

Identification of Novel Human Dipeptidyl Peptidase-IV Inhibitors of Natural Origin (Part II): *In Silico* Prediction in Antidiabetic Extracts

Laura Guasch¹, Esther Sala¹, María José Ojeda¹, Cristina Valls¹, Cinta Bladé¹, Miquel Mulero¹, Mayte Blay¹, Anna Ardévol¹, Santiago Garcia-Vallvé^{1,2}, Gerard Pujadas^{1,2*}

1 Grup de Recerca en Nutrigenòmica, Departament de Bioquímica i Biotecnologia, Universitat Rovira i Virgili, Tarragona, Catalonia, Spain, **2** Centre Tecnològic de Nutrició i Salut (CTNS), TECNIO, CEICS, Reus, Catalonia, Spain

Abstract

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

Methodology/Principal Findings: Using an initial set of 29,779 natural products that are annotated with their natural source and an experimentally validated virtual screening procedure previously developed in our lab (Guasch et al.; 2012) [1], we have predicted 12 potential DPP-IV inhibitors from 12 different plant extracts that are known to have antidiabetic activity. Seven of these molecules are identical or similar to molecules with described antidiabetic activity (although their role as DPP-IV inhibitors has not been suggested as an explanation for their bioactivity). Therefore, it is plausible that these 12 molecules could be responsible, at least in part, for the antidiabetic activity of these extracts through their inhibitory effect on DPP-IV. In addition, we also identified as potential DPP-IV inhibitors 6 molecules from 6 different plants with no described antidiabetic activity but that share the same *genus* as plants with known antidiabetic properties. Moreover, none of the 18 molecules that we predicted as DPP-IV inhibitors exhibits chemical similarity with a group of 2,342 known DPP-IV inhibitors.

Conclusions/Significance: Our study identified 18 potential DPP-IV inhibitors in 18 different plant extracts (12 of these plants have known antidiabetic properties, whereas, for the remaining 6, antidiabetic activity has been reported for other plant species from the same *genus*). Moreover, none of the 18 molecules exhibits chemical similarity with a large group of known DPP-IV inhibitors.

Citation: Guasch L, Sala E, Ojeda MJ, Valls C, Bladé C, et al. (2012) Identification of Novel Human Dipeptidyl Peptidase-IV Inhibitors of Natural Origin (Part II): *In Silico* Prediction in Antidiabetic Extracts. PLoS ONE 7(9): e44972. doi:10.1371/journal.pone.0044972

Editor: Vladimir N. Uversky, University of South Florida College of Medicine, United States of America

Received: September 22, 2011; **Accepted:** August 16, 2012; **Published:** September 21, 2012

Copyright: © 2012 Guasch et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This research was supported by grants from the Ministerio de Educación y Ciencia of the Spanish Government (AGL2011-25831/ALI and AGL2011-23879/ALI) and from the Generalitat de Catalunya (ACC10/TECCT11-1-0012 program and XRQTC grant). N.G.-A. was supported by a pre-doctoral fellowship BES-2009-026450 from the Spanish Government. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: gerard.pujadas@urv.cat

Introduction

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

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

Numerous mechanisms of antidiabetic action have been proposed for extracts of the previous mentioned plants, some of them relate to their ability to stimulate insulin secretion [11]. Regarding the stimulation of insulin secretion, one target of interest for the antidiabetic action of these extracts is the serine protease dipeptidyl peptidase-IV (DPP-IV; EC 3.4.14.5) because the

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

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

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

The goal of the present work was to identify natural extracts with known antidiabetic activity that contain at least one molecule that we predict to be a DPP-IV inhibitor through a slightly modified version of the VS workflow described above [1]. Therefore, in this study, we provide new information about the active molecules in some natural extracts with antidiabetic properties and suggest that, at least in part, the mode of action of these molecules involves stimulating insulin secretion through the inhibition of DPP-IV. We also provide a list of plants with no

previously described antidiabetic activity that may contain DPP-IV inhibitors and that are related to plants with known antidiabetic activity. These plants represent a new source of potential antidiabetic extracts. In addition, the new DPP-IV inhibitors that we identified are chemically different from known DPP-IV inhibitors, and therefore, they could be used as lead-hopping candidates for the development of new antidiabetic drugs.

Results and Discussion

Virtual Screening Description and Application

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

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

This VS protocol was applied to an in-house database of 29,779 NPs with appropriate ADME/Tox properties. The first filter found that 10,883 molecules in our database have at least one

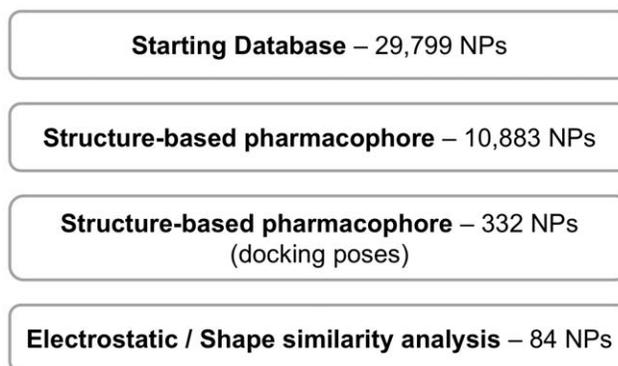


Figure 1. The VS workflow used in the present work. The number beside each VS step is the number of molecules that survived that step.

