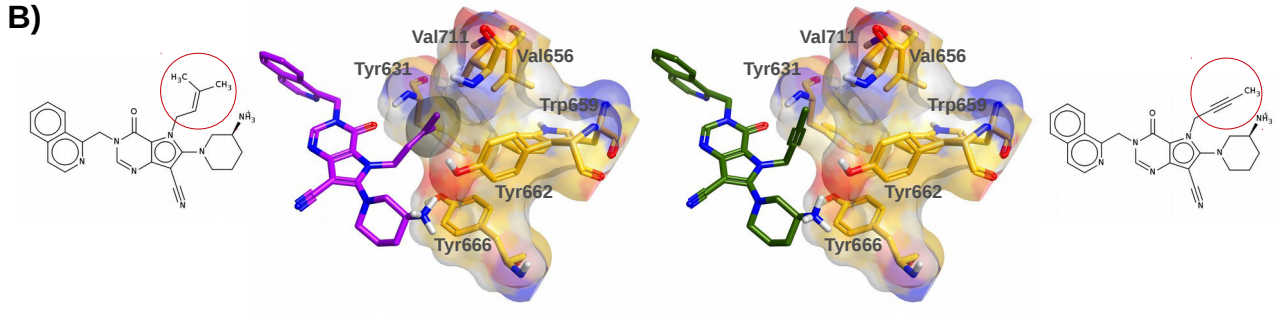
A)



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

similarity: 0.980 **disparity**: 42.0

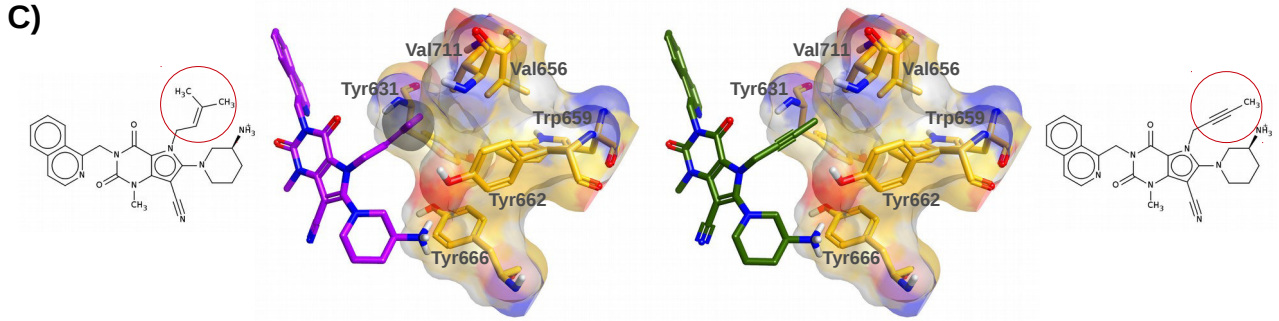
B)



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

similarity: 0.928 **disparity**: 14.0

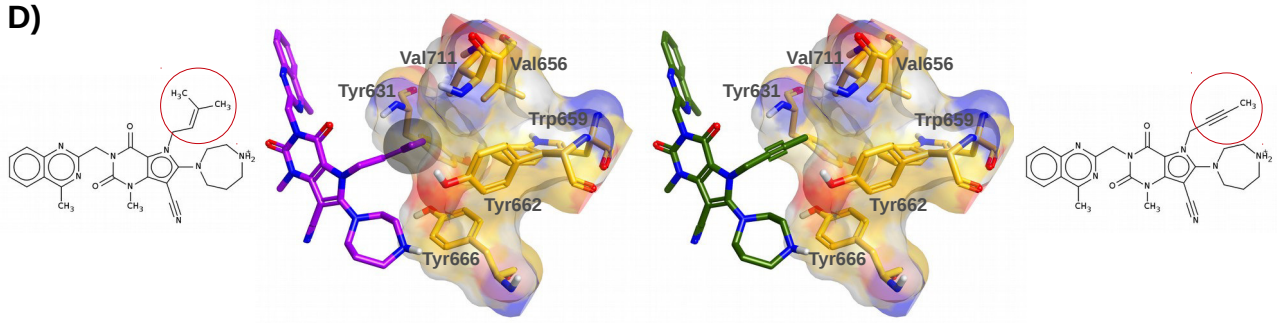
C)



