Available online at www.sciencedirect.com
SCIENCE DIRECT®
European Journal of Medicinal Chemistry 40 (2005) 632–640
ELSEVIER
EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY
www.elsevier.com/locate/ejmech

Construction of a virtual combinatorial library using SMILES strings to discover potential structure-diverse PPAR modulators

Chenzhong Liao [a,b,*], Bing Liu [b], Leming Shi [a], Jiaju Zhou [b], Xian-Ping Lu [a]

[a] *Chipscreen Biosciences, Ltd. Research Institute of Tsinghua University, Suite C301, P.O. Box 28, Shenzhen, Guangdong 518057, China*
[b] *Institute of Process Engineering, Chinese Academy of Sciences, Beijing 100080, China*

Received 2 September 2004; accepted 22 February 2005

Abstract

Based on the structural characters of PPAR modulators, a virtual combinatorial library containing 1226,625 compounds was constructed using SMILES strings. Selected ADME filters were employed to compel compounds having poor drug-like properties from this library. This library was converted to sdf and mol2 files by CONCORD 4.0, and was then docked to PPARγ by DOCK 4.0 to identify new chemical entities that may be potential drug leads against type 2 diabetes and other metabolic diseases. The method to construct virtual combinatorial library using SMILES strings was further visualized by Visual Basic.net that can facilitate the needs of generating other type virtual combinatorial libraries.

Keywords: Virtual combinatorial library; SMILES; PPAR; Virtual screening; Drug discovery

1. Introduction

The needs to generate large amounts of structure-diverse new chemical entities or drug-like chemical moieties are essential for a successful drug discovery program in modern pharmaceutical setting. Combinatorial chemistry is far more rapidly than is possible using conventional approaches to facilitate the identification of biologically active molecules in drug discovery efforts [1–3]. Depending on their use, combinatorial libraries can be divided into two main categories: (1) a library directly focused on a specific target, structural class, or known pharmacophore, and (2) a probe library that is target-independent and is designated to span a wide range of physicochemical and structural characteristics.

Virtual combinatorial chemistry is a conceptually different approach that rests on supramolecular chemistry [4]. It relies on a reversible connection process for the spontaneous and continuous generation of all possible combinations of a set of basic components, thus making virtually available of all the structural and interaction features these combinations may present. There are two ways to construct virtual combinatorial libraries. Briefly, the first one is based on the application of a Markush structure, which represents a common scaffold with variation sites labeled as R-groups, each of which is associated with a list of alternatives [5]. The second one is rather more flexible: it encodes a reaction as a chemical “transform”. The transform specifies the parts of the reacting molecules that undergo chemical transformations and nature of these transformations. This approach mimics more closely the steps that are involved in actual synthesis, does not require a common template or the generation of clipped reagents, and can be applied to a broad spectrum of chemical reaction [6,7]. Accordingly, the programs for constructing virtual combinatorial libraries such as Legion [8] and CombiLibMaker [9] operate in either of the two modes that correspond to one of the two combinatorial processes: (1) the “core + side chains”, and (2) the “combine reactants”. They are generally referred as product-based and reactant-based approaches, respectively, in which the former one appears superior to the latter.

Simplified Molecular Input Line Entry Specification (SMILES) proposed by Weininger [10] in 1988 is a simple yet comprehensive chemical notation for specifying molecules or molecular fragments. Comparing with mol and other

---

* Corresponding author. Tel.: +86 755 2695 7737; fax: +86 755 2695 7291.
E-mail address: czliao@chipscreen.com (C. Liao).

0223-5234/$ - see front matter © 2005 Elsevier SAS. All rights reserved.
doi:10.1016/j.ejmech.2005.02.006

file formats, SMILES is broadly used in saving chemical molecular structures and data exchange due to its very compact volume.

Upon construction of the virtual combinatorial library using above-mentioned methods, virtual screening [11–14], a powerful computational filter for reducing the size of a chemical library, can then be applied to identify new leads from the designed library based on a given three-dimensional structure of the target. Currently, there are several docking programs available for the screening, for instance, the FlexX [15], AutoDock [16], Gold [17] et al., but the most well-known and applied docking program is DOCK [18].

Peroxisome proliferator activated receptor-γ (PPARγ), PPARα and PPARδ are ligand-activated transcriptional factors belonging to the nuclear hormone receptor superfamily, which are essential in control of lipid, glucose, and energy homeostasis [19]. At present, PPARγ agonist rosiglitazone has been successfully prescribed for the patients having type 2 diabetes, and other selective agonists, partial agonists and antagonists of PPARγ, dual agonists of PPARγ and PPARα are in clinical phases of the development as anti-type 2 diabetic and metabolic syndrome treatments. Because of the side effects such as weight gain and water retention associated with agonists of PPARγ [20] and the potential fat burning activity of PPARα and PPARδ [21], balanced triple agonists for PPARα/γ/δ or selective agonists for PPARα or δ are also the pursuits of pharmaceutical companies.

In this study, we have constructed a virtual combinatorial library (anti-type 2 diabetic virtual combinatorial library, A2DVCL) using SMILES strings to represent the fragments of backbones and side chains according to the structural characters of previously published PPAR modulators. Permuting all side chains at the various attachment points of the backbones generated all of the allowed products. The resultant compounds were then filtered by the selected ADME properties that gave a total of 1226,625 molcules in the library. The library was converted to sdf and mol2 files by the CONCORD 4.0 program and screened virtually by DOCK 4.0 program when setting PPARγ as the target to identify potential new chemical entities in the management of type 2 diabetes and other metabolic diseases. This method to construct virtual combinatorial library using SMILES strings was further visualized by the software of Visual Basic.net to meet the needs in generating other type virtual combinatorial libraries.