doi:10.1371/journal.pone.0044972.g001

conformer that after proper reorientation, matches the pharmacophore (see Figure 1). Only 332 out of these 10,883 molecules have docked conformations that without reorientation, are able to match the pharmacophore (see Figure 1). This reduction is useful because it discards those molecules that are predicted to bind in a non-productive way to the DPP-IV binding site. Finally, the later filter (*i.e.*, electrostatic and shape similarity screening) aims to smooth differences in chemical structures and translate them into criteria important for their intermolecular interactions with the ligand-binding site. This filter has been reported to be a valuable VS tool for the discovery of novel scaffolds [21], and in our case, it was applied to rescore the 332 hits that survived the previous filter. Consequently, only those molecules that had at least one docked pose that met the following conditions in the comparisons made by EON were predicted to inhibit DPP-IV: (a) $ET_{pb} \geq 0.468$ and (b) $ST \geq 0.237$. Interestingly, the fact that DPP-IV inhibitors (a) have a significant positive electrostatic potential in the region that interacts with the Glu205/Glu206 dyad (see Figure 1) and (b) that this ligand area matches the mandatory positive/donor site justifies the dominance of the electrostatic contribution over the shape contribution in the selected thresholds. Finally, the VS workflow identified 84 molecules with potential DPP-IV inhibitory activity (see Figure 1).

Virtual Screening Hits in Natural Extracts with Known Antidiabetic Activity

According to the information available in our in-house NPs database, the 84 molecules that were predicted by the VS workflow as potential DPP-IV inhibitors have been isolated from 95 different natural sources. Interestingly, a systematic bibliographic search of PubMed (<http://www.pubmed.org>) revealed that the extracts of 12 out of these 95 natural sources have been reported to exhibit antidiabetic activity (see Table S1). Moreover, among these 12 sources we found 12 VS hits that may, through their role as DPP-IV inhibitors, contribute to the observed antidiabetic activity of their corresponding extracts (see Table S1). In fact, a search using SciFinder (<http://www.cas.org/products/sfacad>) revealed that 6 out of these 12 natural compounds correspond (or are similar in chemical structure) to molecules with known antidiabetic properties (see Table S1). This finding further validates our methodology (in addition to the experimental validation of the original VS workflow [1]) and suggests that the mode of action of these molecules could involve DPP-IV inhibition. The remaining 6 natural compounds not previously reported to have antidiabetic properties represent new molecules that may also exhibit this bioactivity. In the next paragraphs, the most significant compounds found in these 12 antidiabetic extracts are discussed:

- N-Nororientaline (CAS number 29079-44-5, see Table S1) has been isolated from five plant species from genus *Erythrina* (*i.e.*, *Erythrina variegata* [22], *Erythrina crystagalli*, *Erythrina indica*, *Erythrina poeppigiana* [22], and *Erythrina arborescens*) and was identified as a DPP-IV inhibitor by our VS procedure. Extracts from *Erythrina variegata*, *Erythrina senegalensis*, *Erythrina addisoniae*, *Erythrina abyssinica*, and *Erythrina mildbraedii* are reported to enhance antidiabetic activity [23–27]. Isoprenylated flavonoids isolated from *Erythrina mildbraedii* and prenylflavonoids isolated from the *Erythrina senegalensis* roots have been described as inhibitors of two other proteins frequently targeted in T2DM treatment (*i.e.*, protein tyrosine phosphatase-1B [25,26] for the former class of molecules and acyl CoA:diacylglycerol acyltransferase [24] for the latter class). All of these flavonoid compounds are chemically related to our hit. Therefore, the

antidiabetic action of extracts from these plants may be the result of more than one bioactive component and mode of action.