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

similarity: 0.980 **disparity**: 12.0

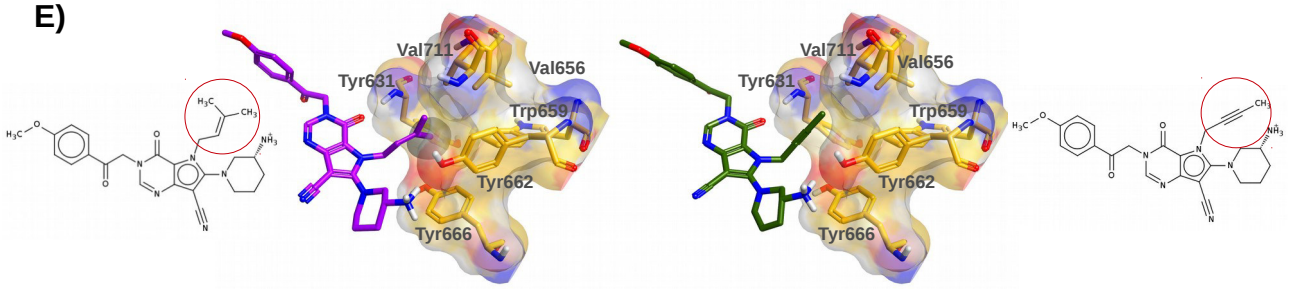
D)



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

similarity: 0.958 **disparity**: 11.6

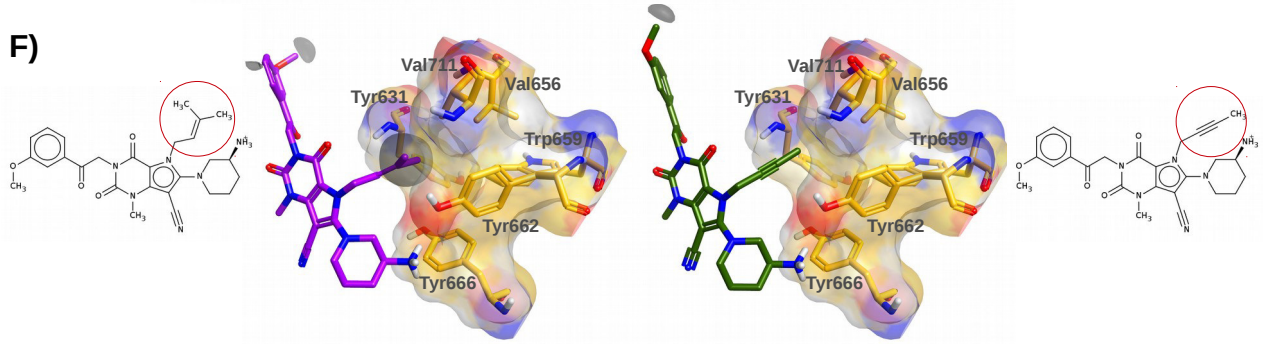
E)



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

similarity: 0.882 **disparity**: 10.0

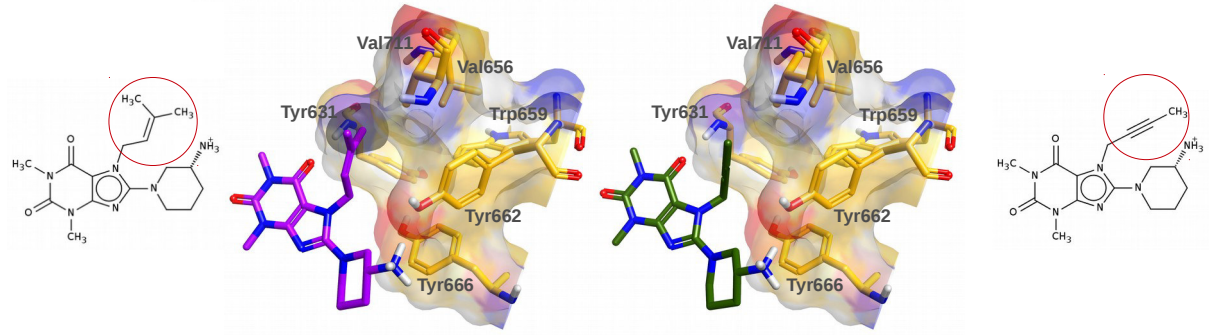
F)



