|**Artículo de investigación**

Application of SMILES strings to identification of functional groups responsible for biological activity in medicinal compounds

Aplicación de cadenas SMILES a la identificación de grupos funcionales responsables de la actividad biológica en compuestos medicinales

Aplicação de SMILES strings na identificação de grupos funcionais responsáveis pela atividade biológica em compostos medicinais

Recibido: 11 de marzo de 2018. Aceptado: 15 de abril de 2018

Escrito por:
Sholeh Maslehat
Soroush Sardari*[68]
Mehdi Soheilizad
Payman Nickchi

Abstract

An efficient and practical approach to identification of important functional groups in the structure of medicinal molecules that are main factor to create biological activity by use of SMILES line notation system is described. Simplicity, high proficiency and fast timing are the main of current method. In this study we aim to find an association between some of the identified functional groups, using SMILES code and their corresponding biological properties in the Canada Drug database. In this study, each functional group and its category which has been tested is presented in the corresponding number of occurrences in the category and the total number is shown as well. The p-value for each functional group – category is calculated using proportion test and R statistical software. The tabular results, the last column indicates the impact of our hypothesis for example, sulfonylurea and 5-thio-1H-tetrazole functional groups are associated with their corresponding category and are significant at 0.05 level. Penicillin and 3-aminopropane-1,2-diol are also significant in the majority of their categories. we have developed a method to create a logical and robust relationship between functional groups and biological activity of molecules. According to existing protocol, finding functional groups responsible for the biological activity of medicinal or chemical compounds is possible.

Resumen

Se describe un enfoque eficiente y práctico para la identificación de grupos funcionales importantes en la estructura de moléculas medicinales que son el factor principal para crear actividad biológica mediante el uso del sistema de notación de líneas SMILES. La simplicidad, la alta competencia y el tiempo rápido son los principales del método actual. En este estudio, buscamos encontrar una asociación entre algunos de los grupos funcionales identificados, utilizando el código SMILES y sus correspondientes propiedades biológicas en la base de datos de Canadá Drug. En este estudio, cada grupo funcional y su categoría que se ha probado se presenta en el número correspondiente de ocurrencias en la categoría y también se muestra el número total. El valor de p para cada categoría de grupo funcional se calcula utilizando la prueba de proporción y el software de estadística R. Los resultados tabulares, la última columna indica el impacto de nuestra hipótesis, por ejemplo, los grupos funcionales sulfonilurea y 5-tio-1H-tetrazol están asociados con su categoría correspondiente y son significativos a nivel 0.05. La penicilina y el 3-aminopropano-1,2-diol también son importantes en la mayoría de sus categorías. hemos desarrollado un método para crear una relación lógica y robusta entre los grupos funcionales y la actividad biológica de las moléculas. De acuerdo con el protocolo existente, es posible encontrar

[68] Drug Design and Bioinformatics Unit, Department of Medical Biotechnology, Biotechnology Research Center, Pasteur Institute of Iran, #69 Pasteur Ave., Tehran, 13164, Iran

Key words

Biological Activity, Functional Group, Medicine, SMILES.

Resumo

Uma abordagem eficiente e prática para a identificação de grupos funcionais importantes na estrutura de moléculas medicinais que são o principal fator para criar atividade biológica pelo uso do sistema de notação de linha SMILES é descrita. Simplicidade, alta proficiência e rapidez no timing são os principais métodos atuais. Neste estudo, pretendemos encontrar uma associação entre alguns dos grupos funcionais identificados, usando o código SMILES e suas propriedades biológicas correspondentes no banco de dados do Canadá. Neste estudo, cada grupo funcional e sua categoria que foi testada é apresentada no número correspondente de ocorrências na categoria e o número total também é mostrado. O valor de p para cada grupo funcional - categoria é calculado usando o teste de proporção e o software estatístico R. Os resultados tabulares, a última coluna indica o impacto da nossa hipótese, por exemplo, grupos funcionais sulfoniluréia e 5-tio-1H-tetrazol estão associados à sua categoria correspondente e são significativos ao nível 0,05. A penicilina e o 3-aminopropano-1,2-diol também são significativos na maioria de suas categorias. Desenvolvemos um método para criar uma relação lógica e robusta entre grupos funcionais e atividade biológica de moléculas. De acordo com o protocolo existente, é possível encontrar grupos funcionais responsáveis pela atividade biológica de compostos medicinais ou químicos.

Palavras-chave:
Atividade Biológica, Grupo Funcional, Medicina, SMILES.