- Tecostamine, which is found in *Tecoma stans*, has hypoglycemic properties similar to those of tecomine [28]. Our results therefore suggest that the hypoglycemic properties of tecostamine could be mediated by the inhibition of DPP-IV. In addition, the *Tecoma stans* aqueous extract possesses at least four antidiabetic-related activities (*i.e.*, intestinal α -glucosidase inhibition, post-prandial antihyperglycemic, hypocholesterolemic and hypotriglyceridemic effects). Therefore, although it is possible that most of these activities are exerted by the phenolic compounds present in the *Tecoma stans* aqueous extract, bioguided studies are necessary to confirm this hypothesis [29].
- Epinephrine (also known as adrenaline), which is found in *Scoparia dulcis*, has been reported to improve hypoglycemia [30]. *Scoparia dulcis* has been described as a folk-medicinal plant and has been traditionally used as a remedy for diabetes mellitus in India and for hypertension in Taiwan [31]. From Indian *Scoparia dulcis*, an antidiabetic compound named amellin was also isolated and characterized by Nath [32].
- From the same family as epinephrine, two additional compounds were predicted by our VS to be potential DPP-IV inhibitors. These compounds are (+)-pseudoephedrine (CAS number 90-82-4) and (–)-ephedrine (CAS number 299-42-3), and both have also been reported to have hypoglycemic activity [33]. Both molecules are found in several *Ephedra* species (*Ephedra alata* [34], *Ephedra distachya* [35], *Ephedra equisetina* [36], *Ephedra gerardiana* [35], *Ephedra shenungiana* [37], *Ephedra sinica*, *Ephedra vulgaris* and *Ephedra pflanze*). However, *Ephedra distachya* and *Ephedra alata* are the only species reported to have antidiabetic properties [33,38]; therefore, we proposed that the remaining *Ephedra* species may be new sources of antidiabetic extracts.
- The molecules ajmaline and isosandwichine, which are enantiomers with the same CAS number, 509-37-5, are found in several *Rauwolfia* species. *Rauwolfia vomitoria* has been investigated for the content of alkaloids, especially those alkaloids with hypotensive and anti-inflammatory properties [39] in addition to the antidiabetic properties. *Rauwolfia serpentina* is also used as antidiabetic extract [40]. The remaining species that contain these molecules, such as *Rauwolfia canescens* [41], *Rauwolfia degeneri* [42], *Rauwolfia densiflora* [43], *Rauwolfia heterophylla* [44], *Rauwolfia indecora*, *Rauwolfia obscura* [45] and *Rauwolfia tetraphylla* [42], are putative antidiabetic extracts.
- Serpinine (CAS number 509-38-6) is isolated from *Vinca major* and several *Rauwolfia* species (*Rauwolfia obscura* [45], *Rauwolfia tetraphylla*, *Rauwolfia serpentina* [46] and *Rauwolfia sellowii* [1]), belongs to the same cluster as ajmaline and isosandwichine (*i.e.*, cluster 78; see Table S1). Therefore, they share similar chemical structures and natural sources. Moreover, *Vinca major* organic leaf extract strongly stimulates glucose utilization [47].
- One interesting hit predicted to be a DPP-IV inhibitor is an epicatechin derivative that is found in *Vitis vinifera*. An antihyperglycemic effect in streptozotocin-induced diabetic rats and insulinomimetic activity in insulin-sensitive cell lines have been described for grape seed procyanidin extracts (GSPE) [48]. In addition, it has been demonstrated that oligomeric procyanidins from GSPE activate the insulin receptor by interacting with and inducing the phosphorylation of the insulin receptor and that this interaction leads to increased glucose uptake [49]. Moreover, several epicatechin

derivates have been reported to have antidiabetic properties (the most studied of which is epigallocatechin gallate [50]). Some findings demonstrate that epigallocatechin gallate may be a novel, plant-derived compound capable of reducing the risk of type 1 diabetes [51]. Therefore, the DPP-IV inhibition induced by this epicatechin derivate may contribute to the antihyperglycemic effect of GSPE [48].

- The remaining 3 molecules predicted to be DPP-IV inhibitors through our VS workflow and that are found in extracts with described antidiabetic properties are hydroxysmimovine from *Galega orientalis* [52], (-)-halosaline (CAS number 26648-71-5) from *Haloxylon salicornicum* [53] and isochanoclavin-(I) (CAS number 1150-44-3) from *Pennisetum typhoideum* [54] (see Table S1). Our results suggest that these molecules could be DPP-IV inhibitors and that extracts containing these molecules could potentiate the glucose-induced insulin response by prolonging the half-lives of GLP-1 and GIP incretins, due to the inhibition of DPP-IV. This information is novel and relevant, as no mechanism that explains the antidiabetic properties of these extracts has been previously suggested.

Virtual Screening Hits in Natural Extracts with No Described Antidiabetic Activity

Taking into account that extracts from closely related species of the same *genus* may share a high number of components, we also determined if any of our VS hits that were isolated from plants with no described antidiabetic activity belong to the same *genus* as species with known antidiabetic properties. We identified 6 molecules isolated from 6 different plants, *Aconitum japonicum*, *Ervatamia officinalis*, *Solanum nudum*, *Solanum sodomaeum*, *Stephania cepharantha* and *Tabernaemontana eglandulosa* (see Table S2), that meet these criteria. The related species with described antidiabetic properties are *Aconitum carnichaelii* [33], *Aconitum moschatum* [2], *Aconitum violaceum* [2], *Ervatamia microphylla* [55], *Solanum lycocarpum* [56], *Solanum nigrum* [57], *Solanum xanthocarpum* [58], *Stephania hernandifolia* [59], *Stephania glabra* [60], *Stephania tetrandra* [61], and *Tabernaemontana divaricata* [55]. Therefore, it is plausible to hypothesize that these 6 plants could also have antidiabetic properties mediated, at least partially, by the inhibition of DPP-IV.

Finding New Scaffolds of Natural Origin for DPP-IV Inhibitors

The 18 molecules in Tables S1 and S2 that were predicted to be DPP-IV inhibitors are of interest. To quantify the number of new scaffolds for DPP-IV inhibitors that were identified in our study, we merged the 18 VS hits with 2,342 known DPP-IV inhibitors, and the resulting set was classified according to structural similarity into 99 clusters (results not shown). Interestingly, the 18 hits were classified in 13 clusters (see Tables S1 and S2) that do not contain known DPP-IV inhibitors (results not shown). Thus, these 18 predicted DPP-IV inhibitors correspond to 13 different chemical scaffolds that are unrelated to those present in known DPP-IV inhibitors, and, consequently, these new scaffolds could be used either in lead-hopping experiments to identify new DPP-IV inhibitors or in structure-activity studies to identify natural-product derivates with stringer DPP-IV inhibition activity than the original NPs from which they are derived.