(S)-11d (IC_{50}= 0.5 nM) **(S)-11i** (IC_{50}= 2 nM)

similarity: 0.904 **disparity**: 6.2

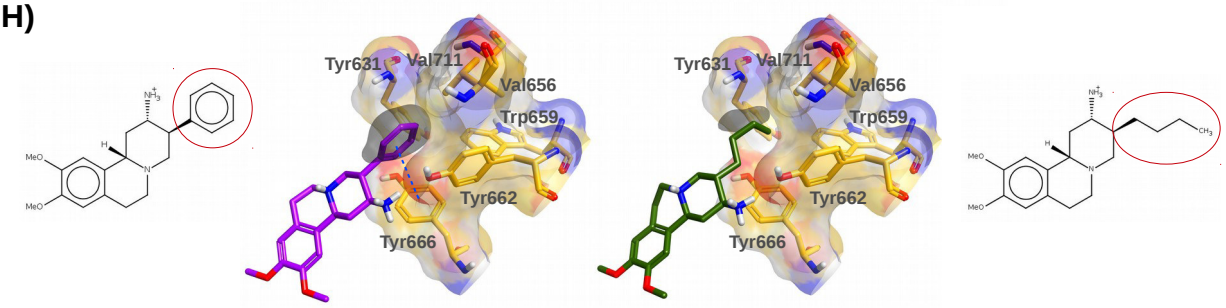
G)



6ac (IC_{50}= 35 nM) **6bc** (IC_{50} = 88 nM)

similarity: 0.916 **disparity**:4.7

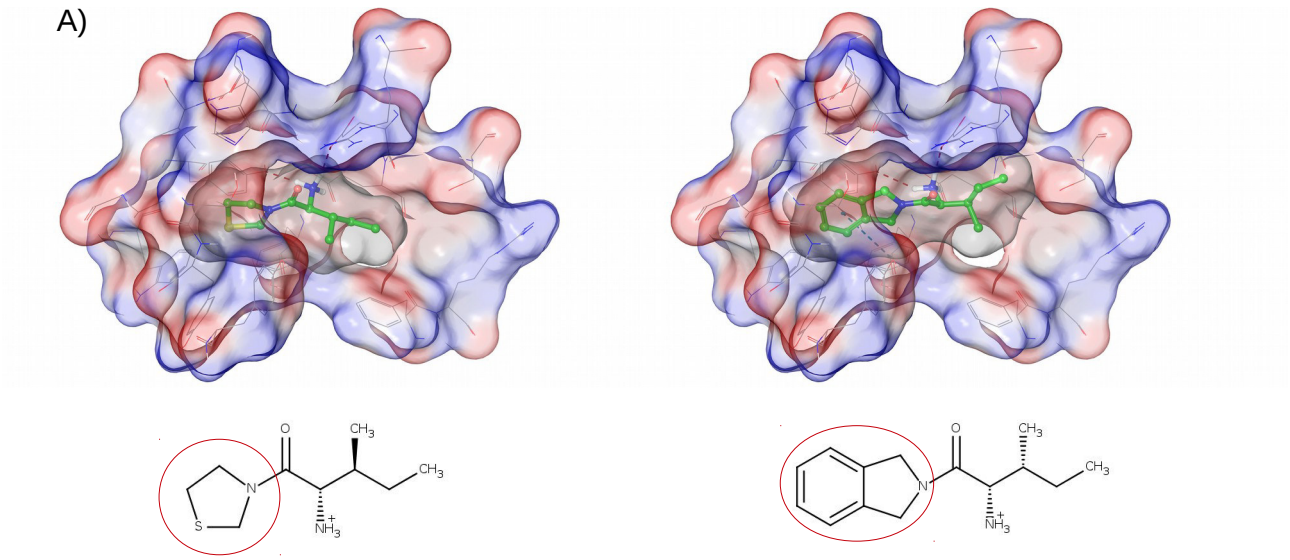
H)



