Available online at www.sciencedirect.com
ScienceDirect
European Journal of Medicinal Chemistry xx (2007) 1–9
EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY
http://www.elsevier.com/locate/ejmech

Original article

A three-dimensional pharmacophore model for dipeptidyl peptidase IV inhibitors

I-Lin Lu [a,b], Keng-Chang Tsai [c], Yi-Kun Chiang [a], Weir-Torn Jiaang [a], Ssu-Hui Wu [a], Neeraj Mahindroo [a], Chia-Hui Chien [a], Shiow-Ju Lee [a], Xin Chen [a], Yu-Sheng Chao [a], Su-Ying Wu [a,*]

[a] *Division of Biotechnology and Pharmaceutical Research, National Health Research Institutes, 35 Keyan Road, Zhunan Town, Miaoli County 350, Taiwan*
[b] *Graduate Institute of Life Sciences, National Defense Medical Center, Taipei 114, Taiwan*
[c] *Genomics Research Center, Academia Sinica, Taipei 115, Taiwan*

Received 7 May 2007; received in revised form 29 October 2007; accepted 12 November 2007

Abstract

Dipeptidyl peptidase IV (DPP-IV) is a valid drug target for type-2 diabetes and DPP-IV inhibitors have been proven to efficiently improve glucose tolerance. In our study, 3D pharmacophore models were generated using a training set of 22 DPP-IV inhibitors. The best model consisted of important chemical features and mapped well into the active site of DPP-IV. The model gave high correlation coefficients of 0.97 and 0.84 for the training set and the test set, respectively, showing its good predictive ability for biological activity. Furthermore, the pharmacophore model demonstrated the capability to retrieve inhibitors from database with a high enrichment factor of 42.58. All results suggest that the model provides a useful tool for designing novel DPP-IV inhibitors.

Keywords: Diabetes; Pharmacophore; CATALYST; Structure-activity relationship; Dipeptidyl peptidase IV

1. Introduction

Dipeptidyl peptidase IV (DPP-IV, EC 3.4.14.5) is an important drug target for type-2 diabetes. The major function of DPP-IV is to degrade incretins including glucagon-like peptide (GLP-1) and glucose-dependent insulinotropic polypeptide (GIP) [1], which regulate insulin in a strictly glucose-dependent manner. Inhibition of DPP-IV prolongs the in vivo half-life of these incretins, leading to the proposal that the DPP-IV inhibitors could enhance insulin secretion and improve glucose tolerance, making them valuable for treating type-2 diabetes. This concept was supported by experimental results in DPP-IV knockout mice [2] and diabetes mice models treated with DPP-IV inhibitors [3].

DPP-IV is a serine protease that specifically cleaves N-terminal dipeptides from polypeptides with Pro and Ala at the penultimate position [4]. Most DPP-IV inhibitors were designed according to the substrate P1 site structure (occupied by proline), namely the proline-like compounds. The majorities of these are peptide-like compounds and contain cyanopyrrolidine moiety, which forms covalent bond to the catalytic residue Ser630 by the nitrile group. In addition to the proline-like compounds, a variety of non-peptide-like and reversible DPP-IV inhibitors [5] were also discovered via high-throughput screens and offered new recognition motifs to DPP-IV. Several inhibitors are currently in clinical trials [5], and the first DPP-IV inhibitor Januvia (sitagliptin), discovered by Merck & Co., was approved by the FDA (U.S. Food and Drug Administration) as a drug for the treatment of type-2 diabetes. Clinical data show that Januvia offers many potential advantages, including low risk of hypoglycemia, no weight gain, etc. [6]. However, some side effects were also observed with Januvia, including upper respiratory tract

* Corresponding author. Tel.: +886 37 246 166x35713; fax: +886 37 586 456.
E-mail address: suying@nhri.org.tw (S.-Y. Wu).

0223-5234/$ - see front matter © 2007 Published by Elsevier Masson SAS.
doi:10.1016/j.ejmech.2007.11.014

infection, sore throat, and diarrhea [7]. Therefore, it is worth developing a faster and more accurate system to discover DPP-IV inhibitors with better therapeutic profiles.

According to published data and our observation in the lab, the hit rate of DPP-IV inhibitors by high throughput screening (HTS) is very low [8], about 0.012% for actives with 30% inhibition at 10 μM. Due to low hit rate of HTS in identifying DPP-IV inhibitors, virtual screening (VS) emerged as an alternative method to discover novel DPP-IV inhibitors. At present, only two publications have reported the virtual screening results for DPP-IV. Ward et al. [8] developed a virtual screening protocol combining pharmacophore filtering with compound docking to discover DPP-IV inhibitors of low molecular weight. The best compound showed inhibition activity of 81.9% at 30 μM. Rummey et al. [9] presented a pharmacophore constrained docking method by screening a small primary aliphatic amines fragment database to identify fragments that could be placed in S1 and S2 sites of DPP-IV. The best binders showed promising result with IC_{50} of 2.3 μM against DPP-IV. Since these compounds were not very potent, the authors defined them as starting points to design better ones. Therefore, a large opportunity still exists for computer-aided drug design against DPP-IV.

The aim of this study is to develop an accurate and efficient method for discovering DPP-IV inhibitors. A pharmacophore model was constructed based on key chemical features of compounds with DPP-IV inhibitory activity. It provides a rational hypothetical picture of the primary chemical features responsible for activity. Herein, a ligand based 3D pharmacophore for DPP-IV inhibitors was constructed through validation by cost analysis and training-set activity correlation. The resulting pharmacophore was then characterized by its complementary with DPP-IV protein structure, activity prediction power for structurally diverse test-set DPP-IV inhibitors, and database screening capability to see if the model could truly reflect chemical features of DPP-IV inhibitors.