2. Methods

2.1. Structural characters of PPAR modulators

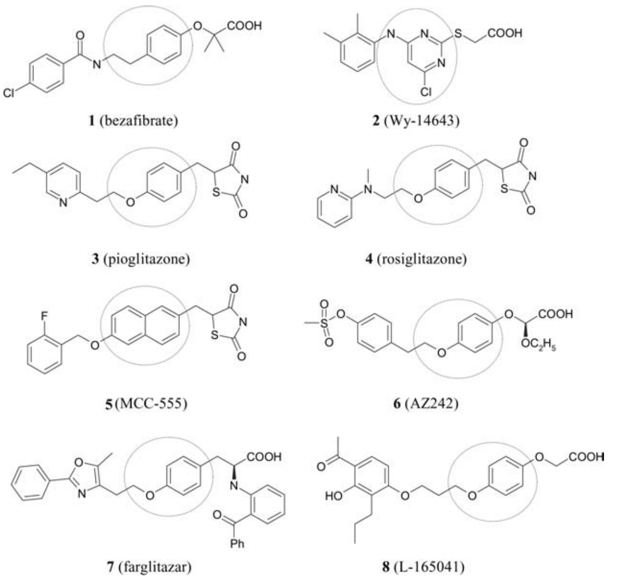

1 (bezafibrate)
2 (Wy-14643)
3 (pioglitazone)
4 (rosiglitazone)
5 (MCC-555)
6 (AZ242)
7 (farglitazar)
8 (L-165041)

Fig. 1. Examples of PPAR modulators.

Examples of the structures of PPAR modulators are shown in Fig. 1 where compounds **1**, **2** are agonists of PPARα; compounds **3**, **4**, **5** are agonists of PPARγ; compounds **6**, **7** are dual agonists of PPARα/γ; compound **8** is a modulator of PPARδ. Kurogi has initially separated the « possible » pharmacophoric structure of thiazolidinediones (TZDs), the typical compounds are **3** and **4**, into the binding and the effector site [22]. That is, the binding site is the thiazolidinedione moiety essential in all TZDs, and the effector site is a secondary region that modifies the biological potency. And there is a linker between the binding and the effector sites, as shown in Fig. 2. Besides TZDs, the pharmacophoric characters existed in most of PPARγ's agonists, a number of PPARγ's partial agonists and antagonist, PPARα and PPARδ's modulators can also be considered to constitute these three partitions (here we called them as binder, linker and effector, respectively). This can be seen in Fig. 1: except compound **2** (Wy-14643), all of the remaining seven compounds are thought to have this trait. Because of the different binding sites and binding modes, there are several kinds of PPAR modulators do not possess such properties, for instance GW 0072 [23], a highly flexible partial agonist of PPARγ.

As to PPARγ's agonists, the binder (polar head) forms hydrogen bonds with PPARγ at the hydrophilic pocket com-

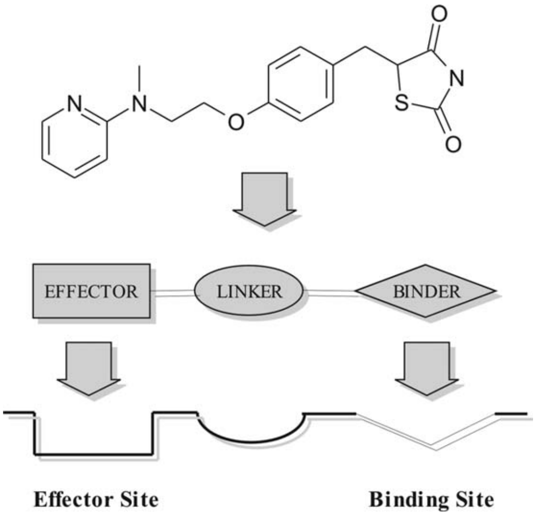


Fig. 2. Pharmacophoric structure from TZD type PPAR agonist.

prised of His449, Tyr473, His323, Gln286, and Ser289. In contrast, the effector (hydrophobic tail) stretches into the hydrophobic pocket of PPARγ with various orientations: they either point up to form hydrophobic interactions with Glu343, Ile341, and Arg288 or direct down to contact with Arg280, Ile281, and Gly284 [19,24]. The tail-up and tail-down configurations have also been observed in the crystal structure of eicosapentaenoic acid when complexed with PPARδ [25]. It indicates that these ligands may adopt multiple tail conformations within the receptor and the hydrophobic pocket has a large space for the structural modification of the modulators of PPAR. The linker also plays an important role in the binding of ligands with PPARγ: it occupies a narrow pocket among Cys285, Tyr327, and Met364, which means that volume of the linkers should not be too big.

2.2. Thinking of A2DVCL

Fig. 3 has summarized the method of constructing A2DVCL and its application in drug discovery.

2.3. Backbones and fragments of A2DVCL

Normally, the first step in the combinatorial process is to define a backbone (also called “core” or “scaffold”) structure to which the side chains will be attached. In this study, we define the linker of the TZDs’ pharmacophoric structure as the backbone. The attachment points (2 in this study) also will be defined by special denotation that will be discussed latter. The binder and effector are considered as the side chains in this study.