Introduction

Simplified Molecular Input Line Entry System (SMILES) is a simple chemical line notation (a typographical method using printable characters) for representing molecules and reactions that originally was introduced by David Weininger in 1987 (1).A SMILES string is a linear text format to represent a two- or three-dimensional molecular structures as a zero-dimensional string, which can describe the connectivity, isomeric and chirality of molecules so that it can be used by the computer. The SMILES is a useful specification and real chemical language with simple vocabulary (atom and bond) and only a few grammar rules. SMILES representations of structure can in turn be used as "words" in the vocabulary of other languages designed for storage of chemical information and chemical intelligence. Also the SMILES representations are generally considered to have the advantage of being more human-readable than other line notation systems. Moreover, it has a wide base of software support with extensive theoretical backing.

Functional groups (FGs) are specific moieties of atoms, or groups of atoms in the structure of molecules that have consistent properties and are responsible for characteristic chemical and biological activity of compounds. It defines the characteristic physical and chemical properties of families of organic compounds. The same functional groups often have the same or similar chemical or biological features and will undergo the same or similar chemical reactions whenever it occurs in different compounds, however the presence of other functional groups and also the size of the molecules can be effective on their properties (2)

Material and Method

In this study we aim to find an association between some of the identified functional groups, using SMILES code and their corresponding biological properties in the Canada Drug database. This relation will help us to elucidate more biological features and properties in the drugs which will enable researchers to focus more on the new emerged properties. At the first stage, the Canada drug bank version 4.5 was downloaded (http://drugbank.ca). This dataset consists of nearly 9,000 unique DRUGCARD identification having the chemical, pharmacological and pharmaceutical features of drugs. The database was in a plain text format entering the information line by line for each DRUGCARD until a pound sing indicates the end of each entry

Palabras clave

Actividad Biológica, Grupo Funcional, Medicina, SMILES.

and the next entry starts in the next lines. In order to selection the desired features and with a better and convenient tabular format, this structure was converted to a tab delimited format (Supplementary file 1). The codes at this stage was developed with Pascal programing language and its corresponding IDE, Delphi 7. At the next stage, this tabular dataset was imported into Microsoft Excel to extract the desired columns, 15 columns were selected considering the goals of this study to process the information. This database and its corresponding column names are provided in Supplementary file 2. This database was filtered considering those drug IDs which their corresponding SMILES code was not null. Having done the final filtering and mining the database, we reached the final database with 4 columns in a Microsoft Access format which is also released as the Supplementary file 3. This database holds the DrugID, Category, Name, and SMILES code of each drug and is ready for future processing.

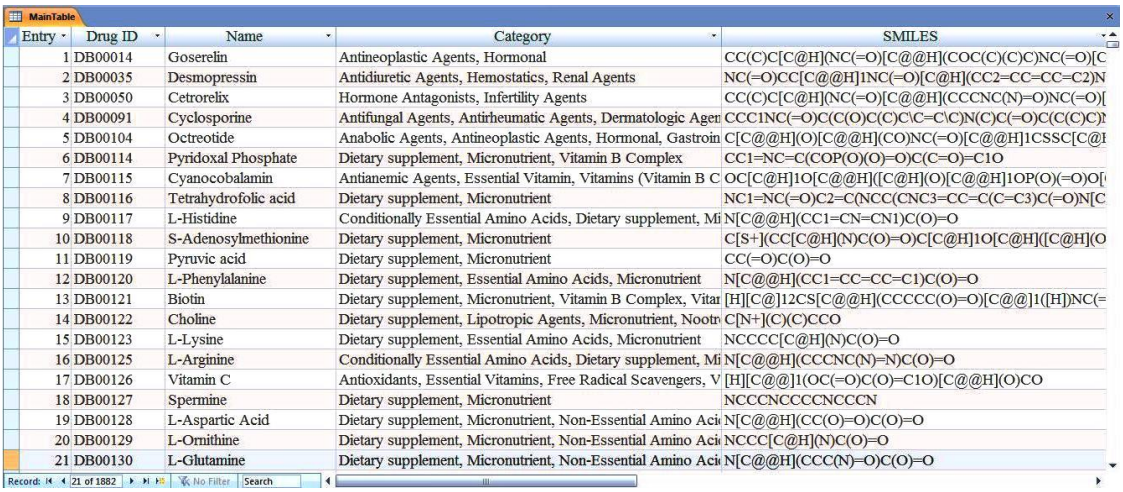
In the next step, we downloaded the image of each structure from the Drug bank using the automatic query sending to Drug bank website with R statistical software. The images were classified for each drug ID and saved in beside the database for each drug ID separately. After mining the downloaded database from databank and completing the database, a search engine was developed under the Microsoft windows operating system using Pascal programing language and its corresponding IDE, Delphi 7. We used this engine to search for SMILES code and detecting their category and functional group.