Conclusions

In a previous study [1], we developed a VS workflow that was able to distinguish successfully molecules that inhibit DPP-IV and molecules that do not inhibit this enzyme. We experimentally

demonstrated that our VS protocol was able to identify DPP-IV inhibitors that **(a)** were not structurally related to any known molecule that inhibits DPP-IV and **(b)** have never been reported to have antidiabetic activity. In the present work, we applied a slightly modified version of this VS workflow to an in-house database of 29,779 NPs annotated with their corresponding natural source(s). From this initial set of NPs, our VS procedure identified as potential DPP-IV inhibitors 84 hit molecules that have been isolated from 95 different natural sources. Interestingly, after an exhaustive bibliographic search, our results demonstrate that we are able to predict (a) 12 DPP-IV inhibitors that are present in 12 plant extracts with known antidiabetic activity and (b) 6 DPP-IV inhibitors that are present in 6 different plants species with no described antidiabetic activity but that share the same *genus* as plants with known antidiabetic properties (consequently, it could be suggested that these plants represent a potential new source of antidiabetic extracts). Moreover, none of these 18 hits exhibits chemical similarity with 2,342 known DPP-IV inhibitors, and, therefore, it is expected that a significant number of these hits could be lead-hopping candidates for the development of new DPP-IV inhibitors. At this point, it is also interesting to note that the analysis of the chemical structures of these 18 NP hits revealed that the majority of them are alkaloids containing basic nitrogen atoms (essential for proper interaction with the Glu205/Glu206 dyad). Moreover, our results provide a new hypothesis about the mechanisms by which, at least partially, these 12 extracts exert their antidiabetic effects (*i.e.*, improving glucose homeostasis by prolonging the activity of GLP-1 and GIP through DPP-IV inhibition).

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

Methods

Initial Dataset of Natural Compounds Used

The database of NPs that was screened by the VS workflow contains 29,779 NPs from different origins (*e.g.*, plants, and fungi) with appropriate ADME/Tox properties and no chiral ambiguities. An important characteristic of this database is that each molecule is annotated with (a) the natural sources from which it has been obtained and (b) the bibliographic references that describe how to extract the relevant molecule from each natural source. The 3D structures of the molecules in this NP database were processed with LigPrep v2.3 (Schrodinger LLC., Portland, USA; <http://www.schrodinger.com>) using the following parameters: (a) the force field used was OPLS 2005; (b) all possible ionization states at pH 7.0 ± 2.0 were generated with Ionizer; (c) the desalt option was activated; (d) tautomers were generated for all ionization states at pH 7.0 ± 2.0 ; (e) chiralities were determined from the 3D structure; and (f) one low-energy ring conformation per ligand was generated. Conformations and sites for the resulting ligand structures were determined during the generation of the corresponding Phase [62] databases with the **Generate Phase Database** graphic front-end. Default parameter values were used during this conformer generation process, with the exception of the maximum number of conformers per structure, which was increased from 100 (the default value) to 200. The conformer sites were generated with definitions made by adding to the default

built-in Phase definitions the ability to consider aromatic rings as hydrophobic groups.

Virtual Screening Workflow

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

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

The first step of the VS workflow uses the common structure-based pharmacophore to screen the conformer database with Phase v3.1 (Schrödinger LLC., Portland, USA; <http://www.schrodinger.com>) and allows the reorientation of the conformers to determine if they match the pharmacophore. Only those ligands with at least one conformer that matches the two compulsory sites of the common pharmacophore and at least one of the optional sites (without sterically colliding with the excluded volumes) survive this VS step. These ligands were docked onto the binding pocket of the PDB file 3C45 with eHits v2009 (SimBioSys Inc., Toronto, Canada; <http://www.simbiosys.ca/ehits>) [63] using default docking conditions, with the exception of the length of the sides of the cubic box encompassing the DPP-IV binding site, which was increased from 10 Å to 15 Å. Next, the resulting docked poses were again filtered with Phase through the common pharmacophore using the same filtering conditions as in the first Phase run (with the exception that no reorientation of the docked poses was allowed during the search). Thus, only docking poses compatible with the pharmacophore survived this filter. Finally, in the last step of the VS protocol, the poses that were hits in this second pharmacophore screen were submitted to a shape and electrostatic potential comparison with the DPP-IV inhibitor present in the 3C45 PDB file to rescore the hits. This comparison was completed with EON v2.0.1 (OpenEye Scientific Software, Inc., Santa Fe, New Mexico, USA; <http://www.eyesopen.com>) and used the Electrostatic Tanimoto combo (ET_combo) score as the similarity criterion. The ET_combo score is the sum of two calculations: (a) the Shape Tanimoto (ST) score, which is a quantitative measure of three-dimensional overlap (where 1 corresponds to a perfect overlap; *i.e.* the same shape), and (b) the Poisson-Boltzmann Electrostatic Tanimoto (ET_pb) score, which compares the electrostatic potential of two small molecules and ranges from 1 (identical potential) to a negative value that results from the overlap of positive and negative charges. In this work, we determined the EON threshold scores by comparing the inhibitor in 3C45 with the inhibitors found in a group of 24 PDB complexes

of DPP-IV and reversible drugs that inhibit this protein (*i.e.*, 1N1M, 20GZ, 2FJP, 2HHA, 2I78, 2IIT, 2IIV, 2OLE, 2ONC, 2OQI, 2OQV, 2QOE, 2QT9, 2QTB, 2RGU, 2RIP, 3C43, 3CCC, 3D4L, 3F8S, 3H0C, 3HAB, and 3HAC). This comparison of the experimental poses of DPP-IV inhibitors yielded threshold values for DPP-IV inhibition that were slightly lower than the ones used in the original VS workflow [1] ($ET_{pb} \geq 0.468$ and $ST \geq 0.237$ instead of $ET_{pb} \geq 0.623$ and $ST \geq 0.244$), probably because the original threshold values were obtained exclusively using potent DPP-IV inhibitors (*i.e.*, $IC_{50} \leq 10$ nM); in this study, this condition was relaxed slightly. After rescoring with EON, only those NPs with at least one conformation with $ET_{pb} \geq 0.468$ and $ST \geq 0.237$ relative to 3C45's inhibitor were considered to be DPP-IV inhibitor candidates.