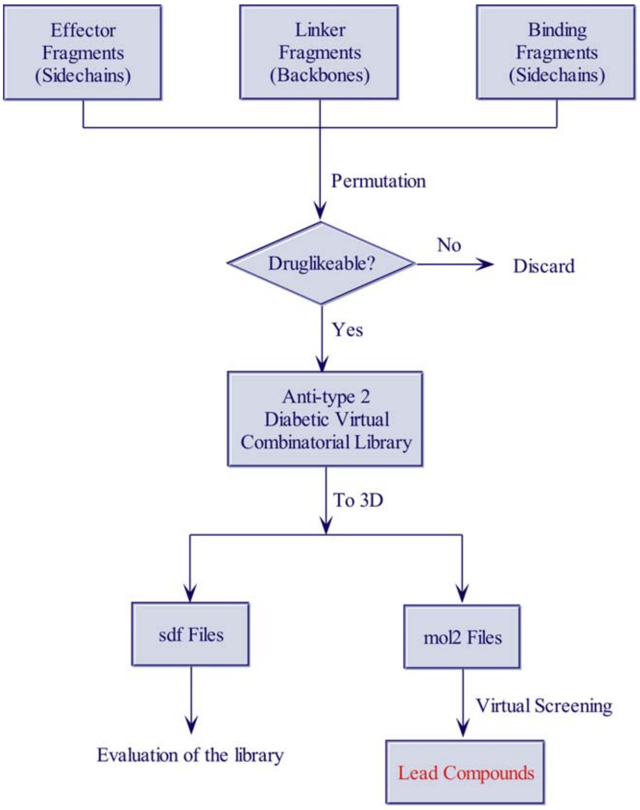


Fig. 3. Construction method of A2DVCL.

The backbones and side chains presented by SMILE strings were collected from literature and patents related to PPAR. Besides these, the authors have designed some fragments according to the physical chemical properties of the PPARs’ binding pockets. We exemplified this progress with the eight compounds as shown in Fig. 1 (although we can not get effector fragment from the compound **2**, we can still get linker and binder fragments from it). Tables 1,2 and 3 list separately the molecular fragments and corresponding SMILES strings representing the effector, linker, and binder decomposed from these eight molcules. It needs to point out that in the SMILES of the linker (backbone), “(&)” and “($)” imply the attachment points.

Every fragment of the backbones has 2 attachment positions. Here we use “($)” and “(&)” to indicate these 2 attachment points: “($)” represents the site of effector and “(&)” represents the site of binder. If the attachments are in rings, “$” and “&” should be marked. But if effector is not in a ring, “($)” no need to write out. In this case, the fragment of the backbone should be written according to the principle that the first character of the string can connect the effector directly. By the same token, if the binding site is not in a ring, “(&)”is also no need to write out. In this case, the fragment of the backbone should be written according to the principle that the last character of the string can connect the binder directly. Also, the fragments of the effectors should be written accord-

Table 1
Molecular fragments and SMILES strings of effector from eight modulators of PPAR

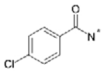
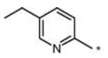
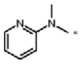
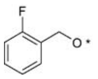
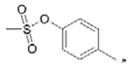
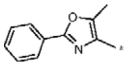
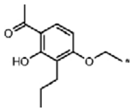
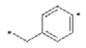
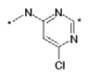
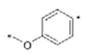
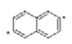
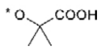
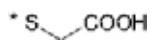
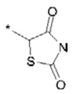
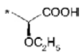
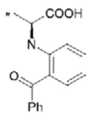
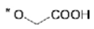
Fragments	Coming from	SMILES strings
	**1**	c1cc(Cl)ccc1C(=O)N
	**3**	c1cc(CC)cnc1C
	**4**	n1ccccc1 N(C)C
	**5**	c1(F)ccccc1CO
	**6**	c1cc(OS(=O)(=O)C)ccc1C
	**7**	c1(C)oc(c4ccccc4)nc1C
	**8**	c1cc(C(=O)C)c(O)c(CCC)c1OCC

Table 2
Molecular fragments and SMILES strings of linker from eight modulators of PPAR

Fragments	Coming from	SMILES strings[a]
	**1**	CCc2ccc(&)cc2
	**2**	CNc2cc(Cl)nc(&)n2
	**3**, **4**, **6**, **7**, **8**	COc2ccc(&)cc2
	**5**	c2($)ccc5cc(&)ccc5c2

[a] "$" and "&" implying the attachment points of effector and binder to the linker.

Table 3
Molecular fragments and SMILES strings of binder from eight modulators of PPAR

Fragments	Coming from	SMILES strings
	**1**	OC(C)(C)C(=O)O
	**2**	SCC(=O)O
	**3**, **4**, **5**	CC3SC(=O)NC3=O
	**6**	C[C@@H](OCC)C(=O)O
	**7**	C[C@@H](C(=O)O)Nc3ccccc3C(=O)c6ccccc6
	**8**	OCC(=O)O

ing to the principle that the last character of the string can connect the backbone directly; the fragments of the binders should be written according to the role that the first character of the string can connect the backbone directly.

Although some programs such as JME [26] can generate SMILES of a compound automatically, it is not recommended to use them to generate the fragments of this library because the results often are not the things we hoped originally. The SMILES of the fragments should be written manually and then we can use some useful tools [27] to verify them before they enter into fragment libraries.

We have constructed three ASCII formatted text files; each of them comprises effector, linker, and binder fragments, respectively. In each fragment library file, every row represents a molecular fragment, so the number of rows in a file is equal to the number of molecular fragments.

When combining the fragments of the three different sectors, the virtual compounds are generated according to the manners described here as shown in Fig. 4: (1) if the backbone does not contain "$" and "&", the program just connects the 3 SMILES strings directly from left to right; (2) if the backbone does not contain "$" but contains "&", the program connects the SMILES strings of effector and linker, and replace the character of "&" by the SMIELS string of binder; (3) if the backbone contain "$", the SMILES string of effector should be converted firstly, and then replace the character of "$".

