BIOSCIENCES BIOTECHNOLOGY RESEARCH ASIA, September 2018. Vol. **15**(3), p. 541-548

Frequency and Importance of Six Functional Groups that Play a Role in Drug Discovery

Sholeh Maslehat, Soroush Sardari* and Mahboube Ganji Arjenaki

[1]Drug Design and Bioinformatics Unit, Department of Medical Biotechnology, Biotechnology Research Center, Pasteur Institute of Iran, Tehran, Iran

http://dx.doi.org/10.13005/bbra/2659

(Received: 05 July 2018; accepted: 18 September 2018)

Small molecules are composed of chemical functional groups; they are sets of connected atoms or atom groups that determine properties and reactivity of the parent molecule. DrugBank is a rich source of information that containing molecular data about small molecules, their mechanisms, pharmaceutical interaction and targets. In this study, After collecting data of small drug molecules from DrugBank database and classifying them in different categories based on their mechanism of action, the therapeutic properties of the molecules were recorded. Finally, the functional group from the pharmaceutical structures were elucidated and registered for each group. The functional groups were divided into five distinct groups in drug design, and a correlation between identified functional group to pharmaceutical structure were indicated according to the classified functional groups of small molecule and drug categories; then defined their frequency in categories, at high abundant functional group present in categories reported. The most frequent rings were benzene and cyclohexane; the common acid functionality had been acetate (carboxy-); three most repeated saturated heterocyles are piperidine, piperazine and azetidine; among the unsaturated heterocyles, pyridine, imidazole and indole are noticed; This database, that may be guidance for researchers with the aim at designing new drugs.

Keywords: Small molecules, functional groups.

The navigate of functional groups (FGs) are sets of connected atoms or atom groups that determine properties and reactivity of parent molecule, forms basis of organic chemistry, analytical chemistry, medicinal chemistry and also chemical nomenclature. Therefore, It defines the characteristic physical, chemical properties of families of organic compounds and indicates the classification of their[1-3]. Functional groups have relatively constant characteristics even when connected to different structures[4]. Small molecules can be composed of independent chemical functional groups, therefore a classification for small molecules could also be obtained by identifying the molecules chemical functional groups using objective and computable criteria[5]. The common characteristic of the chemical structures of the global small molecular drugs could be used as a criterion to guide the selection, design and optimization of drug purpose compounds and candidates in the early stages of preclinical research, therefore it would be increase the success rate of drug development and eliminate those poor drug-like compounds in advance and avoid more research and development expenses. Therefore, the study of the key chemical structure characteristics

*Corresponding author E-mail: sorosh.sardari98@gmail.com

of small molecule drugs has significant theoretical and practical value[6]. Researchers who are interested in chemical compounds for drug design, small molecule drugs especially face several problems. They might want to find the common features of a category of drugs for design a new drug with similar targeting, or how a group of small molecule is converted into other types of compounds, even when the total structures are not specified and how a small molecule is fined or designed, when biological molecular target is defined.

DrugBank (>6000 approved and experimental drugs) (http://www.drugbank.ca) is a web-enabled database and a comprehensive online database containing molecular information about drugs, their mechanisms, interactions and targets. Number of investigational drugs, small molecule drugs, drug-drug interactions, SNP-associated drug effects has grown in the database in more than 10 years. DrugBank 3.0 has been contained data on drug metabolism, absorption, distribution, metabolism, excretion and toxicity (ADMET) and other kinds of quantitative structure activity relationships (QSAR) information also numbers of small molecule drugs in GrugBank 3.0 more than previous versions (4, 7-10). Therefore, its database is a good data resource for small molecule drugs, that may be guidance for researchers with aim design new drugs. Based on our knowledge, no evidence or software defined that, how are frequency drug functional groups between small molecules and relative drug categories available at DrugBank 3.0. In this study, we classified functional groups of small molecule drugs base on drug categories, then defined their frequency in categories, at result high abundant functional group into categories reported.

MATERIAL AND METHOD

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 was introduced by David Weininger in 1987[11] originally. SMILES string is a linear text format to represent the two- or three-dimensional of molecular structures as a zero-dimensional string, which can described the connectivity, isomeric and chirality of molecules so that it can be used by the computer. The SMILES is a useful specification, real chemical language with simple vocabulary (atoms and bonds) and only a few grammar rules. representations of structure of SMILES can be used in turn as "words" in the vocabulary of other languages that designed for storage of chemical information and chemical intelligence. Also the representations of SMILES are generally considered for 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 the size of the molecules can be effective on their properties[11].

RESULTS

In this article, we studied 114 functional groups that were more important in drug design. In the next step, the functional group extracts from the pharmaceutical structures.