13 (IC_{50}= 200 nM) **7** (IC_{50}= 520 nM)

similarity: 0.873 **disparity**: 3.3

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.[177,181,184] 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[184] (calculated with the help of Maestro[200] using default options). In panels **A** to **G**, a but-2-yn-1-yl substituent in the S_1 subsite is replaced by a prenyl substituent,[177,181] while in panel **H** a monobutyl group is replaced by a phenyl group.[184] The ligand orientations are the result of their superposition with co-crystallized ligands from the same or very similar chemical series (*i.e.*, 4A5S[181] for panels **A** to **F**, 2RGU[177] for panel **G**, and 3KWF[179] for panel **H**; residue locations in each panel are also taken from the corresponding PDB file). This figure was obtained with the help of the Forge[187] and MarvinSketch programs.[213]



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).[60,139] 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[138] using GlideXP.[169] 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.[200]

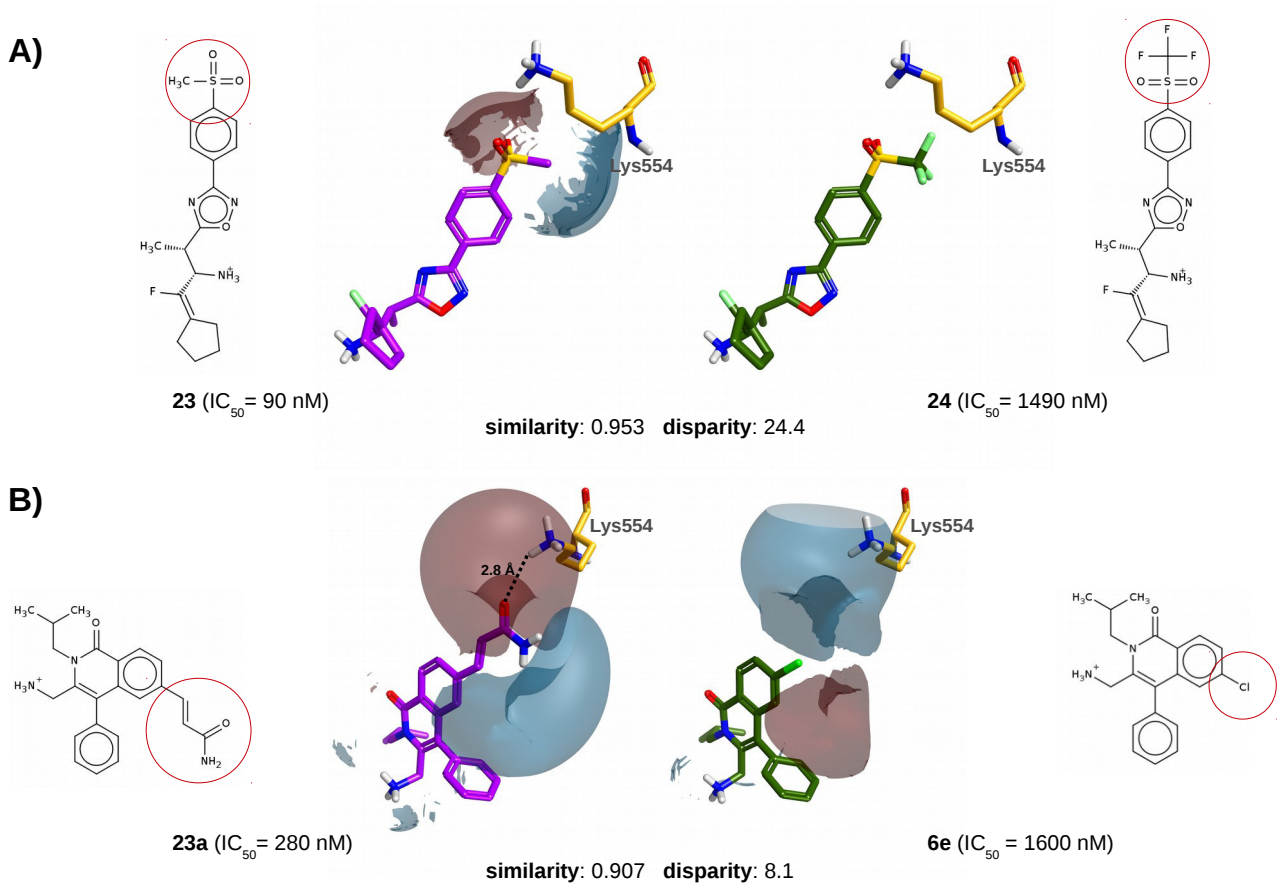
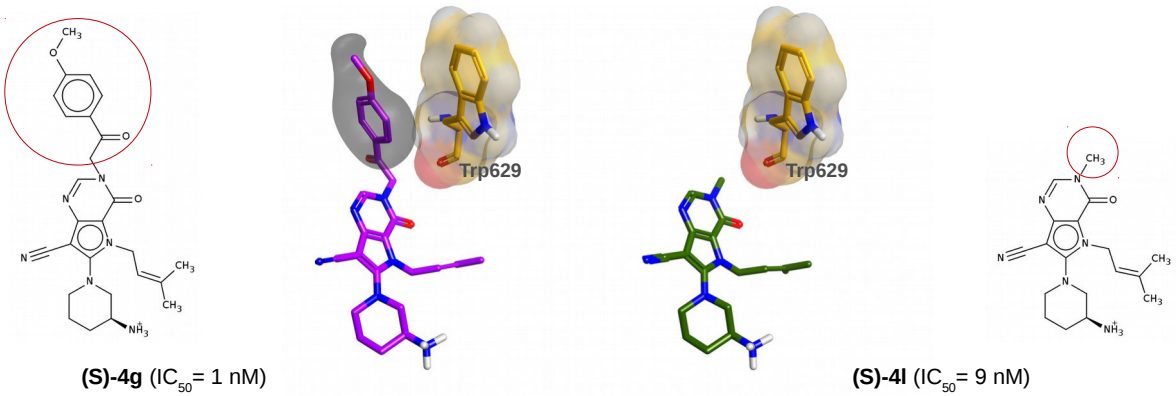