Structural Similarity Analysis

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

Supporting Information

Table S1 Natural extracts with reported antidiabetic activity that contain molecules that were predicted to be DPP-IV inhibitors by our VS protocol. The first column shows the 2D structure of each molecule and, when available, the corresponding common name and/or CAS number. The second column shows the number of the cluster in which the corresponding molecule was classified when its structure was compared with those of a group of 2,342 known DPP-IV inhibitors. The third column shows the scientific name of one of the sources in which the antidiabetic activity has been reported (rows in that table are alphabetically sorted based on this column). Bibliographic references for each molecule are divided into three columns in which (a) the first column presents papers that describe the purification of the molecule from the corresponding extract; (b) the second column lists papers that describe the antidiabetic activity of the corresponding extract; and (c) the third column lists papers, when available, that describe the antidiabetic activity of the corresponding molecule or one that is very similar to it. (DOC)

Table S2 Natural extracts with no described antidiabetic activity (but from the same genus as plants with extracts with described antidiabetic activity) that contain molecules that are predicted to be DPP-IV inhibitors by our VS protocol. The first column shows the 2D structure of each molecule and, when available, the corresponding common name or CAS number. The second column shows the number of the cluster in which the corresponding molecule was classified when its structure was compared with those of a group of 2,342 known DPP-IV inhibitors. The third column lists the source from which the VS hits have been purified (rows in that table are alphabetically sorted based on this column). The fourth column lists the papers that describe the purification of the each molecule from the corresponding extract. The fifth column shows which are the extracts from the same genus where the antidiabetic activity has

been described. Finally, the last column lists papers that describe the antidiabetic activity of the corresponding extract. (DOC)

Acknowledgments

We thank the Servei de Disseny de Fàrmacs (Drug Design Service) of the Centre de Serveis Científics i Acadèmics de Catalunya (CESCA; <http://www.cesca.cat>) for providing access to Schrödinger software and OpenEye Scientific Software, Inc. (Santa Fe, New Mexico, USA; <http://www>.