2.4. ADME filters

Lipinski et al. [28] have concluded that poor absorption or permeation of a compound are more likely to occur when (1) the molecular weight is over 500; (2) the octanol/water partition coefficient (log$P$) is over 5; (3) there are more than five hydrogen bond donors (expressed as the sum of the O–H and N–H groups), and (4) there are more than 10 hydrogen bond acceptors (expressed as the sum of N and O atoms). This is the well-known "rule of 5" (RO5). In addition to the RO5 criteria, other properties also can be used to filter virtual libraries prior to the enumeration stage. With filtering, one can avoid, for example, the presence of large numbers of halogens in the products; the presence of highly flexible unsubstituted, unbranched alkyl chains with, e.g. more than eight CH_2 groups; the formation of compounds that exceeds the maximum accepted number of rings, or ionizable groups, in the product library.

According to the above-mentioned ADME filters, a molecule can not enter into the A2DVCL when (1) it does not accord with the drug-like principle; (2) the number of rings is greater than 7; (3) the number of halogen elements is greater than 7; (4) the number of unsubstituted, unbranched alkyl chains is greater than 8. The so-called drug-like principle here means when a molecule in the virtual library meets any of two conditions from the followings: (1) the molecules weight is less than 650; (2) the calculated log$P$ is less than 6; (3) the number of hydrogen bond acceptors is equal or less than 10; (4) the number of hydrogen bond donors is equal or less than 5.

Since all three of PPAR subtypes have a more spacious pocket (~1400 $Å^3$) than any other nuclear hormone receptors [29], we, therefore, set molecular weight of 625 but not 500 as the cutoff value as the upper limitation to reflect such difference. For instance, the molecular weight of farglitazar [30], a dual agonist of PPARα/γ, is 546.6.

We can accurately enumerate the number of rings, the number of halogens, the number of unsubstituted and unbranched alkyl chains, the number of hydrogen bond acceptors and the number of hydrogen bond donors, and molecular weight, of the molecule expressed by SMILES since these parameters are additive. But the definite value of log$P$ needs to be mea-

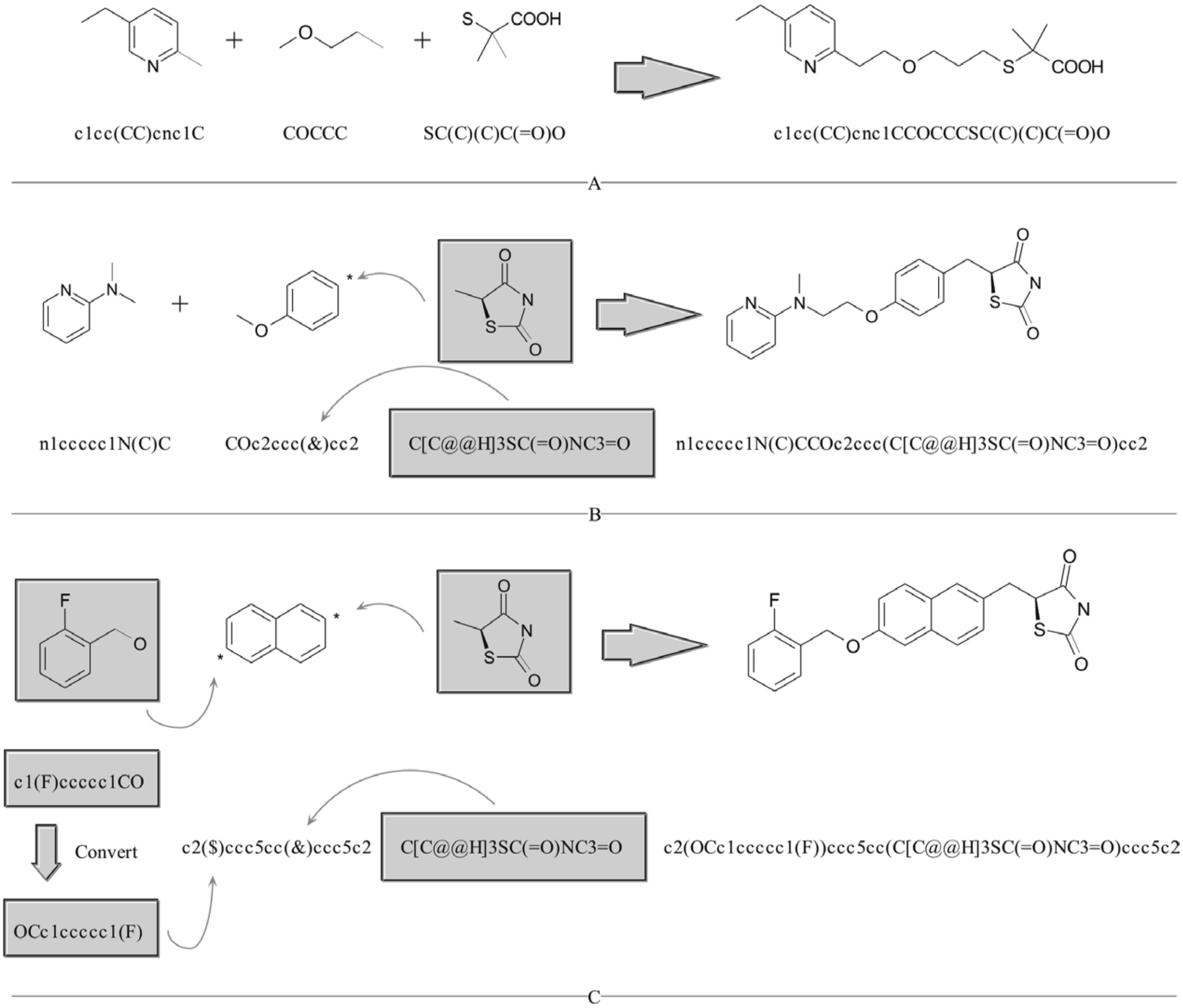


Fig. 4. Generation of virtual molecules.

sured in laboratory. Some algorithms, for example, KLOGP [31] etc. can estimate the log*P* value. The log*P* value of every molecule in A2DVCL is calculated by XLOGP [32] that is based on atom-additive method.