2. Methods

2.1. DPP-IV inhibition assay

Recombinant DPP-IV purification and activity assay was carried out as previously described [10]. The enzyme used in these studies was soluble human protein produced in a baculovirus expression system. To measure the activity of DPP-IV, Gly-Pro-pNA (Gly-Pro-*p*-nitroanilide, Bachem, Torrance, CA, USA) was used as the substrate and cleaved by the enzyme to release the chromogenic *p*-nitroanilide. The activity inhibition assay was carried out at 37 °C in 2 mM Tris–HCl (pH 8.0) buffer with a substrate concentration of 500 μM. The release of *p*-nitroaniline was monitored at 405 nm to calculate the IC_{50} values for inhibitors.

2.2. Compounds preparation

The two-dimensional (2D) chemical structures of the compounds were sketched using ChemDraw Ultra 6.0 (Chembridge Soft Corp. (www.cambridgesoft.Corn), USA) and saved in MDL-sdfile format. Subsequently, they were imported into CATALYST 4.10 (Accelrys Inc. (www.accelrys.com), USA), and converted into corresponding standard 3D structures by using the best conformation model generation method with CHARMm force field and Poling algorithm to ensure energy-minimized conformation for each compound and reduce conformational redundancies. Default parameters were employed in the conformation generation procedure. The conformations with energy higher than 20 kcal/mol from the global minimum were rejected, and the maximum number of conformations was set to 250. For compounds of where the stereochemistry was not known, all possible conformers were generated.

2.3. Pharmacophore generation

Based on the conformations for each training-set compound, the Hypogen algorithm in CATALYST was employed to construct pharmacophore models. Three features were selected for hypotheses generation; hydrophobic, hydrogen bond donor and hydrogen bond acceptor. The HypoRefine function was applied to generate the excluded volumes. All the parameters were set as default except for the uncertainty value was set to 2.4 for compound activity, representing the ratio of the uncertainty range of measured biology activity against the actual activity for each compound.

2.4. Database preparation and searching

For stockroom database preparation, compounds were saved in MDL-sdfile format and then converted to CATALYST database format by DBconvert module using the default parameter settings. Thereafter, selected pharmacophores was used as a query to retrieve active inhibitors from the compound database by the best flexible search method. The performance of hypotheses to a database search was evaluated by the enrichment factor defined as $(H_a/H_t)/(A/D)$, in which D is the number of compounds in the whole database, A is the total number of active compounds in the database, H_t is the number of compounds by X% of the scored and ranked database and H_a is active compounds among H_t. We defined the X% at 10% to compare the database search power of different pharmacophore models.

Maybridge database (Maybrdige Chemical Co., Cornwall, UK), Chembridge database (ChemBridge Corp., San Diego, CA, USA) and Specs database (Specs, Delft, The Netherlands) were prepared following the same procedure as described for stockroom database and screened by flexible search method in CATALYST.

3. Results and discussion

3.1. Pharmacophore generation for DPP-IV inhibitors

The Hypogen module of CATALYST v4.10 was used for pharmacophore generation. A training set of 22 compounds

Please cite this article in press as: I. Lu et al., Eur. J. Med. Chem. (2007), doi:10.1016/j.ejmech.2007.11.014