The Functional groups use in this study

We divided Functional groups based on the structural similarities of their molecules to five familiar groups: Saturated cyclic, Unsaturated cyclic, Heterocyclic compound unsaturated, Acids, Amide, Aldehydes (the carbonyl group is polar), Neutral aromatic (Polar Deactivated). Then we sketched their graphs

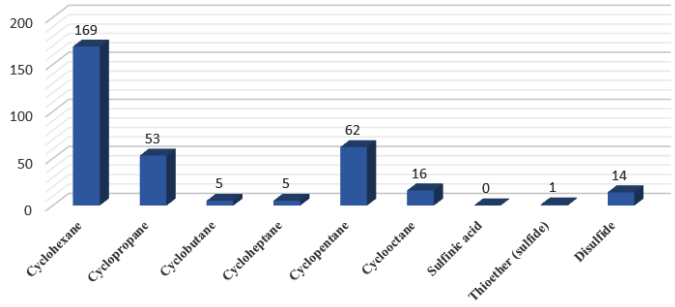
Group 1: Functional groups that have saturated cyclic properties include fragmentation: Cyclohexane ,cyclopropane, cyclobutane, cycloheptane, cyclopentane, cyclooctane, that are distinguished by the yellow color in (Fig 1).

Group 2: Functional Groups that have unsaturated cyclic properties include fragmentation: benzene, naphthalene that are distinguished orange in (Fig 1).

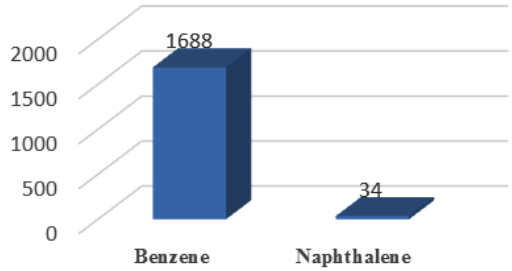
Group 3: Functional groups that are Heterocyclic compound unsaturated , which include fragmentation aziridine, dioxirane, oxirane, thiirane, azetidine, oxetane, thietane. Pyrrolidine, oxolane, thiolane, piperidine, piperazin, 1,4-dioxane, 1,4-dithiane, morpholine, thiomorpholine, Tropane, azapentacyclo, oxane, 1,3-diazinane, azabicyclo, Imidazolidine, Imidazolidine, phenothiazin that are distinguished by red in (Fig 1).

Group 4: : Functional groups that are Heterocyclic compound saturated , which include fragmentation PYRIDINE , Thiophene, THIAZOLE , isoxazole, oxazole, pyrazole , Indol, imidazole, furan, pyrrole, acridine, purin, pyrimidine, pentadeca, benzodiazepin, imidazolidino , oxazolidin, Tetrazole, triazol , triazol, Pteridin-4-ol, pyrazine, Benzopyran=(chromene), benzodiazol, Benzothiazole, 1,4-Dihydroquinoline, Which are distinguished by purple in (Fig 1).

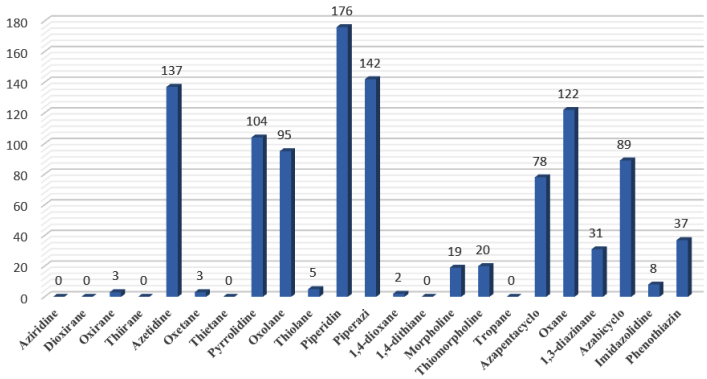
Group 5: Functional groups that are Acids, which include fragmentation carboxylate Anion, sulfinic acid, sulfonic acid, Vendors, phosphodiester, phosphonic acid, phosphate,



Group 1. Total number of saturated cyclic moieties in various therapeutic categories mentioned in DrugBank