In this study, we are interested in the association of functional groups and their biological activity considering their SMILES code. To reach this goal, we searched in the prepared database for some of the known and predefined functional groups considering their SMILES code. Having searched in this database, we were capable of classifying and identifying the number of drugs in each category for each functional group. At the next stage, we performed a proportion test to test our hypothesis regarding the association of functional groups and their biological activity in drugs. Proportion test is a useful test to test if the occurrence of an bi-result event is according to chance and has been considered diligently in(3, 4). Here, we are interested to, based on our findings, if the effect of each functional group and its corresponding biological activity is by chance or is related to its properties based on our findings. Proportion test is a good candidate to measure the accuracy of this hypothesis and we consider p as the probability of occurring a related biological activity to a functional group. The null hypothesis, in this study, postulates that p is equal to 0.05 (this event happens totally by chance) and the alternative hypothesis assumes that the p is greater than 0.05 (not by chance). This conducted proportion test was performed using R statistical software (5).

Results and discussion

As a part of our studies to search practical and convenient approaches for simplicity scientific problems, especially in the fields of bioinformatics(6), in this paper, we wish to introduce an efficient and easy method for finding biological active functional groups in the backbone of medicinal compounds by use of SMILES string specifications and Then, we are going to introduce a rational and efficient relationship between detected functional groups and biological properties in investigated molecules.

At first, to achieve our desired goals, we have obtained a very complete medicinal database contains a comprehensive information of nearly 9,000 medicinal molecules from Canada drug bank(7, 8). Our initial studies indicate that only drug category information of 1,882 records containing SMILES string is available in this database, so to accomplish our desired purposes, we create a new drug database based on that database. Then, among abundant existed information in this new database, we have extracted mere four cases includes drug ID, name, category and SMILES string of drug (Fig. 1).



MainTable

Entry	Drug ID	Name	Category	SMILES
1	DB00014	Goserelin	Antineoplastic Agents, Hormonal	CC(C)C[C@H](NC(=O)[C@@H](COC(C)(C)C)NC(=O)[C
2	DB00035	Desmopressin	Antidiuretic Agents, Hemostatics, Renal Agents	NC(=O)CC[C@@H]1NC(=O)[C@H](CC2=CC=CC=C2)N
3	DB00050	Cetrorelix	Hormone Antagonists, Infertility Agents	CC(C)C[C@H](NC(=O)[C@@H](CCCNC(N)=O)NC(=O)[
4	DB00091	Cyclosporine	Antifungal Agents, Antirheumatic Agents, Dermatologic Agen	CCC1NC(=O)C(C(O)C(C)C\C=C\C)N(C)C(=O)C(C(C)C)I
5	DB00104	Octreotide	Anabolic Agents, Antineoplastic Agents, Hormonal, Gastroin	C[C@@H](O)[C@@H](CO)NC(=O)[C@@H]1CSSC[C@I
6	DB00114	Pyridoxal Phosphate	Dietary supplement, Micronutrient, Vitamin B Complex	CC1=NC=C(COP(O)(O)=O)C(C=O)=C1O
7	DB00115	Cyanocobalamin	Antianemic Agents, Essential Vitamin, Vitamins (Vitamin B C	OC[C@H]1O[C@@H]([C@H](O)[C@@H]1OP(O)(=O)O[
8	DB00116	Tetrahydrofolic acid	Dietary supplement, Micronutrient	NC1=NC(=O)C2=C(NCC(CNC3=CC=C(C=C3)C(=O)N[C
9	DB00117	L-Histidine	Conditionally Essential Amino Acids, Dietary supplement, Mi	N[C@@H](CC1=CN=CN1)C(O)=O
10	DB00118	S-Adenosylmethionine	Dietary supplement, Micronutrient	C[S+](CC[C@H](N)C(O)=O)C[C@H]1O[C@H]([C@H](O
11	DB00119	Pyruvic acid	Dietary supplement, Micronutrient	CC(=O)C(O)=O
12	DB00120	L-Phenylalanine	Dietary supplement, Essential Amino Acids, Micronutrient	N[C@@H](CC1=CC=CC=C1)C(O)=O
13	DB00121	Biotin	Dietary supplement, Micronutrient, Vitamin B Complex, Vitar	[H][C@]12CS[C@@H](CCCCC(O)=O)[C@@]1([H])NC(=
14	DB00122	Choline	Dietary supplement, Lipotropic Agents, Micronutrient, Nootr	C[N+](C)(C)CCO
15	DB00123	L-Lysine	Dietary supplement, Essential Amino Acids, Micronutrient	NCCCC[C@H](N)C(O)=O
16	DB00125	L-Arginine	Conditionally Essential Amino Acids, Dietary supplement, Mi	N[C@@H](CCCNC(N)=N)C(O)=O
17	DB00126	Vitamin C	Antioxidants, Essential Vitamins, Free Radical Scavengers, V	[H][C@@]1(OC(=O)C(O)=C1O)[C@@H](O)CO
18	DB00127	Spermine	Dietary supplement, Micronutrient	NCCCNCCCCNCCCN
19	DB00128	L-Aspartic Acid	Dietary supplement, Micronutrient, Non-Essential Amino Aci	N[C@@H](CC(O)=O)C(O)=O
20	DB00129	L-Ornithine	Dietary supplement, Micronutrient, Non-Essential Amino Aci	NCCC[C@H](N)C(O)=O
21	DB00130	L-Glutamine	Dietary supplement, Micronutrient, Non-Essential Amino Aci	N[C@@H](CCC(N)=O)C(O)=O