Figure S7. Comparison of the distribution of electrostatic surfaces between pairs of compounds that differ in their interactions with residue Lys554. For each panel the compound with the highest activity is shown in purple on the left and the compound with the lowest activity is shown in green on the right. Molecules are labeled with the same names that identify them in the corresponding paper.[159,180] 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;[159] and in panel **B**, a 2-carbamoyleth-1-en-1-yl group replaces a chlorine substituent.[180] 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[159] for panel **A** and 3OPM[180] 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[187] and MarvinSketch programs.[213]

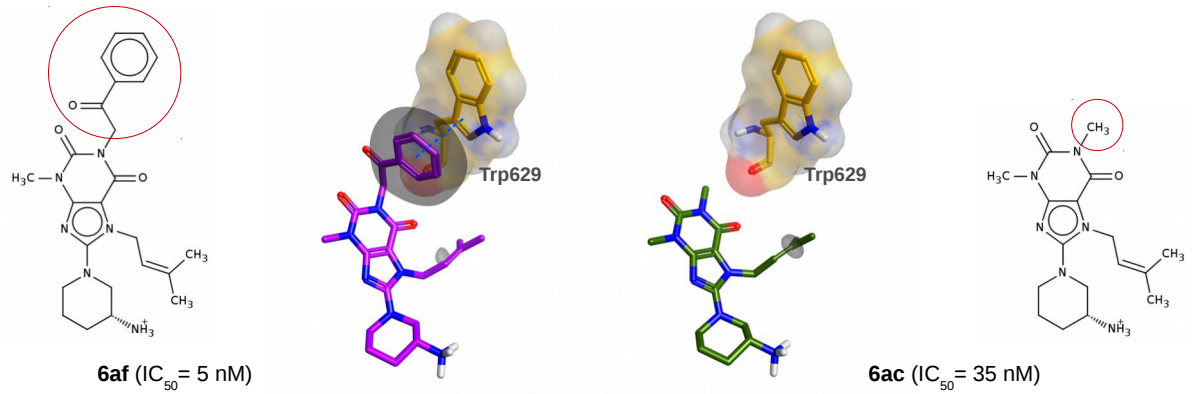
A)