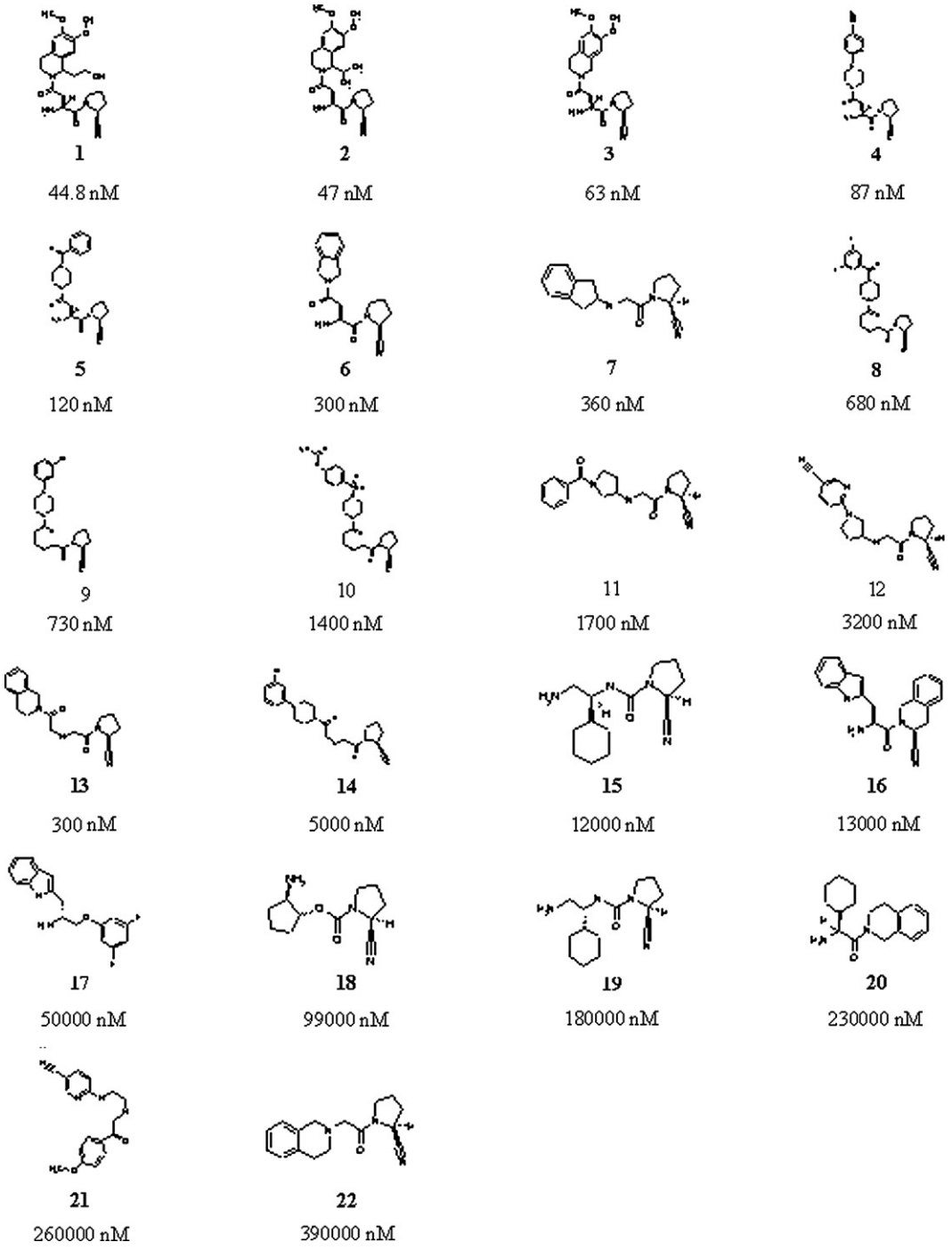

Fig. 1. The chemical structures of 22 training-set inhibitors and their corresponding IC_{50} values against DPP-IV.

with inhibition ranging from 44.8 nM to 390,000 nM was used to generate models (Fig. 1). The training set included 4 compounds found in the literature [11–13] and 18 identified in our laboratories. By analyzing the structures of DPP-IV in complex with valine-pyrrolidine [14], chemical features were selected, including hydrogen bond acceptor, donor, and hydrophobic interactions. In addition, to further evaluate the steric effects that contribute to the biological activities, the features of excluded volumes were also generated by the HypoRefine module and included in the pharmacophore models. Ten top-scored hypotheses were generated and were then subject to cost analysis, which gives an idea about the statistical significance of the hypotheses. Two index values are used for cost analysis: one is the difference between fixed and null costs and the other is the Δcost (the difference between null cost and total cost). The fixed cost represented an ideal hypothesis, which could perfectly predict the activity of compounds in the training set with minimum deviation, while the null cost represented a random hypothesis that was unable to predict activity of the compounds. The difference between these two costs should be $\geq$70 bits to show the over 90% statistical significance of the model. As for the Δcost (the

Please cite this article in press as: I. Lu et al., Eur. J. Med. Chem. (2007), doi:10.1016/j.ejmech.2007.11.014

difference between null cost and total cost), it should be greater than 60 bits to represent a true correlation of the data. The cost difference between null and fixed cost of the run was 107.6926 and the Δcost were more than 60 bits for all ten hypotheses (Table 1). The hypotheses were subject to further evaluation for their capability to predict the activity of the training set compounds. All hypotheses had a correlation coefficient of higher than 0.9, with Hypo1 showed the highest correlation coefficient value of 0.972, demonstrating good predictive ability of these hypotheses (Table 1). In addition, the RMSD (root mean squared deviation) values for all hypotheses were lower than 1.5, further underlining the good predictive quality of these hypotheses (Table 1). Interestingly, all the ten hypotheses share with the same hydrogen-bond acceptor, hydrogen-bond donor and hydrophobic features but different excluded volumes features. Among the ten hypotheses, hypothesis 1 (Hypo1) gave the best statistics; the lowest total cost, the highest Δcost, the highest correlation coefficient and significantly low RMSD value. As a result, Hypo1 was chosen as the best hypothesis for further analysis, which consists of two hydrogen-bond acceptors, one hydrogen-bond donor, one hydrophobic feature and three excluded volumes (Fig. 2).

3.2. Activity prediction and mapping of Hypo1 to training-set compounds

To verify the prediction accuracy of Hypo1 to training set compounds, the activity of each training set compound is estimated by regression analysis. The training-set compounds were separated into three groups: active (≤100 nM, +++), moderately active (100−10,000 nM, ++) and inactive (>10,000 nM) (Table 2). From Table 2, we can see that the estimated activity of all training-set compounds was predicted to the same activity scale to the measured activity. In addition, for feature fitting, the four potent compounds (≤100 nM), **1**, **2**, **3**, and **4**, have the highest fitting score and mapped well to all the chemical features on Hypo1. As for moderately and inactive compounds, at least one chemical feature is miss-mapped. Interestingly, all the highly active compounds mapped into the hydrophobic feature whereas all the moderately and inactive compounds could not fit into the hydrophobic feature except compound **21**, suggesting the importance of the hydrophobic feature to DPP-IV activity. These results indicated that Hypo1 reflects as a reliable model to accurately predict the activities of training-set compounds.

3.3. Validation of Hypo1

Hypo1 was further validated by the CatScramble module in CATALYST to evaluate its statistical significance using the method based on Fisher's randomization test. The validation was performed by generating random spreadsheets among training set molecules, which randomly reassigned activity values to each compounds and subsequently generated pharmacophore hypotheses using the same features and parameters developed for Hypo1. To achieve a confidence level of 99%, 99 random spreadsheets (random hypotheses) needed to be generated (Significance = $[1 - ((1 + 0)/(99 + 1))] * 100\% = 99\%$). The statistics of the lowest 10 random spreadsheets together with Hypo1 data are summarized in Table 3. Results clearly show that the Hypo1 hypothesis was not generated by chance, because its statistics are far more superior to the 10 lowest random hypotheses and hence the other 89 random hypotheses.