Record: 21 of 1882 | No Filter | Search

Fig 1. The used database in this work

After generating our database, next we needed a convenient and proficient searching engine that can find functional groups based on SMILES strings. Since in nomenclature base on SMILES system some of the functional groups can be searched by more than one keyboard and usually these keyboards must be separated from each other, we needed a specialized tool that can perform search based on our intentions. Our developed search engine allows users to search multiple functional groups at the same time. A general overview of the program is shown in Fig 2.

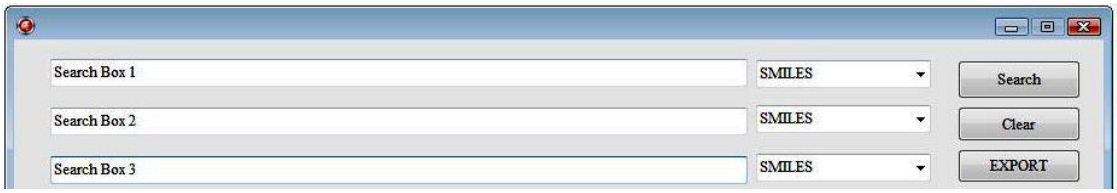
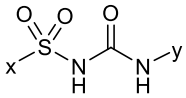
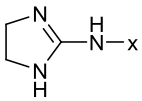
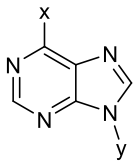
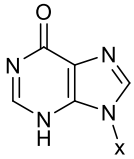
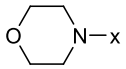
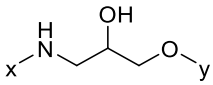
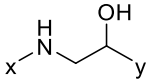
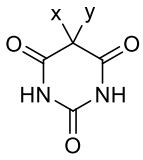
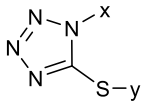
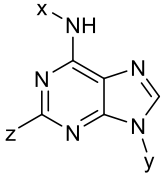
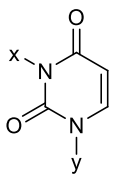


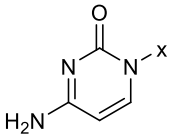
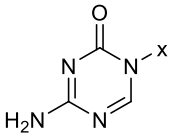
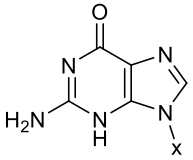
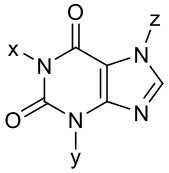
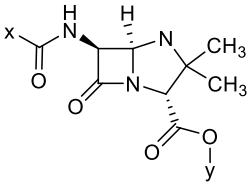
Fig 2. General overview of designed searching tool for finding FGs based on SMILES strings.

We have designed only three search boxes in the program because our surveys showed that for finding most of functional groups in the structure of molecules there would not be a need for more than three search boxes. For example for finding a small-sized functional group such as carboxylic acid (-COOH) or cyano (-CN) in the structure of a large molecule, only one search box is required. The relevant SMILES codes for -COOH and –CN, will be C(=O)O or OC(=O) and C#N, respectively (Table 1, entries 1 and 2). The number of search boxes to finding medium or large-sized functional groups will be two or more. For example to barbiturate functional group, two search boxes and for Xanthine functional group, three search boxes are required. For barbiturate, the corresponding SMILES keywords are C1 and C(=O)NC(=O)NC1=O (Table 1, entry 9). For Xanthine, the relevant SMILES keywords are N1C2=C, C(=O)N and C1=O (Table 1, entry 16).