References

- Guasch L, Ojeda MJ, González-Abuín N, Sala E, Ceretó A, et al. (2012) Identification of novel human dipeptidyl peptidase-IV inhibitors of natural origin (part I): Virtual screening and activity assays.
- Howes M, Simmonds M (2005) Plants used in the treatment of diabetes. In: Soumyanath A, editor. *Traditional Medicines for Modern Times*. Press.
- Yin J, Zhang H, Ye J (2008) Traditional chinese medicine in treatment of metabolic syndrome. *Endocr Metab Immune Disord Drug Targets* 8(2): 99–111.
- Modak M, Dixit P, Londhe J, Ghaskadbi S, Paul A Devasagayam T (2007) Indian herbs and herbal drugs used for the treatment of diabetes. *J Clin Biochem Nutr* 40(3): 163–173.
- Andrade-Cetto Adolfo, Heinrich M (2005) Mexican plants with hypoglycaemic effect used in the treatment of diabetes. *J Ethnopharmacol* 99(3): 325–348.
- Haque N, Salma U, Nurunnabi T, Uddin M, Jahangir M, et al. (2011) Management of type 2 diabetes mellitus by lifestyle, diet and medicinal plants. *Pak J Biol Sci* 14(1): 13–24.
- Malviya N, Jain S, Malviya S (2010) Antidiabetic potential of medicinal plants. *Acta Pol Pharm* 67(2): 113–118.
- Prabhakar PK, Doble M (2008) A target based therapeutic approach towards diabetes mellitus using medicinal plants. *Curr Diabetes Rev* 4(4): 291–308.
- Qi L, Liu E, Chu C, Peng Y, Cai H, et al. (2010) Anti-diabetic agents from natural products—an update from 2004 to 2009. *Curr Top Med Chem* 10(4): 434–457.
- Bnouham M, Ziyyat A, Mekhfi H, Tahri A, Legssyer A (2006) Medicinal plants with potential antidiabetic activity - A review of ten years of herbal medicine research (1990–2000). *Int J Diabetes Metab* 14(1).
- Prabhakar PK, Doble M (2011) Mechanism of action of natural products used in the treatment of diabetes mellitus. *Chin J Integr Med* 17(8): 563–574.
- Havale SH, Pal M (2009) Medicinal chemistry approaches to the inhibition of dipeptidyl peptidase-4 for the treatment of type 2 diabetes. *Bioorg Med Chem* 17(5): 1783–1802.
- Mentlein R, Gallwitz B, Schmidt W (1993) Dipeptidyl-peptidase IV hydrolyses gastric inhibitory polypeptide, glucagon-like peptide-1(7–36)amide, peptide histidine methionine and is responsible for their degradation in human serum. *Eur J Biochem* 214(3): 829–835.
- Brubaker PL, Drucker DJ (2004) Minireview: Glucagon-like peptides regulate cell proliferation and apoptosis in the pancreas, gut, and central nervous system. *Endocrinology* 145(6): 2653–2659.
- Meier JJ, Nauck MA, Schmidt WE, Gallwitz B (2002) Gastric inhibitory polypeptide: The neglected incretin revisited. *Regul Pept* 107(1–3): 1–13.
- Kim D, Wang L, Beconi M, Eiermann GJ, Fisher MH, et al. (2005) (2R)-4-oxo-4-[3-(trifluoromethyl)-5,6-dihydro[1,2,4]triazolo[4,3-a]pyrazin-7(8H)-yl]-1-(24,5-trifluorophenyl)butan-2-amine: A potent, orally active dipeptidyl peptidase IV inhibitor for the treatment of type 2 diabetes. *J Med Chem* 48(1): 141–151.
- Villhauer EB, Brinkman JA, Naderi GB, Burkey BF, Dunning BE, et al. (2003) 1-[[[3-hydroxy-1-adamantylamino]acetyl]-2-cyano-(S)-pyrrolidine: A potent, selective, and orally bioavailable dipeptidyl peptidase IV inhibitor with antihyperglycemic properties. *J Med Chem* 46(13): 2774–2789.
- Augeri DJ, Robl JA, Betebenner DA, Magnin DR, Khanna A, et al. (2005) Discovery and preclinical profile of saxagliptin (BMS-477118): A highly potent, long-acting, orally active dipeptidyl peptidase IV inhibitor for the treatment of type 2 diabetes. *J Med Chem* 48(15): 5025–5037.
- Zettl H, Schubert-Zsilavec Manfred, Steinhilber D (2010) Medicinal chemistry of incretin mimetics and DPP-4 inhibitors. *ChemMedChem* 5(2): 179–185.
- Kuhn B, Hennig M, Mattei P (2007) Molecular recognition of ligands in dipeptidyl peptidase IV. *Current Topics in Medicinal Chemistry* 7(6): 609–619.
- Sala E, Guasch L, Iwaszkiewicz J, Mulero M, Salvado M, et al. (2011) Identification of human IKK-2 inhibitors of natural origin (part I): Modeling of the IKK-2 kinase domain, virtual screening and activity assays. *PLoS One* 6(2): e16903.
- Ito K, Haruna M, Furukawa H (1975) [Studies on the erythrina alkaloids. X. alkaloids of several erythrina plants from singapore]. *Yakugaku Zasshi* 95(3): 358–362.
- Kumar A, Lingadurai S, Shrivastava TP, Bhattacharya S, Haldar PK (2011) Hypoglycemic activity of erythrina variegata leaf in streptozotocin-induced diabetic rats. *Pharm Biol* 49(6): 577–582.
- Oh WK, Lee CH, Seo JH, Chung MY, Cui L, et al. (2009) Diacylglycerol acyltransferase-inhibitory compounds from erythrina senegalensis. *Arch Pharm Res* 32(1): 43–47.