As the training-set compounds were cyanopyrrolidine-type compounds (Fig. 1), it was useful to verify whether Hypo1 could also predict the activity of compounds that are structurally distinct from those in the training set. A set of 20 compounds with diverse structures and various activities was selected from literature [5,12,15,16]. Compounds **23**−**42** with inhibition ranging from 2.6 nM to 710 nM are listed in Fig. 3. The compounds were subjected to further analyses by Hypo1. The experimental and predictive activities of Hypo1 as applied to test-set inhibitors are shown in Table 4. The predicted activities for all of the compounds are in the same activity scale with the experimental activities with the exception of three compounds, demonstrating the predicted

Table 1
Results of statistical significance, predictive power to training set compounds and the features of top-10 hypotheses[a]

Hypothesis	Correlation	RMSD[b]	Total cost	Δcost[c]	Features[d,e]				
					HA1	HA2	HD	HP	EX[f]
1	0.972	0.798	93.8863	100.03	✓	✓	✓	✓	3
2	0.947	1.08	99.6034	94.279	✓	✓	✓	✓	
3	0.945	1.11	100.08	93.833	✓	✓	✓	✓	1
4	0.948	1.07	100.675	93.238	✓	✓	✓	✓	2
5	0.936	1.19	104.721	89.192	✓	✓	✓	✓	
6	0.922	1.30	105.388	88.525	✓	✓	✓	✓	
7	0.934	1.22	105.654	88.259	✓	✓	✓	✓	
8	0.918	1.34	106.468	87.445	✓	✓	✓	✓	1
9	0.916	1.35	106.73	87.183	✓	✓	✓	✓	2
10	0.912	1.34	107.858	86.055	✓	✓	✓	✓	

[a] Fixed cost of top-10 scoring hypotheses is 86.6204. Null cost value is 193.913 and Configuration cost value is 16.4911.
[b] RMSD, the deviation of the log (estimated activities) from the log (measured activities) normalized by the log (uncertainties).
[c] Δcost = null cost − total cost.
[d] HA, hydrogen bond acceptor; HD, hydrogen bond donor; HP, hydrophobic feature; EX, number of excluded volumes.
[e] ✓: The feature was included in the hypothesis.
[f] Number of excluded volumes of the hypothesis.

Please cite this article in press as: I. Lu et al., Eur. J. Med. Chem. (2007), doi:10.1016/j.ejmech.2007.11.014

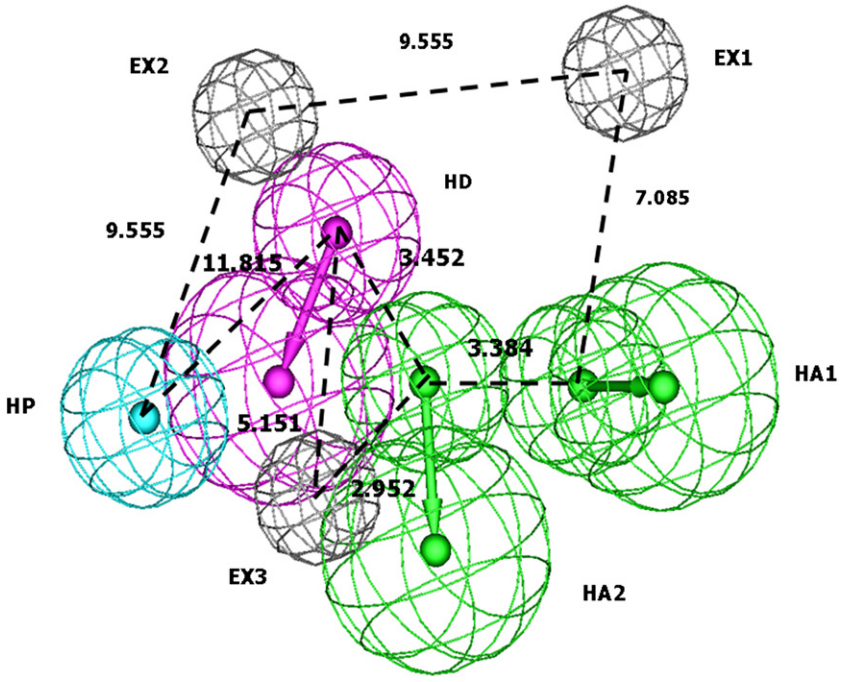

Fig. 2. The features of best hypothesis model Hypo1. Hypo1 consists of three excluded volumes (EX, black), two hydrogen-bond acceptors (HA, green), one hydrogen-bond donor (HD, magenta) and one hydrophobic feature (HP, cyan).

accuracy of 85% for the Hypo1. Moreover, the correlation coefficients (r) for the test-set given by Hypo1 were 0.84. The results suggest that Hypo1 is not just fit for cyanopyrrolidine-type of compounds but also to more structurally diverse compounds. Combined with the training-set compounds prediction results, these results highlight that the activities predicted by Hypo1 are in agreement with the experimental data, revealing the predictive power of Hypo1.

Table 2
The prediction power and mapping of Hypo1 to training-set compounds

Compound	True IC_{50} (nM)	Estimated IC_{50} (nM)[a]	Error factor[b]	Fit value[c]	Activity scale[d]	Estimated activity scale	Mapped features			
							HA1	HA2	HD	HP
1	45	37	−1.2	7.84	+++	+++	✓	✓	✓	✓
2	47	29	−1.6	7.94	+++	+++	✓	✓	✓	✓
3	63	56	−1.1	7.66	+++	+++	✓	✓	✓	✓
4	87	100	+1.2	7.40	+++	+++	✓	✓	✓	✓
5	120	130	+1.1	7.29	++	++	✓	✓	✓	
6	300	1300	+4.3	6.30	++	++	✓	✓	✓	
7	360	1500	+4.1	6.25	++	++	✓	✓	✓	
8	680	1400	+2.2	6.24	++	++	✓	✓	✓	
9	730	1300	+1.9	6.26	++	++	✓	✓	✓	
10	1400	1300	−1.1	6.30	++	++	✓	✓	✓	
11	1700	1300	−1.4	6.30	++	++	✓	✓	✓	
12	3200	2400	−1.3	6.02	++	++	✓	✓	✓	
13	3300	1200	−2.8	6.33	++	++	✓	✓	✓	
14	5000	1400	−3.5	6.26	++	++	✓	✓	✓	
15	12000	13000	+1.1	5.28	+	+	✓	✓	✓	
16	13000	15000	+1.1	5.25	+	+	✓	✓	✓	
17	50000	100000	+2	4.40	+	+		✓	✓	
18	99000	120000	+1.2	4.32	+	+	✓	✓	✓	
19	180000	140000	−1.3	4.25	+	+		✓	✓	
20	230000	280000	+1.2	3.97	+	+	✓		✓	
21	260000	98000	−2.6	4.42	+	+	✓			✓
22	390000	150000	−2.6	4.23	+	+	✓	✓		