It should be noted sometimes for finding functional groups is not required to importing all relevant parts of one specific SMILES keyword, especially in the case of large-sized functional groups. In other words, only by entering some SMILES parts of a functional group, it can be found those functional groups. For example, penicillin has a large-size functional group. To find this functional group among the existing molecules of database, it is not necessary to import all SMILES letters relevant to penicillin (Table 1, entry 17). Some of the results are summarized in Table 1.

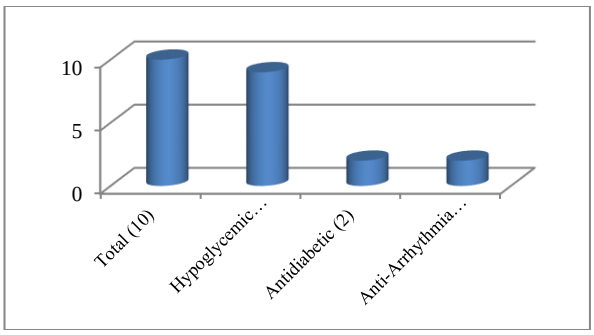
Table 1. Examples of the functional groups and their relevant SMILES keyword(s)

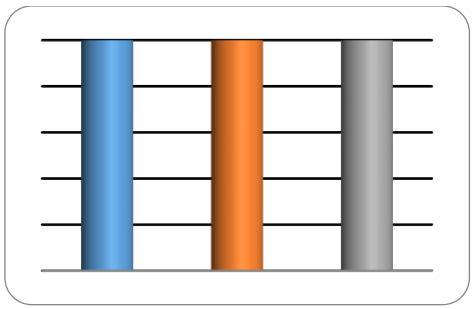
Entry	Functional Group	Structure[a]	Relevant SMILES keyword(s)	Required search boxes
1	Carboxylic acid		C(=O)O or OC(=O)	1
2	Cyano		C#N	1
3	Triflouromethyl		C(F)(F)F	1
4	Sulfonylurea		S(=O)(=O)NC(=O)N or NC(=O)NS(=O)(=O)	1
5	4,5-dihydro-1H-imidazol-2-amine		NC1=NCCN1 or NC2=NCCN2	1
6	Purine		N1C=NC2=C(N)N=CN=C12 or N1C=NC2=C1N=CN=C2N	1
7	3H-purin-6(9H)-one		N1C=NC2=C1NC=NC2=O	1
8	Morpholine		N1CCO and CC1 or N2CCO and CC2 or N3CCO and CC3	1
9	3-aminopropane-1,2-diol		NCC(O)CO or OCC(O)CN	1
10	2-aminoethanol		NCC(O)C or CC(O)CN	1
12	Barbiturate		C1 and C(=O)NC(=O)NC1=O	2
13	5-thio-1*H*-tetrazole		SC and NN=NN	2
14	Adenine		N1C=NC2=C and N=CN=C12 or NC1=C2N=CN and C2=NC=N1 or NC1=NC=N and C2=C1N=CN2	2
15	Uracil		N1C=C and C(=O)NC1=O or C1=CN and C(=O)NC1=O	2

16	Cytosine		NC1=NC(=O)N and C=C1	2
17	4-amino-1,2-dihydro-1,3,5-triazin-2-one		NC1=NC(=O)N and C=N1	2
18	2-amino-3H-purin-6(9H)-one		NC1=NC(=O)C2=C(N1)N and C=N2 or N1C=NC2=C1NC(N)=NC2=O	2
19	Xanthine		N1C2=C and C(=O)N and C1=O or N1C=N and C2=C1C(=O)N and C(=O)N2C	3
20	Penicillin		SC(C)(C) and [C@@H] and C(O)=O or C(C)(C)S and [C@@H] and C(O)=O	3

[a] x, y and z are other parts of molecule that can be same or different.