2.5. Virtual screening

The A2DVCL was transformed to sdf and mol2 files by CONCORD standalone 4.0 [33] at the SGI workstation. The sdf file was used to evaluate the diversity of the library, and the mol2 file was used to screen virtually in order to discover potent leads for type 2 diabetes and other metabolic diseases.

In the present study, the program DOCK 4.0 was employed for the primary screening. 1FM9, a X-ray crystal structure of PPARγ downloaded from Brookhaven Protein Data Bank (PDB) [34] was used as the target for virtual screening. Residues around the ligand, farglitazar, at radius of 5 Å was isolated for constructing the grids of docking screening, and the pocket composed by these residues was larger enough to include residues of the binding pocket. The conformational flexibility of the compounds from the database was considered in the docking searching. The DOCK program is designed to find possible orientations of a ligand in a receptor site [35]. The orientation of a ligand is evaluated with a shape scoring function and/or a function approximating the ligand–receptor binding energy.

3. Results and discussions

3.1. The fragments

Except a small portion of fragments are designed by the authors, most of the fragments were collected from the literature and patents published before May 1, 2003. The numbers of fragments obtained for effector, linker, and binder are 283, 40, and 151, respectively. All of them contain (1) 2D connecting information; (2) double bond orientation information if it has; (3) tetrahedral chirality information if it has.

3.2. Generation of the virtual library

The permutation of fragments of the three parts of PPAR modulators will generate numerous molecules. For example, the fragments collected from eight compounds depicted in Fig. 1 will create 168 molcules if none will be discarded by the ADME filters. Among the 168 molcules, 160 are new.

According to the thinking mentioned above, a C++ program containing one main program and nine sub programs was compiled to fulfill the following tasks: judge the need to convert the SMILES string of effector, if so, convert it according to some rules; permutate and combine the fragments to form the molecules; calculate the number of rings; calculate the number of halogen elements; examine the presence of highly flexible unsubstituted and unbranched alkyl chains with eight or more CH_2 group; judge whether or not the compound is an aliphatic or aromatic hydrocarbon, or an amino acid, if so, calculate the correction value of log*P*; judge

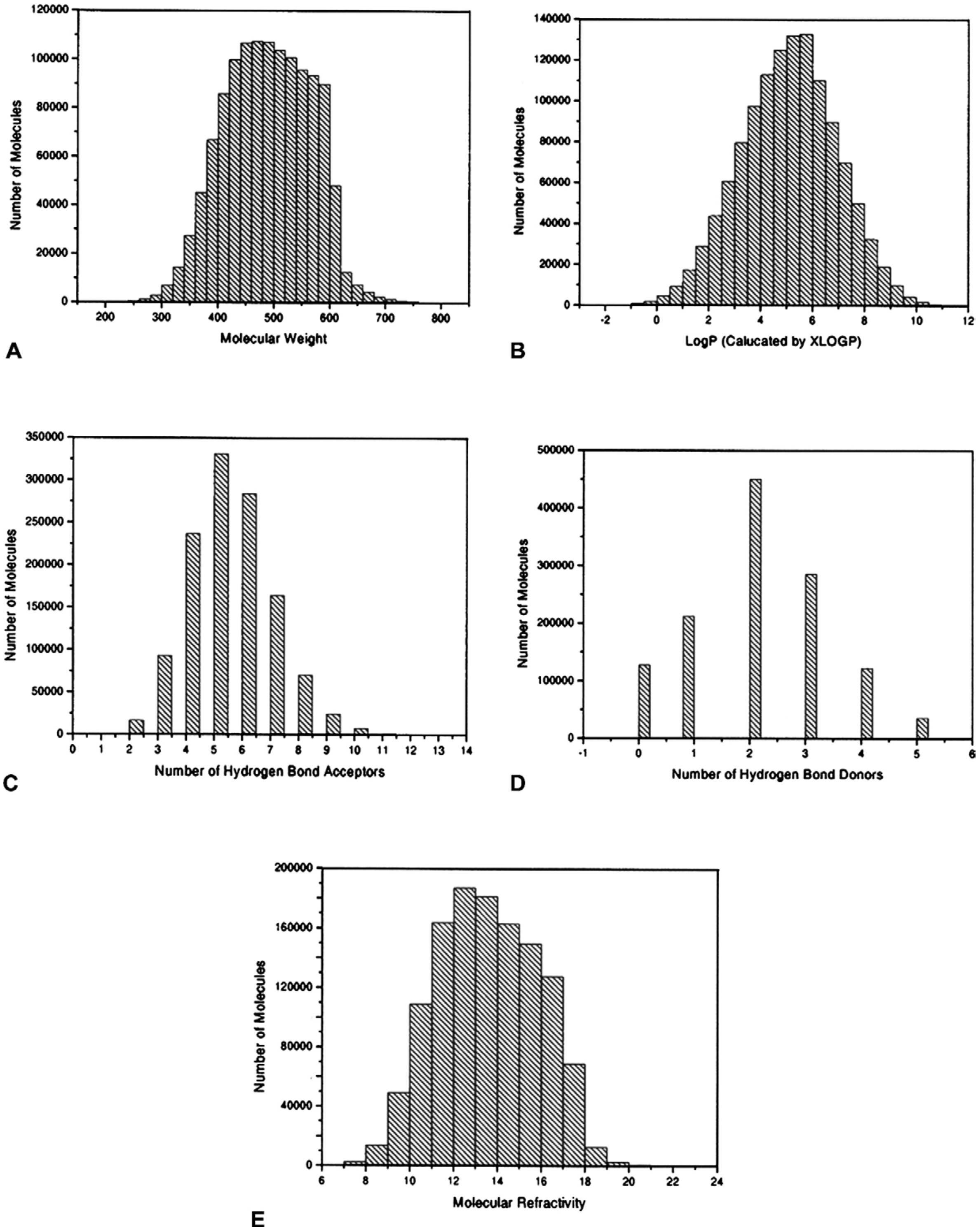


Fig. 5. Distribution of drug-like properties of the A2DVCL. (*A*) the distribution of molecular weight; (*B*) the distribution of log*P* (calculated by XLOGP); (*C*) the distribution of number of hydrogen bond acceptors; (*D*) the distribution of number of hydrogen bond donors; (*E*) the distribution of molecular refractivity.