[a] The value was predicted by Hypo1.
[b] The error factor is calculated as the ratio of the measured activity to the estimated activity or the inverse if estimated is greater than measured. If the measured activity is higher than estimated activity, the error factor is negative and vice versa.
[c] Fit value calculated the geometry fitting between the hypothesis to compounds. The higher the value, the better the fitting.
[d] Activity scale: highly active, +++, $IC_{50} \leq 100$ nM; moderately active, ++, 100 nM $< IC_{50} <$ 10,000 nM; inactive, +, $IC_{50} >$ 10,000 nM.

Please cite this article in press as: I. Lu et al., Eur. J. Med. Chem. (2007), doi:10.1016/j.ejmech.2007.11.014

Table 3
Validation of Hypo4-2 using CatScramble program

Validation no.	Total cost	Fixed cost	RMSD	Correlation coefficient	Configuration cost
Hypo1	93.8863	86.6204	0.798	0.97	16.49
Trial 01	101.038	81.5147	1.186	0.94	11.39
Trial 02	116.722	81.6879	1.510	0.90	11.56
Trial 03	117.801	83.1107	1.688	0.87	12.98
Trial 04	117.868	83.3564	1.482	0.91	13.23
Trial 05	118.43	80.3449	1.708	0.87	10.22
Trial 06	118.722	81.4211	1.718	0.86	11.29
Trial 07	120.275	81.0036	1.766	0.86	10.87
Trial 08	120.36	84.4023	1.618	0.88	14.27
Trial 09	120.401	83.6571	1.631	0.88	13.53
Trial 10	121.161	83.2135	1.770	0.85	13.08

Since the purpose of this study is not only to construct the pharmacophore hypothesis to predict compound activities, but also to employ the hypothesis model on virtual screening to search for compound hits, the capability of Hypo1 to retrieve active compounds from large database was further examined. We prepared so-called stockroom databases (DB), consisting of 57 DPP-IV inhibitors collected from the public source [17–20], 30 DPP-IV inhibitors and 62,979 inactive compounds tested in our laboratory. Hypo1 was then applied to screen DB using the BEST flexible search algorithm implemented in CATALYST. Screening result was reported as enrichment factors to evaluate the effectiveness of Hypo1. Enrichment factors assess the screening power of the hypotheses to retrieve actives from a large database [21]; its definition is given in Section 2. Higher enrichment factors correspond