- Bae EY, Na M, Njamen D, Mbafor JT, Fomum ZT, et al. (2006) Inhibition of protein tyrosine phosphatase 1B by prenylated isoflavonoids isolated from the stem bark of erythrina addisoniae. *Planta Med* 72(10): 945–948.
- Na M, Jang J, Njamen D, Mbafor JT, Fomum ZT, et al. (2006) Protein tyrosine phosphatase-1B inhibitory activity of isoprenylated flavonoids isolated from erythrina mildbraedii. *J Nat Prod* 69(11): 1572–1576.
- Nguyen P, Nguyen T, Dao T, Kang H, Ndieth D, et al. (2010) AMP-activated protein kinase (AMPK) activation by benzofurans and coumestans isolated from erythrina abyssinica. *J Nat Prod* 73(4): 598–602.
- Hammouda Y, Rashid K, Amer S (1964) Hypoglycaemic properties of tecomine and tecostamine. *J Pharm Pharmacol* 16: 833–834.
- Aguilar-Santamaria L, Ramirez G, Nicasio P, Alegria-Reyes C, Herrera-Arellano A (2009) Antidiabetic activities of tecoma stans (L.) juss. ex kunth. *J Ethnopharmacol* 124(2): 284–288.
- Ly TT, Hewitt J, Davey RJ, Lim EM, Davis EA, et al. (2011) Improving epinephrine responses in hypoglycemia unawareness with real-time continuous glucose monitoring in adolescents with type 1 diabetes. *Diabetes Care* 34(1): 50.
- Latha M, Pari L, Sitasawad S, Bhone R (2004) Scoparia dulcis, a traditional antidiabetic plant, protects against streptozotocin induced oxidative stress and apoptosis in vitro and in vivo. *J Biochem Mol Toxicol* 18(5): 261–272.
- Nath CM, Chakraborty KM, Brahmachari DH (1945) Investigations on the new antidiabetic principle (amellin); its role in the reduction of acetone bodies and the increase of alkali reserve of the blood of diabetics. *Ann Biochem Exp Med* 5(3): 101–104.
- Kono C, Mizuno T, Hikino H (1985) Isolation and hypoglycemic activity of ephedrans A, B, C, D and E, glycans of ephedra distachya herbs. *Planta Med* (2): 162–163.
- Black OF, Kelly JW (1927) Pseudo ephedrine from Ephedra alata. *Amer. Jour. Pharm.* 99: 748–751.
- Grue-Sorensen G, Spenser ID (1989) The biosynthesis of ephedrine. *Can J Chem* 67: 998–1009.
- Osadchii SA, Shults EE, Polukhina EV, Shakirov MM, Vasilevskii SF, et al. (2007) Study of alkaloids of the siberian and altai flora 14. synthesis of alkaloid-based tertiary N-(3-arylprop-2-ynyl)amines. *Russian Chemical Bulletin* 56(6): 1261–1267.
- Gilg E, Schürhoff P (1930) Die ephedrinhaltigen stammpflanzen der "mauhang"-droge. *Arch Pharm (Weinheim)* 233–239.
- Shabana MM, Mirhom YW, Genenah AA, Aboutabl EA, Amer HA (1990) Study into wild egyptian plants of potential medicinal activity. ninth communication: Hypoglycaemic activity of some selected plants in normal fasting and alloxanised rats. *Arch Exp Veterinarmed* 44(3): 389–394.
- Campbell J, Mortensen A, Molgaard P (2006) Tissue lipid lowering-effect of a traditional nigerian anti-diabetic infusion of rauwolfia vomitoria foliage and citrus aurantium fruit. *J Ethnopharmacol* 104(3): 379–386.
- Benzi G, Villa R, Dossena M, Vercesi L, Gorini A, et al. (1984) Cerebral and cerebellar metabolic changes induced by drugs during the recovery period after profound hypoglycemia. *Farmacol Sci* 39(1): 44–56.
- Gosh BP (1958) Isolation of serpentine and ajmaline from the root of rauwolfia Canescens. *Naturwissenschaften* 45(15): 365.
- Gorman M, Neuss N, Djerassi C, Kutney JP, Scheuer PJ (1957) Alkaloid studies—XIX: Alkaloids of some hawaiian rauwolfia species: The structure of sandwicine and its interconversion with ajmaline and ajmalidine. *Tetrahedron* 1(4): 328–337.
- Chatterjee A, Talapatra S (1955) Alkaloids of the roots of rauwolfia densiflora benth. and hook, rauwolfia perakensis King and gamble, rauwolfia canescens Linn. and rauwolfia serpentina benth. *Naturwissenschaften* 77: 3551–3553.
- Hochstein F, Murai K, Boegemann W (1955) Alkaloids of rauwolfia heterophylla. *J Am Chem Soc* 77(13): 3551–3554.
- Roland M (1959) [The alkaloids of rauwolfia obscura K. schum]. *J Pharm Belg* 14: 347–364.
- Sheludko Y, Gerasimenko I, Kolshorn H, Stackigt J (2002) New alkaloids of the sarpagine group from rauwolfia serpentina hairy root culture. *J Nat Prod* 65(7): 1006–1010.
- van de Venter M, Roux S, Bungu LC, Louw J, Crouch NR, et al. (2008) Antidiabetic screening and scoring of 11 plants traditionally used in south africa. *J Ethnopharmacol* 119(1): 81–86.
- Pinent M, Blay M, Bladé MC, Salvadó MJ, Arola L, et al. (2004) Grape seed-derived procyanidins have an antihyperglycemic effect in streptozotocin-induced