After identification of functional groups in the structure of molecules using this simple tool, the next step was to create a convenient relationship between detected functional groups and biological properties of molecules containing functional groups. To achieve this goal, we randomly chose a number of functional groups, by accident and examine the biological category of molecules containing those functional groups. Some of the results are shown in Fig 3. For example, our investigations based on SMILES search showed that there are 10 medicines with sulfonylurea functional group in our database. Among them 9 medicines are hypoglycemic agents, so it can be said with certainty sulfonylurea functional group acts as a hypoglycemic agent (Fig 3. Graph A). Another example is 5-thio-1*H*-tetrazole functional group that is based on SMILES search was seen in the structure of 10 molecules in our library. By examining the biological category of these molecules, we observed that all of them are anti-bacterial agents and cephalosporins (Fig 3. Graph B). Also the moiety of barbiturate functional group was seen in the structure of 10 molecules that 8 cases are hypnotics and sedatives (Fig 3. Graph C). In the case of Penicillin functional group, all 22 existed molecules in our database are anti-bacterial agents (Fig 3. Graph D). More examples are shown in the Table 2.





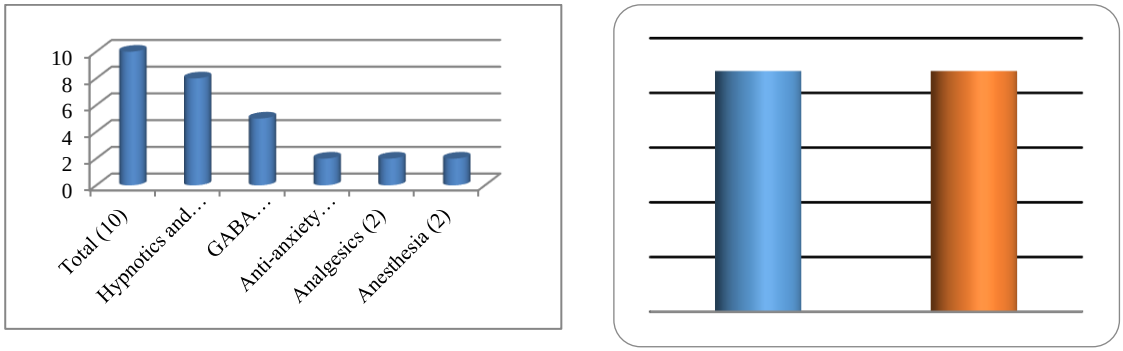
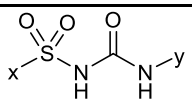
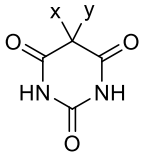
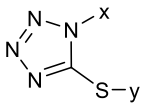
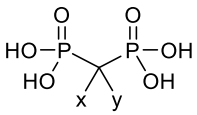
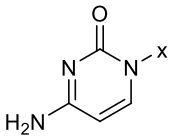
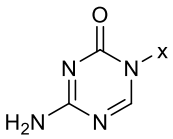
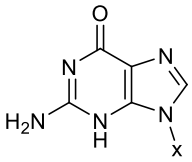
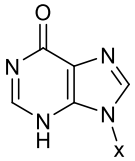
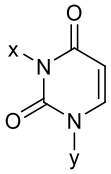
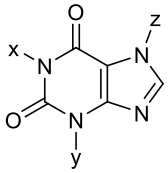
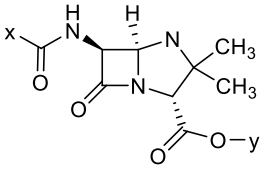
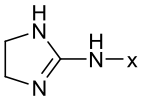
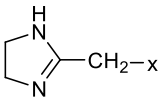
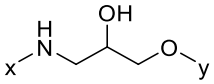
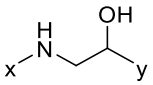


Fig 3. Statistical view of a number of functional groups and their major biological activities, (A) Sulfonylurea, (B) 5-thio-1*H*-tetrazole, (C) Barbiturate and (D) Penicillin.

Table 2. Examples of the functional groups and their relevant major biological categories

Entry	Functional Group (Total Medicine)	Structure	Major biological category (No)
1	Sulfonylurea (10)		Hypoglycemic Agents (9)
2	Barbiturate (10)		Hypnotic and Sedative (8), GABA Modulator (5)
3	5-thio-1*H*-tetrazole (10)		Anti-Bacterial Agent (10), Cephalosporin (10)
4	Methylenediphosphonic acid (8)		Bone Density Conservation Agents (8), Antihypocalcemic Agents (7), Antiresorptives (4)
5	Cytosine (7)		Anti-HIV Agents (4), Antimetabolites (4), Antineoplastic (4), Antiviral Agents (4)
6	4-amino-1,2-dihydro-1,3,5-triazin-2-one (2)		Antimetabolites (2), Antineoplastic (2)
7	2-amino-3H-purin-6(9H)-one (6)		Antiviral Agents (6)
8	3H-purin-6(9H)-one (2)		Purine Nucleoside Phosphorylase inhibitor (2)