whether or not the compound can form intramolecular hydrogen bond or has halogen–halogen geminal interaction, if so, calculate the correction value of log*P* value; calculate the molecular weight of the compound, the number of hydrogen bond acceptors, the number of hydrogen bond donors; calculate the number of different descriptors in atom classification by XLOGP, and then calculate the log*P* value; judge the suitability of designed molecule entering into the library; and at last create the virtual combinatorial library containing millions of compounds.

The formed A2DVCL is an ASCII format file. Permutation can generate 1709,320 molcules, but after filtering, the library contains 1226,625 molcules. The rate of molecules being filtered is 28.24%.

3.3. The distributions of drug-like properties

We have analyzed the distribution of molecular weight, log*P* (calculated by XLOGP), and number of the hydrogen bond acceptors and number of hydrogen bond donors of the final A2DVCL. Since molecular refractivity is a very important steric parameter that influences the actions of the ligand and acceptor, we also calculated this parameter for every compound in this library. The distribution diagrams of these parameters are shown in Fig. 5.

All of the five charts are like bell, which implies all of them fit the Gaussian distribution. The minimal and maximal of the molecular weight is 199.25 and 828.07, respectively. Most of the molecular weight is at the span of 400–600. The minimal of log*P* is –2.19, and the maximal reaches a value of 11.92 that corresponds to the compound having only two hydrogen bond acceptors. Most of the log*P* is at the span of 2–8. Although increasing the log*P* value at some extent will increase the affinity to PPAR, it is not a good idea to argument log*P* too much since this will induce bad distribution of the compound on fat and body fluid. The number of hydrogen bond acceptors and donors are ordinal spanning from 2 to 13 and 0–5, respectively. Most of the compounds have four to seven hydrogen bond acceptors and one to three hydrogen bond donors while only one compound has 13 hydrogen bond acceptors in this library. The minimal and maximal of molecular refractivity are 6.25 and 23.11, respectively. Most of this value is concentrated on the scope from 9 to 18.

3.4. Diversity evaluation

The following parameters are often used to describe the diversity profiling of a combinatorial library [36]: two-dimensional descriptors, three-dimensional descriptors, physicochemical descriptors, electronic descriptors, etc. Dr Xu [37] has presented a new term topological scaffold to classify the structures. This arithmetic is called scaffold-based classification approach (SCA). The steps to define the topological scaffold can be described as follows: (1) Define ring bond. A ring bond is a bond with both of its atoms in the same ring. (2) Define linker bond. A linker bond is not a ring bond, but both of its atoms are connect to ring directly or indirectly. (3) Define chain bond. A chain bond is a bond neither a ring bond nor a linker bond. (4) Define topological scaffold. A topological scaffold is a structure that contains ring bond and linker bond but no chain bond. Fig. 6 gives the illustration of the definition of a topological scaffold.

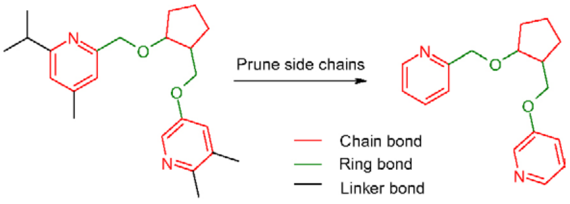


Fig. 6. Definition of topological scaffold.

Unlike most conventional approaches for clustering chemical compounds, SCA is based on topological scaffold but not structural descriptors. The main idea is that if the structures have the same topological scaffold they will be in the same group. In this study, we used SCA to calculate the diversity of our library, and the result are compared with other well-known public database such as NCI3D (1994 released) [38], MDDR-3D [39], AIDS (1999 released) [40], CANCER (1999 released) [41], Traditional Chinese Medicine Database (TCMD) developed by Zhou et al. [42]. The databases, their structure numbers, structure scaffold diversity (SSD) are listed in Table 4 [43].

From Table 4, we can see that the SSD value of A2DVCL is 20.03%, more than the value obtained from NCI3D, but less than the value obtained from CANCER, AIDS, MDDR-3D, and TCMD. It is possible that the limited number of the backbones leads to finite scaffolds of this virtual library.

3.5. Result of virtual screening

Before we put the A2DVCL into virtual screening, we have docked some ligands that are crystallized with PPARγ to check the interactions between the ligands and PPARγ by DOCK 4.0 program. The results are showed in Table 5.

It appears that the main interaction between ligand and PPARγ is hydrophobic force (van der Waals component) regardless of being an agonist or a partial agonist. Although the hydrogen bond (electrostatic component) is important to the system, it only takes up a small portion of the energy score.