- diabetic rats and insulinomimetic activity in insulin-sensitive cell lines. *Endocrinology* 145(11): 4985–4990.
49. Montagut G, Onnockx S, Vaque M, Blade C, Blay M, et al. (2010) Oligomers of grape-seed procyanidin extract activate the insulin receptor and key targets of the insulin signaling pathway differently from insulin. *J Nutr Biochem* 21(6): 476–481.
 50. Song E, Hur H, Han M (2003) Epigallocatechin gallate prevents autoimmune diabetes induced by multiple low doses of streptozotocin in mice. *Arch Pharm Res* 26(7): 559–563.
 51. Fu Z, Zhen W, Yuskavage J, Liu D (2011) Epigallocatechin gallate delays the onset of type 1 diabetes in spontaneous non-obese diabetic mice. *Br J Nutr* 105(8): 1218–1225.
 52. Benn MH, Shustov G, Shustova L, Majak W, Bai Y, et al. (1996) Isolation and characterization of two guanidines from *galega orientalis lam. cv. gale* (fodder galega). *J Agric Food Chem* 44(9): 2779–2781.
 53. Michel K, Sandberg F, Haglid F, Norin T (1967) Alkaloids of *haloxylon salicornicum* (moq.-tand.) boiss. *Acta Pharm Suec* 4(2): 97–116.
 54. Brack A (1962) Verlauf der Alkaloidbildung durch den Clavicepsstamm von *Pennisetum typhoideum* Rich. in saprophytischer Kultur. 54. Mitteilung über Mutterkornalkaloide. *Arch Pharm Pharm Med Chem* 295(7): 510–515.
 55. Fujii M, Takei I, Umezawa K (2009) Antidiabetic effect of orally administered conophylline-containing plant extract on streptozotocin-treated and gotokakizaki rats. *Biomed Pharmacother* 63(10): 710–716.
 56. Yoshikawa M, Nakamura S, Ozaki K, Kumahara A, Morikawa T, et al. (2007) Structures of steroidal alkaloid oligoglycosides, robencosides A and B, and antidiabetogenic constituents from the brazilian medicinal plant *solanum lycocarpum*. *J Nat Prod* 70(2): 210–214.
 57. Villaseñor I, Lamadrid M (2006) Comparative anti-hyperglycemic potentials of medicinal plants. *J Ethnopharmacol* 104(1–2): 129–131.
 58. Kar DM, Maharana L, Pattnaik S, Dash GK (2006) Studies on hypoglycaemic activity of *solanum xanthocarpum* schrad. & wendl. fruit extract in rats. *J Ethnopharmacol* 108(2): 251–256.
 59. Mosihuzzaman M, Nahar N, Ali L, Rokeya B, Khan AK, et al. (1994) Hypoglycemic effects of three plants from eastern himalayan belt. *Diabetes Res* 26(3): 127–138.
 60. Semwal DK, Rawat U, Semwal R, Singh R, Singh GJP (2010) Anti-hyperglycemic effect of 11-hydroxypalmitine, a palmitate derivative from *stephania glabra* tubers. *J Asian Nat Prod Res* 12(2): 99–105.
 61. Tsutsumi T, Kobayashi S, Liu YY, Kontani H (2003) Anti-hyperglycemic effect of fangchinoline isolated from *stephania tetrandra* radix in streptozotocin-diabetic mice. *Biol Pharm Bull* 26(3): 313–317.
 62. Dixon SL, Smondyrev AM, Knoll EH, Rao SN, Shaw DE, et al. (2006) PHASE: A new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. methodology and preliminary results. *J Comput Aided Mol Des* 20(10–11): 647–671.
 63. Zsoldos Z, Reid D, Simon A, Sadjad SB, Johnson AP (2007) eHiTS: A new fast, exhaustive flexible ligand docking system. *J Mol Graph Model* 26(1): 198–212.
 64. Liu T, Lin Y, Wen X, Jorissen RN, Gilson MK (2007) BindingDB: A web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res* 35(Database issue): D198–201.
 65. Duan J, Dixon SL, Lowrie JF, Sherman W (2010) Analysis and comparison of 2D fingerprints: Insights into database screening performance using eight fingerprint methods. *J Mol Graph Model* 29(2): 157–170.
 66. Kelley LA, Gardner SP, Sutcliffe MJ (1996) An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies. *Protein Eng* 9(11): 1063–1065.
 67. Vuksan V, Sievenpiper JL (2005) Herbal remedies in the management of diabetes: Lessons learned from the study of ginseng. *Nutr Metab Cardiovasc Dis* 15(3): 149–160.
 68. Shukla K, Narain JP, Puri P, Gupta A, Bijlani RL, et al. (1991) Glycaemic response to maize, bajra and barley. *Indian J Physiol Pharmacol* 35(4): 249–254.
 69. Ronchetti F, Russo G, Bombardelli E, Bonati A (1971) A new alkaloid from *rauwolfia vomitoria*. *Phytochemistry* 10(6): 1385.
 70. Phan MG, Phan TS, Matsunami K, Otsuka H (2006) Chemical and biological evaluation on scopadulane-type diterpenoids from *scoparia dulcis* of vietnamese origin. *Chem Pharm Bull* 54(4): 546–549.
 71. Costantino L, Raimondi L, Pirisino R, Brunetti T, Pessotto P, et al. (2003) Isolation and pharmacological activities of the *tecoma stans* alkaloids. *Farmaco* 58(9): 781–785.
 72. Torres JL, Bobet R (2001) New flavanol derivatives from grape (*vitis vinifera*) byproducts. antioxidant Aminoethylthio Flavan-3-ol conjugates from a polymeric waste fraction used as a source of flavanols. *J Agric Food Chem* 49(10): 4627–4634.
 73. Takayama H, Okazaki T, Yamaguchi K, Aimi N, Haginiwa J, et al. (1988) Structure of two new diterpene alkaloids, 3-epi-ignavinol and 2,3-dehydrodelcosine. *Chem Pharm Bull (Tokyo)* 36(8): 3210–3212.
 74. Konno C, Murayama M, Sugiyama K, Arai M, Murakami M, et al. (1985) Isolation and hypoglycemic activity of aconitans A, B, C and D, glycans of *aconitum carmichaeli* roots. *Planta Med* (2): 160–161.
 75. Zhang H, Wang X, Lin L, Ding J, Yue J (2007) Indole alkaloids from three species of the *ervatamia* genus: *E. officinalis*, *E. divaricata*, and *E. divaricata gouyahua*. *J Nat Prod* 70(1): 54–59.
 76. Usubillaga A (1988) Solanudine, a steroidal alkaloid from *solanum nudum*. *Phytochemistry* 27(9): 3031.
 77. El Sayed KA, Hamann MT, Abd El-Rahman HA, Zaghoul AM (1998) New pyrrole alkaloids from *solanum sodomeum*. *J Nat Prod* 61(6): 848–850.
 78. Kashiwaba N, Morooka S, Ono M, Toda J, Suzuki H, et al. (1997) Alkaloidal constituents of the leaves of *stephania cepharantha* cultivated in japan: Structure of cephasugine, a new morphinane alkaloid. *Chem Pharm Bull (Tokyo)* 45(3): 545–548.
 79. Beek TAV, Verpoorte R, Svendsen AB (1984) Alkaloids of *tabernaemontana eglandulosa*. *Tetrahedron* 40(4): 737.