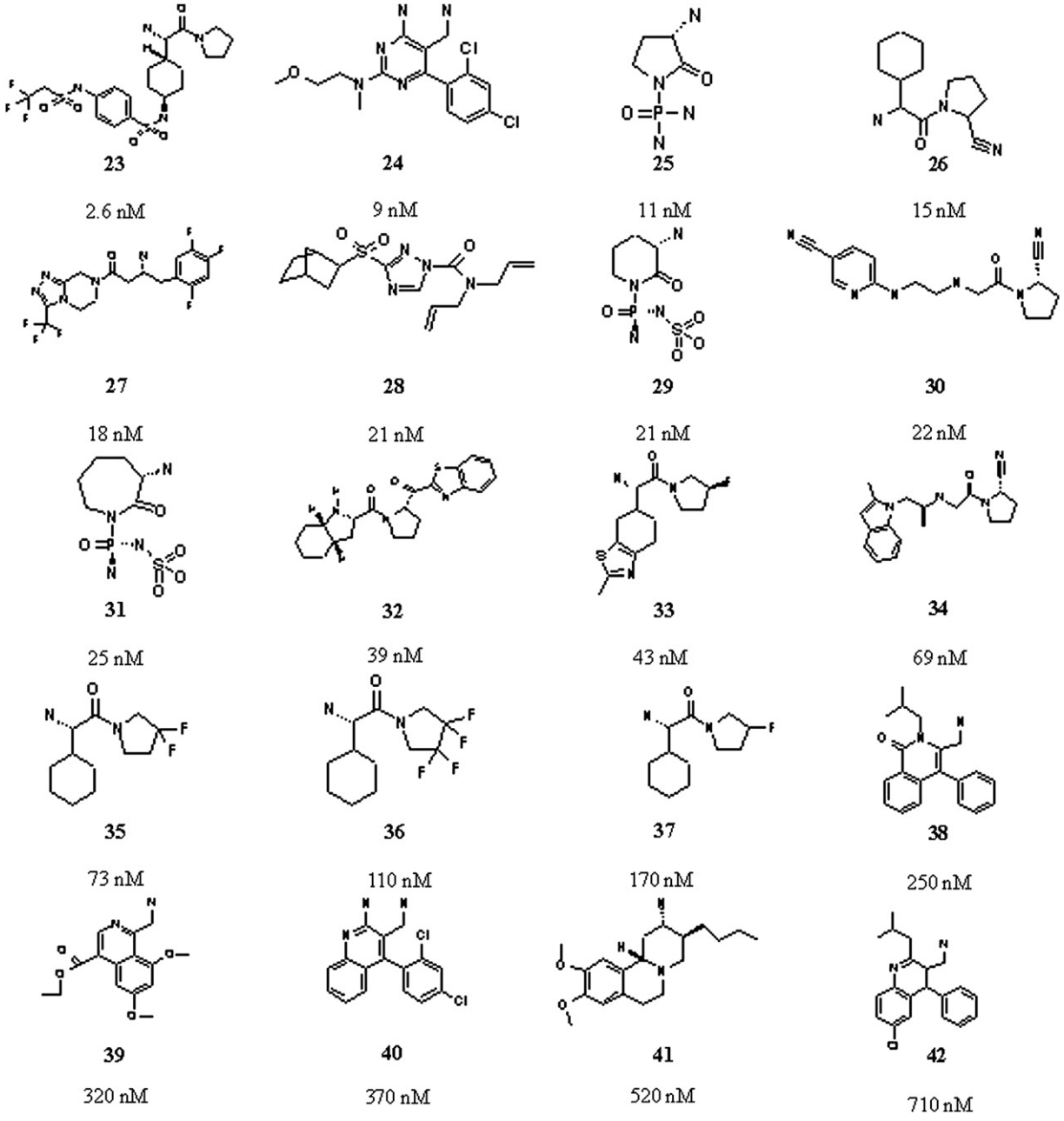

Fig. 3. The chemical structures of 20 test-set inhibitors and their corresponding IC_{50} values against DPP-IV.

Table 4
The prediction power of Hypo1 to test-set compounds[a]

Compound	True IC_{50} (nM)	Estimated IC_{50} (nM)	Error factor[b]	Activity scale[c]	Estimated activity scale
23	2.6	1.7	−1.5	+++	+++
24	9	46	+5.1	+++	+++
25	11	10	−1.1	+++	+++
26	12	13	+1.0	+++	+++
27	18	120	+6.4	+++	++
28	21	11	−2	+++	+++
29	21	11	−1.8	+++	+++
30	22	35	+1.6	+++	+++
31	25	11	−2.3	+++	+++
32	39	110	+2.9	+++	++
33	43	31	−1.4	+++	+++
34	69	67	−1	+++	+++
35	73	130	+1.7	+++	++
36	110	130	+1.2	++	++
37	170	130	−1.3	++	++
38	250	200	−1.2	++	++
39	320	170	−1.9	++	++
40	370	860	+2.3	++	++
41	520	180	−2.9	++	++
42	710	200	−3.5	++	++

[a] Correlation coefficient: 0.84.
[b] The error factor is calculated as the ratio of the measured activity to the estimated activity or the inverse if estimated is greater than measured. If the measured activity is higher than estimated activity, the error factor is negative and vice versa.
[c] Activity scale: highly active, +++, $IC_{50} \leq 100$ nM; moderately active, ++, 100 nM $< IC_{50} <$ 10,000 nM; inactive, +, $IC_{50} >$ 10,000 nM.

to more efficient hypotheses in database screening. In this study, the top 10% of scored and ranked DB compounds were used to calculate enrichment factors for Hypo1. The enrichment factors of Hypo1 was 42.58, which retrieved 37 out of 87 DPP-IV inhibitors in the top 10% of the ranked database, suggesting that Hypo1 could retrieve DPP-IV inhibitors with satisfactory enrichment toward database.

Finally, Hypo1 was applied to screen the Maybridge, Chembridge and Specs databases consisting of 510,170 compounds. Sixty-five compounds fitted into all features described by Hypo1 were predicted as highly active compounds ($IC_{50} \leq 100$ nM) and could be worth exploring further. They have been classified into several chemotypes and some examples are shown in figure 1 in the supplementary data.

Another approach to search for new DPPIV inhibitors by screening chemical databases is the high-throughput docking method. Two studies have applied the docking method to discover novel DPP-IV inhibitors [8,9]. However, these studies all incorporated pharmacophore constraints in docking procedures. This could be due to the specific features required for DPP-IV inhibitors in order to accommodate S1 site and S2 site. These features may not be automatically recognized by docking programs, thus result in the low hit rate of high-throughput docking to DPP-IV. In our study, Hypo1 could accurately describe the key features of DPP-IV inhibitors and also retrieve actives from chemical database with high enrichment. Therefore, the pharmacophore search using Hypo1 is expected to identify compounds with important chemical features essential for DPPIV inhibition. In addition, pharmacophore searching could screen a large chemical database at high speed. However, it is worthwhile to note that pharmacophore searching might decrease the diversity of the compound subset as they are biased by the chemical features of known inhibitors. New interactions potentially favorable for inhibition might be neglected. Therefore, it would be useful to combine pharmacophore searching with high-throughput docking method to search for novel and potent DPPIV inhibitors. The pharmacophore searching of Hypo1 could be firstly used as a filter to remove compounds lacking essential features for DPP-IV binding. The filtered compound database would be subsequently subject to docking programs to carry out extensive docking to explore other potential chemical space in the protein. Furthermore, the top-score compounds from docking screening could be evaluated by Hypo1 to predict the activities in prior to biological testing to increase the hit rate and efficiency in computational compound screening.

3.4. Mapping of Hypo1 onto the protein active site and inhibitors

Hypo1 was characterized by two hydrogen bond acceptors, one hydrogen bond donor, one hydrophobic feature and three excluded volumes. All features were located at defined positions and were surrounded by a spatial tolerance sphere of 1 Å radius (Fig. 2).