(S)-4g (IC_{50}= 1 nM) **(S)-4l** (IC_{50}= 9 nM)

similarity: 0.830 **disparity**: 5.6

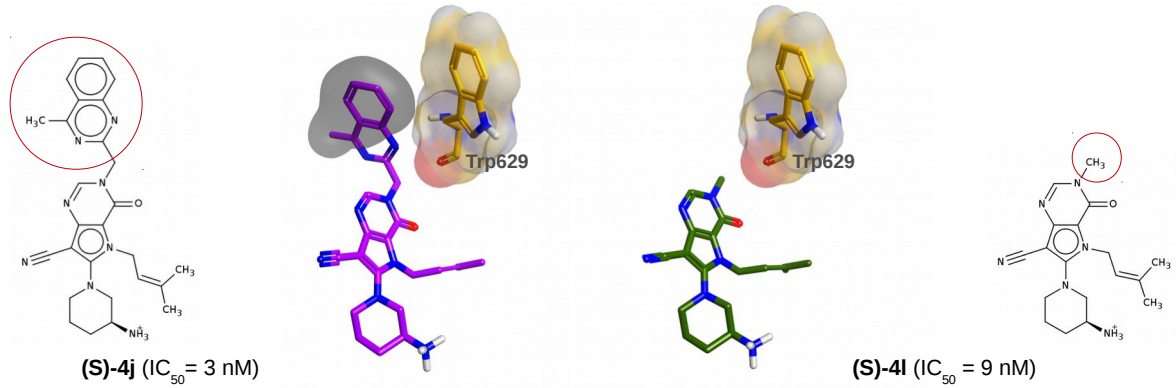
B)



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

similarity: 0.771 **disparity**: 3.7

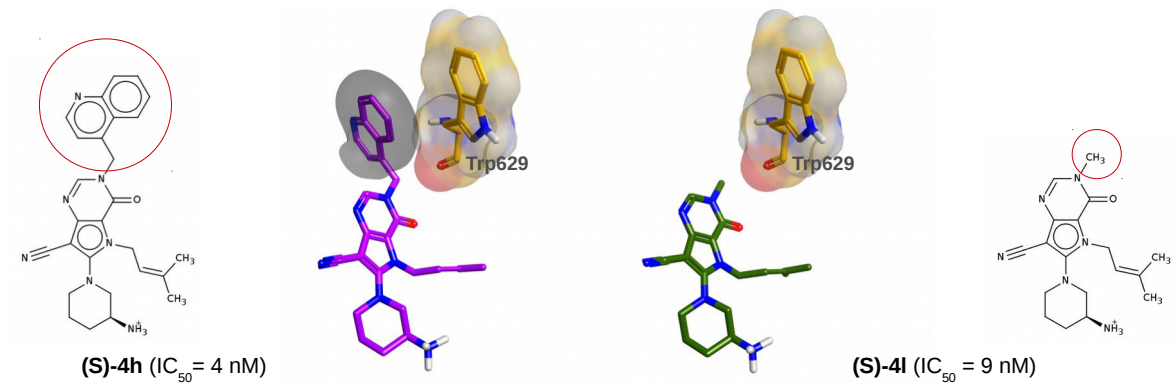
C)



(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.[177,181] 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.[177,181] The blue dashed line shows the π-π interaction between **6af** and Trp629 and was calculated with the help of Maestro[200] 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[181] for panels **A**, **C** and **D** and 2RGU[177] 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[187] and MarvinSketch programs.[213]