9	Uracil (10)		Antimetabolites (6), Antiviral Agents (4)
10	Xanthine (10)		Bronchodilator Agents (6), Phosphodiesterase Inhibitors (4), Vasodilator Agents (4)
11	Penicillin (22)		Anti-Bacterial Agent (22)
12	4,5-dihydro-1H-imidazol-2-amine (4)		Adrenergic alpha-Agonists (4)
13	4,5-dihydro-1H-imidazole (7)		Adrenergic alpha-Antagonists (4), Antihypertensive Agents (3), Sympatholytics (3)
14	3-aminopropane-1,2-diol (19)		Adrenergic beta-Antagonists (18), Antihypertensive Agents (16), Anti-Arrhythmia Agents (13)
15	2-aminoethanol (12)		Adrenergic beta-Agonists (9), Sympathomimetics (9), Bronchodilator Agents (8)

[a] x, y and z are other parts of molecule that can be same or different.

Two interesting examples of functional groups are 3-aminopropane-1,2-diol and 2-aminoethanol (Table 2, entries 14, 15). We observed that these functional groups are existed in the structure of a number of drugs in our database. Atenolol, Carteolol and Practolol are a number of commercial drugs containing 3-aminopropane-1,2-diol functional group (Fig. 4). Also, Sotalol, Isoproterenol, Salbutamol and Clenbuterol are Examples of commercial drugs includes 2-aminoethanol functional group (Fig. 5).

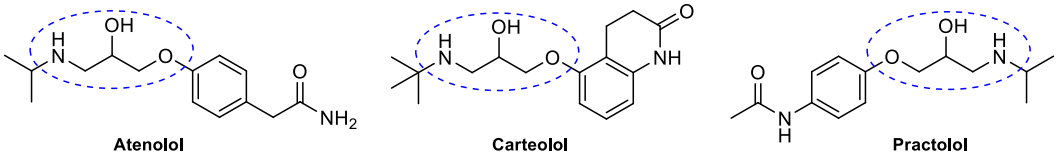


Fig. 4. Examples of commercial drugs containing 3-aminopropane-1,2-diol functional group.

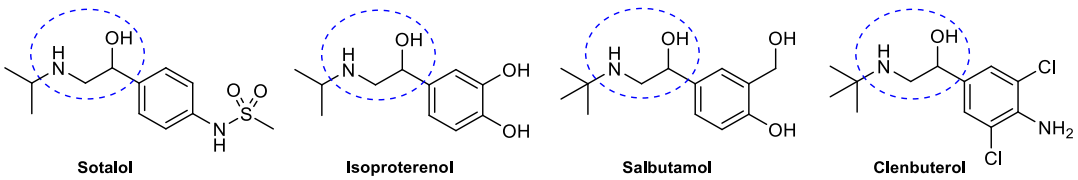


Fig. 5. Examples of commercial drugs containing 2-aminoethanol functional group.

As shown in Fig. 6, among the 19 drugs available in our database containing 2-aminoethanol moiety, 18 drugs shows adrenergic beta-antagonist activity. While, among the 12 drugs with 2-aminoethanol functional group, 9 of them are adrenergic beta-agonist.