Hypo1 hypothesis was superimposed on the active site of DPP-IV for comparison. As shown in Fig. 4A, the two hydrogen bond acceptor features, represented by green spheres, perfectly directed to two hydrogen bond donor residues, Arg125 and Ser630. The hydrogen bond donor feature, shown in magenta sphere, was inserted into the area between side chain of Glu205 and Glu206 that served as hydrogen acceptors. Additionally, the hydrophobic feature, represented by cyan sphere, was close to the Phe357 and implied the preferred hydrophobic interaction of DPP-IV inhibitors to Phe357. As mentioned before, this hydrophobic feature was critical for DPP-IV inhibitor activity, revealing inhibitors that interact with Phe357 could be potent DPP-IV inhibitors. As for the three excluded volumes, two of them (EX1 and EX2) were located near the S1 pocket consisting of residues Tyr662 and Tyr666. Based on the biological assay data and crystal structure of DPP-IV [16], only amino acids with smaller side chains (proline, alanine or glycine) will be able to fit into this narrow pocket. Amino acids with bulky side chains were not able to access to this pocket due to their steric effects with Thyr662 and Tyr666. Thereby, the space arrangement of EX1 and EX2, consistent with the function of Tyr662 and Tyr666, is to restrict the size of DPP-IV inhibitors at this region. The third excluded volume (EX3) was located between Arg125 and Glu205 that limited inhibitors to adopt the conformation favored for forming hydrogen bond interactions to Arg125, Glu205 and Glu206. In conclusion, Hypo1 possessed the key features required for DPP-IV inhibition activity and the spatial arrangement of these features complement well with the important residues, including Arg125, Glu205, Glu206, Ser630, Tyr662, Tyr666 and Phe357. These results consisted with the

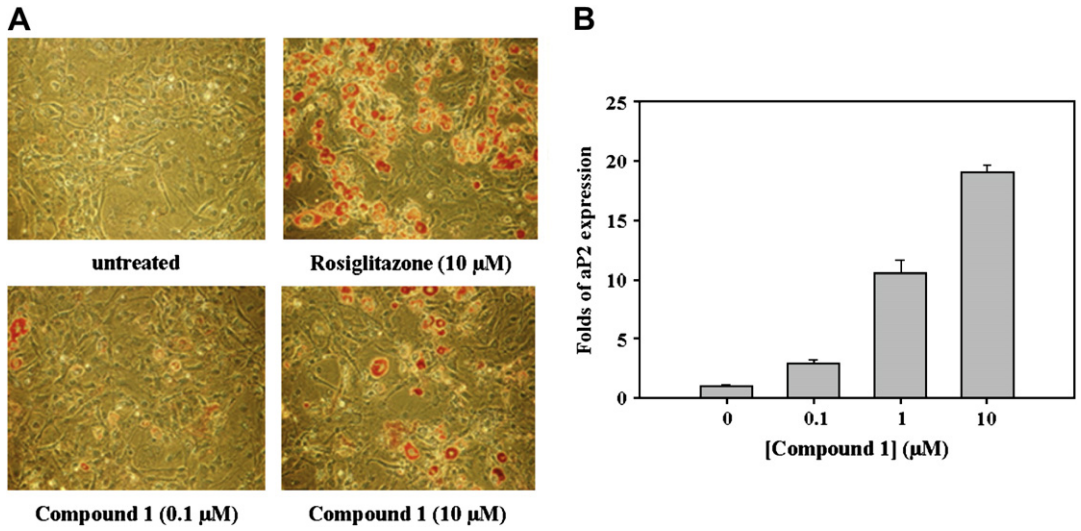

Fig. 4. Mapping of Hypo1 to protein structure and selected compounds. (A) Mapping of Hypo1 onto the DPP-IV active site. The chemical features of Hypo1 were complementary to DPP-IV active site. The two H-bond acceptor and one H-bond donor features made close contact to the opposite donor residues Arg125 and Ser630, and acceptor residues Glu205 and Glu206. The hydrophobic feature was close to Phe357. Two of the three excluded volumes were mapped to the residues Tyr662 and Tyr666 and the third one located between Arg125 and Glu205. Protein residues were shown in grey, and the H-bond acceptor, donor and hydrophobic features of Hypo1 were shown in green, magenta, cyan and black individually. (B) Mapping of Hypo1 to selected compounds. The compounds are shown in gray and the Hypo1 features colors are the same as shown in (A).

observation by Engel et al. [22], suggesting these key residues contributed the tight interactions for clamping inhibitors to the DPP-IV active site.

To further explore the features of the pharmacophore, Hypo1 was mapped onto the structures of DPP-IV inhibitors. Three-dimensional mapping of Hypo1 onto the most active compounds in the training-set (**1**) and test-set (**23**) are depicted in Fig. 4B. For compound **1**, the hydrophobic feature was mapped onto the methyl group and the hydrogen-bond donor feature was placed over the N atom of the amine group. The two hydrogen-bond acceptor features correctly located on the nitrile and carbonyl groups. For compound **23**, the O atom of the carbonyl group was mapped to the hydrogen bond acceptor feature. The amine NH_2 group was located to the hydrogen bond donor feature while the trifluoromethyl group overlapped the hydrophobic feature.

4. Conclusion

In this study, pharmacophore models were constructed and carefully validated. The best output hypothesis, Hypo1, consisted of 7 features, including two hydrogen-bond acceptors, one hydrogen-bond donor, one hydrophobic feature and three excluded volumes. The features were perfectly complementary to the DPP-IV active site and directed to relative protein residues, showing that Hypo1 represents the characteristics of DPP-IV active sites. For predicting activity, the correlation coefficient of Hypo1 with training and test DPP-IV inhibitors sets were 0.97 and 0.84 respectively (Table 5), suggesting a good predictive power of the model for the majority of DPP-IV inhibitors. When further applied to large database screening, enrichment factor of the ranked top 10% lists to the constructed database were greater than 40-fold (Table 5), indicating that actives can be retrieved from databases efficiently, making Hypo1 a useful virtual screening tool for discovery of novel DPP-IV inhibitors. In conclusion, we generated a pharmacophore that truly reflects the features of DPP-IV inhibitors. The pharmacophore could be used as a fast and accurate tool to assist discovery of novel DPP-IV inhibitors either at the hit discovery (database screening) or lead optimization (activity prediction) stages.

Table 5
Summary of Hypo1 results

Hypo1						
Statistics value	**Null cost**	**Fixed cost**	**Total cost**	**Δcost**[a]	**Configuration cost**	**RMS**
	193.913	86.6204	93.8863	100.03	16.4911	0.798
Correlation coefficient	**Training set**	**Test set**				
	0.97	0.84				
Database screen[b] **(enrichment factor)**	**DB**					
	42.58					

[a] Δcost was define as (null cost) − (total cost).
[b] The enrichment factor was calculated for the ranked top10% list of DB.