The docking result after considering the impacts of heavy atoms etc. on scoring shows that the affinities of lots of com-

Table 4
Topological scaffold numbers and SSD values derived from different database

Database	Number of structures	Scaffold number	SSD (%)
NCI3D	126,089	23,776	18.86
CANCER	32,440	13,458	41.49
AIDS	42,389	17,986	42.43
MDDR-3D	132,726	51,124	38.52
TCMD	9126	3183	34.88
A2DVCL	1226,625	245,693	20.03

Table 5
Docking results of some ligands crystallized with PPARγ

PDB code	Ligand	Type of ligand	Energy score	van der Waals	Electrostatic
1FM6	Rosiglitazone	Selective agonist of γ	–39.94	–38.63	–1.31
1FM9	Farglitazar	Dual agonist of α/γ	–54.17	–53.89	–0.28
1I7I	AZ 242	Dual agonist of α/γ	–37.14	–39.35	2.21
1K74	GW 409544	Dual agonist of α/γ	–57.40	–56.78	–0.63
4PRG	GW 0722	Partial agonist of γ	–61.16	–62.47	1.31

pounds in the library are better than that of farglitazar, a dual agonist of PPARα/γ (nanomol level to PPARγ). The remaining thing is to identify the compounds that have good bioactivity to PPARγ but are not covered by existing patents. Since Schapira, M. et al. previously reported that virtual screening is so far unable to reliably discriminate among PPAR subtypes [44], we did not try to dock the library to PPARα and PPARδ.

3.6. Visualization of the program

To make this program more robust and to apply this program to construct other virtual combinatorial libraries, we have written this program in Visual Basic.net. The interface of this program, we called it virLibLinker, is shown in Fig. 7.

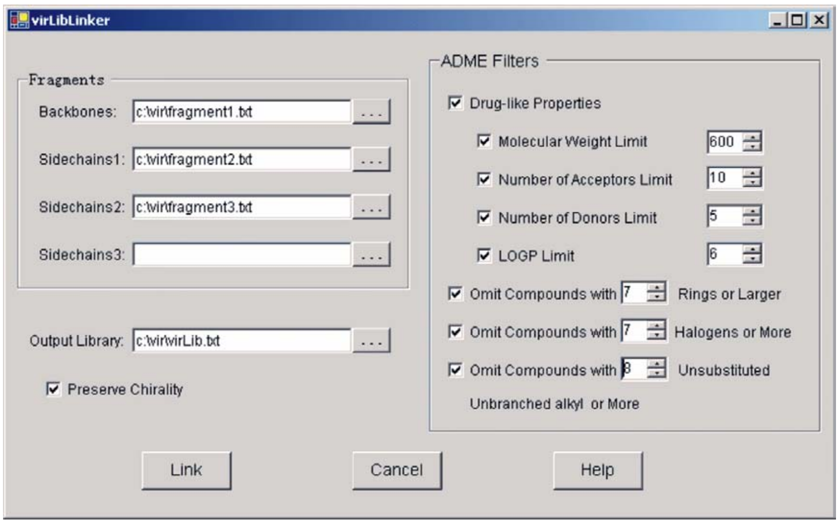
Through this program, we can choose a file containing backbone fragments and at most three files containing side-chain fragments to form a virtual combinatorial library. VirLibLinker is capable of preserving all aspects of whatever stereochemistry specified when defining the backbone(s) and side-chain through checking the “Preserve Chirality” checkbox. And, as we can see at the right part of the interface, we can set ADME filters to repel molecules having poor ADME proprieties in theory when creating the virtual combinatorial library.

Other filters, for example, rotatable bond limit, single bond chain limit etc. are worth to add into the later version of VirLibLinker to obtain more meaningful library. Since the structure generated though SMILES can not be seen by chemical drawing tools, additional modules may need to add in the future.

4. Conclusions

The further development based on other success paradigm is called “me-too” principle in the curatorial research. At present, the data concerning diabetes, PPAR, anti-diabetic drugs has been booming very quickly. Effectively getting and using these information can fasten drug design and drug discovery process. Based on published works, the aim of this article is to discover new knowledge by collecting the information from existing literature and patents in an attempt to identify new chemical entities that have potential in the management of type 2 diabetes and other metabolic diseases.

In this study, we have constructed a virtual combinatorial library according to the structural characters of PPAR modulators using SMILES strings, a chemical nomenclature for specifying molecules or molecular fragments. This method for constructing virtual combinatorial library has also been visualized by Visual Basic.net. The library composed by SMILES has been converted to sdf and mol2 file by CONCORD 4.0. The virtual screening by DOCK 4.0 results in identification of significant number of compounds with stronger binding potential against PPARγ than farglitazar, an nanomole agonist of PPARα/γ.



Backbones: c:\vir\fragment1.txt
Sidechains1: c:\vir\fragment2.txt
Sidechains2: c:\vir\fragment3.txt
Sidechains3:
Output Library: c:\vir\virLib.txt
☑ Preserve Chirality

☑ Drug-like Properties
☑ Molecular Weight Limit 600
☑ Number of Acceptors Limit 10
☑ Number of Donors Limit 5
☑ LOGP Limit 6
☑ Omit Compounds with 7 Rings or Larger
☑ Omit Compounds with 7 Halogens or More
☑ Omit Compounds with 8 Unsubstituted Unbranched alkyl or More

Fig. 7. Interface of the virLibLinker.

Acknowledgements

The authors gratefully acknowledge financial supports from the national “863” project of People’s Republic of China and project of “Most Important Biopharmaceutical Projects” by Guangdong province government in China. The authors also want to give special thanks to Drs Zhibin Li, Zhiqiang Ning and Weiming Hu for their supports of the project. In addition, the authors appreciate the help from Zhenming Liu, a Ph.D. candidate of Peking University for useful discussion and valuable advises for the virtual screening of the A2DVCL.