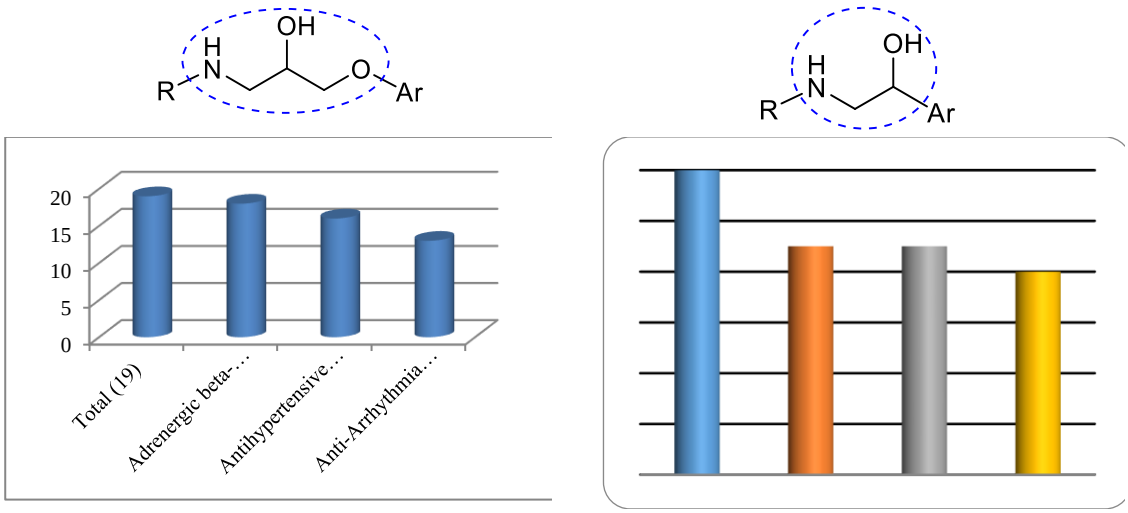


Fig. 6. Major biological activities of drugs containing (A) 3-aminopropane-1,2-diol, (B) 2-aminoethanol functional groups.

Proportion test measures the accuracy of our results

As it was mentioned in the material and method part of this article, we seek to test the accuracy of our hypothesis using the well-known proportion test. The results of this test are shown in table 3. In this table, each functional group and its category which has been tested is presented in rows and the corresponding number of occurrences in the category and the total number is shown as well. The p-value for each functional group – category is calculated using proportion test and R statistical software(5). The last column indicates the result of our hypothesis. Sulfonylurea and 5-thio-1H-tetrazole functional groups are associated with their corresponding category and are significant at 0.05 level. Penicillin and 3-aminopropane-1,2-diol are also significant in the majority of their categories. Although some functional groups are well associated with their corresponding category, some of them are not associated and more validation approaches are required to test the accuracy of hypothesis for these categories.

Table 3. Results of proportion test in each functional group. The significant level is set to 0.05 in all cases.

Funtional Group	Category	Total	Found in Category	p-value	Significant(*)
Sulfonylurea	hypoglycemic agents	10	9	0.013428348	Yes
5-thio-1H-tetrazole	anti-bacterial agents	10	10	0.002213263	Yes
	Cephalosporins	10	10	0.002213263	Yes
Barbiturate	hypnotics and sedatives	10	8	0.056923149	No
Penicillin	anti-bacterial agents	22	22	3.78E-06	Yes
3-aminopropane-1,2-diol	Adrenergic beta-antagonists	19	18	1.21E-04	Yes
	Antihypertensive agents	19	16	0.002952696	Yes
	Antiarrhythmia agents	19	13	0.084334309	No
2-aminoethanol	Adrenergic beta-antagonists	12	9	0.074457337	No
	Sympathomimetics	12	9	0.074457337	No
	Bronchodiletor agents	12	8	0.193238115	No
* Significancy level is set to 0.05					

Conclusion

In summary, we have introduced a practical and efficient approach to finding special functional groups in the backbone of medicinal compounds using SMILES string specifications. Then, we have developed our introduced method to create a logical and robust relationship between functional groups and biological activity of molecules. According to existing protocol, finding functional groups responsible for the biological activity of medicinal or chemical compounds easily is possible by use of SMILES specifications. This approach easily can be extended or can be applied for other similar applications.

References

Favre, H, Powell, W. (2013). Nomenclature of organic chemistry: IUPAC recommendations and preferred names 2013: Royal Society of Chemistry.

Law, V., Knox, C., Djoumbou, Y., Jewison, T., Guo, A., Liu, Y., Maciejewski, A., Arndt, D., Wilson, M., Neveu, V., Tang, A., Gabriel, G. y Adamjee, C, Dame, Z., Han, B., Zhou, y Wishart, D. (2013. DrugBank 4.0: shedding new light on drug metabolism. *Nucleic acids research* 2013;42(D1):D1091-D7.

Lawal, B. (2014). Applied statistical methods in agriculture, health and life Sciences: Springer.

Lehmann, E., Romano, J. (2005). Testing Statistical Hypotheses (Springer Texts in Statistics).
Lemke, T. (2003). Review of organic functional groups: introduction to medicinal organic chemistry: Lippincott Williams & Wilkins.

Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013. ISBN 3-900051-07-0; 2014.

Weininger D. SMILES (1988). A chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of chemical information and computer sciences*; 28(1): 31-6.

Wishar, D., Knox, C., Guo, A., Shrivastava, S., Hassanali, M., Stothard, P., Chang, Z., Woolsey, J. (2006). DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Research*; 34(suppl_1): D668-D72.