Please cite this article in press as: I. Lu et al., Eur. J. Med. Chem. (2007), doi:10.1016/j.ejmech.2007.11.014

Acknowledgments

The authors thank Ms. Hsiao-Wen Edith Chu, Ms. Chi-Wen Chen, Mr. Tai-Tsung Chen, Ms. Hsiu-Hsiu Huang, Ms. Huai-Tzu Chang and Ms. Pey-Yea Yang for administrative support, and the National Health Research Institutes, Taiwan for financial support.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version, at doi:10.1016/j.ejmech.2007.11.014.

References

[1] R. Mentlein, B. Gallwitz, W.E. Schmidt, Eur. J. Biochem. 214 (1993) 829–835.
[2] D. Marguet, L. Baggio, T. Kobayashi, A.M. Bernard, M. Pierres, P.F. Nielsen, U. Ribel, T. Watanabe, D.J. Drucker, N. Wagtmann, Proc. Natl. Acad. Sci. U.S.A. 97 (2000) 6874–6879.
[3] B. Ahren, E. Simonsson, H. Larsson, M. Landin-Olsson, H. Torgeirsson, P.-A. Jansson, M. Sandqvist, P. Bavenholm, S. Efendic, J.W. Eriksson, S. Dickinson, D. Holmes, Diabetes Care 25 (2002) 869–875.
[4] A. Yaron, F. Naider, Crit. Rev. Biochem. Mol. Biol. 28 (1993) 31–81.
[5] D. Hunziker, M. Hennig, J.U. Peters, Curr. Top. Med. Chem. 5 (2005) 1623–1637.
[6] S.A. Miller, E.L. St. Onge, Ann. Pharmacother. 40 (2006) 1336–1343.
[7] A. Barnett, Int. J. Clin. Pract. 60 (2006) 1454–1470.
[8] R.A. Ward, T.D. Perkins, J. Stafford, J. Med. Chem. 48 (2005) 6991–6996.
[9] C. Rummey, S. Nordhoff, M. Thiemann, G. Metz, Bioorg. Med. Chem. Lett. 16 (2006) 1405–1409.
[10] C.H. Chien, L.H. Huang, C.Y. Chou, Y.S. Chen, Y.S. Han, G.G. Chang, P.H. Liang, X. Chen, J. Biol. Chem. 279 (2004) 52338–52345.
[11] D.M. Ashworth, B. Atrash, G.R. Baker, A.J. Baxter, P.D. Jenkins, D.M. Jones, M. Szelke, Bioorg. Med. Chem. Lett. 6 (1996) 1163–1166.
[12] I.L. Lu, S.J. Lee, H. Tsu, S.Y. Wu, K.H. Kao, C.H. Chien, Y.Y. Chang, Y.S. Chen, J.H. Cheng, C.N. Chang, T.W. Chen, S.P. Chang, X. Chen, W.T. Jiaang, Bioorg. Med. Chem. Lett. 15 (2005) 3271–3275.
[13] W.T. Jiaang, Y.S. Chen, T. Hsu, S.H. Wu, C.H. Chien, C.N. Chang, S.P. Chang, S.J. Lee, X. Chen, Bioorg. Med. Chem. Lett. 15 (2005) 687–691.
[14] H.B. Rasmussen, S. Branner, F.C. Wiberg, N. Wagtmann, Nat. Struct. Biol. 10 (2003) 19–25.
[15] T.-Y. Tsai, M.S. Coumar, T. Hsu, Hsing-Pang Hsieh, C.-H. Chien, C.-T. Chen, C.-N. Chang, Y.-K. Lo, Ssu-Hui Wu, C.-Y. Huang, Y.-W. Huang, M.-H. Wang, H.-Y. Wu, H.-J. Lee, X. Chen, Y.-S. Chao, W.-T. Jiaang, Bioorg. Med. Chem. Lett. 16 (2006) 3268–3272.
[16] T. Hsu, X. Chen, C.T. Chen, S.J. Lee, C.N. Chang, K.H. Kao, M.S. Coumar, Y.T. Yeh, C.H. Chien, H.S. Wang, K.T. Lin, Y.Y. Chang, S.H. Wu, Y.S. Chen, I.L. Lu, S.Y. Wu, T.Y. Tsai, W.C. Chen, H.P. Hsieh, Y.S. Chao, W.T. Jiaang, J. Med. Chem. 49 (2006) 373–380.
[17] F. Himmelsbach, E. Langkopf, M. Eckhardt, M. Mark, R. Maier, R.R.H. Lotz, M. Tadayyon, Germany patent WO 2004018468. 2004.
[18] K.L. Longenecker, K.D. Stewart, D.J. Madar, C.G. Jakob, E.H. Fry, S. Wilk, C.W. Lin, S.J. Ballaron, M.A. Stashko, T.H. Lubben, H. Yong, D. Pireh, Z. Pei, F. Basha, P.E. Wiedeman, T.W. von Geldern, J.M. Trevillyan, V.S. Stoll, Biochemistry 45 (2006) 7474–7482.
[19] H.-U. Demuth, M. Hoffmann, T. Hoffmann, A.J. Niestroj, S. Schilling, U. Heiser, UK PCT Int. Appl. (2004).
[20] C.D. Haffner, D.L. McDougald, J.M. Lenhard, PCT Int. Appl. (2003).
[21] D.A. Pearlman, P.S. Charifson, J. Med. Chem. 44 (2001) 502–511.
[22] M. Engel, T. Hoffmann, S. Manhart, U. Heiser, S. Chambre, R. Huber, H.-U. Demuth, J. Mol. Biol. (2006) 768–783.