References

[1] D.K. Agrafiotis, V.S. Lobanov, F.R. Salemme, Nat. Rev. Drug Discov. 1 (2002) 337–346.
[2] M.A. Gallop, R.W. Barrett, W.J. Dower, S.P.A. Fodor, E.M. Gordon, J. Med. Chem. 37 (1994) 1233–1251.
[3] E.M. Gordon, R.W. Barrett, W.J. Dower, S.P.A. Fodor, M.A. Gallop, J. Med. Chem. 37 (1994) 1385–1399.
[4] I. Huc, J.M. Lean, Proc. Natl. Acad. Sci. USA 94 (1997) 2106–2110.
[5] B.A. Leland, B.D. Christie, J.G. Nourse, D.L. Grier, R.E. Carhart, T. Maffett, et al., J. Chem. Inf. Comput. Sci. 37 (1997) 62–70.
[6] V.S. Lobanov, D.K. Agrafiotis, Combin. Chem., High-Throughput Screen. 5 (2002) 167–178.
[7] A.R. Leach, J. Bradshaw, D.V.S. Green, M.M. Hann, J.J. Delany, J. Chem. Inf. Comput. Sci. 39 (1999) 1161–1172.
[8] Tripos, Inc, Legion User’s Manual, 2002.
[9] Tripos, Inc, CombiLibMaker User’s Manual, 1999.
[10] D. Weininger, J. Chem. Inf. Comput. Sci. 28 (1988) 31–36.
[11] W.P. Walters, M.T. Stahl, M.A. Murcko, Drug Discov. Today 3 (1998) 160–178.
[12] H.F. Xu, D.K. Agrafiotis, Curr. Top. Med. Chem. 2 (2002) 1305–1320.
[13] T.L. Blundell, H. Jhoti, C. Abell, Nat. Rev. Drug Discov. 1 (2002) 45–54.
[14] G. Klebe, J. Mol. Med. 78 (2000) 269–281.
[15] M. Rarey, B. Kramer, T. Lengauer, G. Klebe, J. Mol. Biol. 261 (1996) 470–489.
[16] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, et al., J. Comput. Chem. 19 (1998) 1639–1662.
[17] G. Jones, P. Wilett, R.C. Glen, A.R. Leach, R. Taylor, J. Mol. Biol. 267 (1997) 727–748.
[18] I.D. Kuntz, Science 257 (1992) 1078–1082.
[19] T.M. Willson, P.J. Brown, D.D. Sternbach, B.R. Henke, J. Med. Chem. 43 (2000) 527–550.
[20] T.M. Larsen, S. Toubro, A. Astrup, Int. J. Obes. 27 (2003) 147–161.
[21] Y.X. Wang, C.H. Lee, S. Tiep, R.T. Yu, J. Ham, H. Kang, et al., Cell 113 (2003) 159–170.
[22] Y. Kurogi, Drug Des. Discov. 16 (1999) 109–118.
[23] J.L. Oberfield, J.L. Collins, C.P. Holmes, D.M. Goreham, J.P. Cooper, J.E. Cobb, et al., Proc. Natl. Acad. Sci. USA 96 (1999) 6102–6106.
[24] X.Y. Xu, F. Cheng, J.H. Shen, X.M. Luo, L.L. Chen, L.D. Yue, et al., Int. J. Quantum Chem. 93 (2003) 405–410.
[25] H.E. Xu, M.H. Lambert, V.G. Montana, D.J. Parks, S.G. Blanchard, P.J. Brown, et al., Mol. Cell 3 (1999) 397–403.
[26] Web Address, http://www.molinspiration.com/jme/index.html.
[27] Web Address, http://www.daylight.com/daycgi/depict.
[28] C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, Adv. Drug Deliv. Rev. 23 (1997) 3–25.
[29] H.E. Xu, M.H. Lambert, V.G. Montana, G.M. Plunket, L.B. Moore, J.L. Collins, et al., Proc. Natl. Acad. Sci. USA 98 (2001) 13919–13924.
[30] R.T. Gampe Jr., V.G. Montana, M.H. Lambert, A.B. Miller, R.K. Bledsoe, M.V. Milburn, S.A. Kliewer, T.M. Willson, H.E. Xu, Mol. Cell 5 (2000) 545–555.
[31] G. Klopman, J.Y. Li, S. Wang, M. Dimayuga, J. Chem. Inf. Comput. Sci. 34 (1994) 752–781.
[32] R.X. Wang, Y. Fu, L.H. Lai, J. Chem. Inf. Comput. Sci. 37 (1997) 615–621.
[33] Tripos, Inc, CONCORD User’s Manual, 1998.
[34] F.C. Bernstein, T.F. Koetzle, G.J. Williams Jr., E.E. Meyer, M.D. Brice, J.R. Rodgers, O. Kennard, T. Shimanouchi, M. Tasumi, Mol. Biol. 112 (1977) 535–542.
[35] T. Ewing, I.D. Kuntz, J. Comput. Chem. 18 (1997) 1175–1189.
[36] D.K. Agrafiotis, J.C. Myslik, F.R. Salemme, Mol. Divers. 4 (1999) 1–22.
[37] J. Xu, J. Med. Chem. 45 (2002) 5311–5320.
[38] Web address: http://www.cactus.nci.nih.gov/ncidb2/download.html.
[39] Web address: http://www.mdli.com/pdfs/MDDRds.pdf.
[40] Web address: http://www.dtp.nci.nih.gov/docs/aids/aids_data.html.
[41] Web address: http://www.dtp.nci.nih.gov/docs/cancer/cancer_data.html.
[42] X. Yan, Z. Xu, J. Zhou, J. Chem. Inf. Comput. Sci. 39 (1999) 86–89.
[43] B. Liu, A. Lu, L. Zhang, H. Liu, Z. Liu, J.J. Zhou, Int. Electr. J. Mol. Design 3 (2004) 143–149.
[44] M. Schapira, R. Abagyan, M. Totrov, J. Med. Chem. 46 (2003) 3045–3059.