Group 2. Saturated cyclic rings

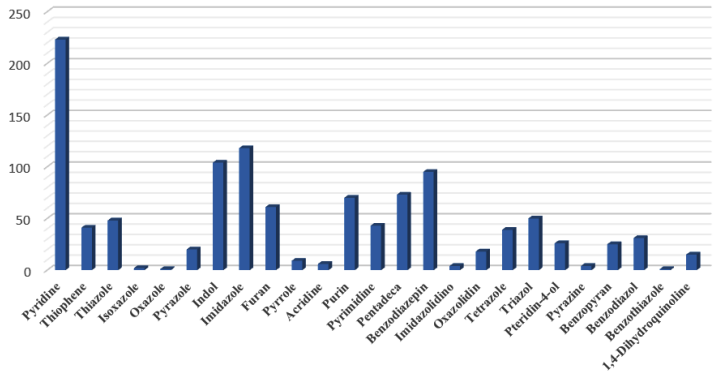


Group 3. Heterocyclic saturated moieties

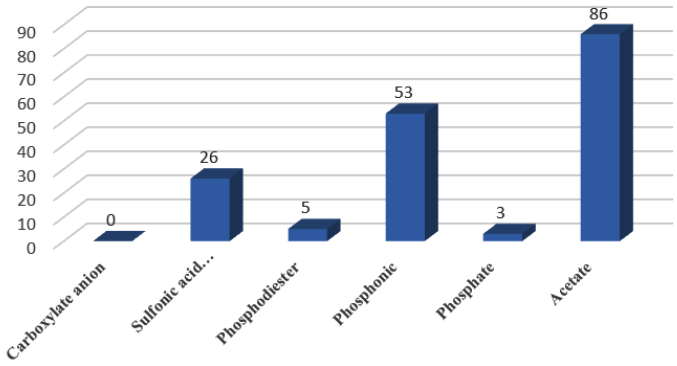
acetate, Which are distinguished in dark red in (Fig 1).

Group 6: : Functional groups that are Amides, which include fragmentation Amine NH3, THIOUREA, Amine NH2 , Thioamide Which are distinguished in green with (Fig 1).

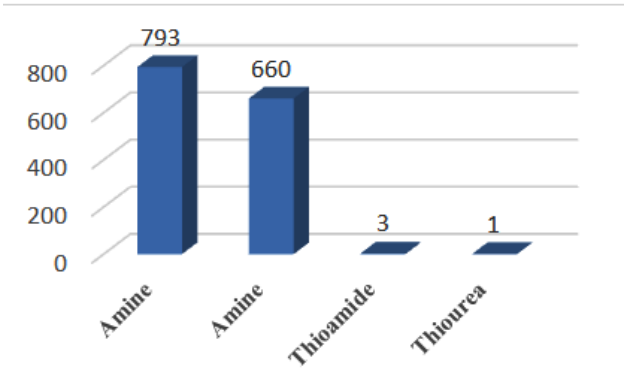
Group 7 : Functional Groups that have Aldehydes the carbonyl group is polar properties include fragmentation ether, HYDROPEROXIDE, peroxide, peroxide, Alcohol, Tertiary Amine, imine, azide, azo compound, Nitrate, Nitrite, Nitrile, isocyanide, NITROSO, aldehyde,



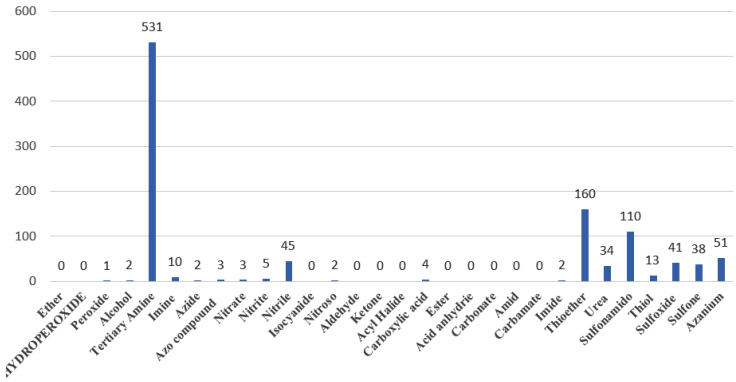
Group 4. Heterocyclic unsaturated moieties



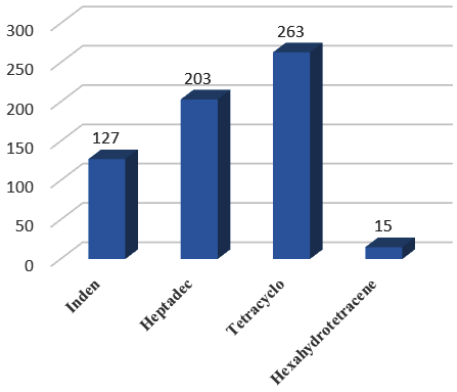
Group 5. Acids



Group 6. Amides



Group 7. Aldehydes



Group 8. Neutral aromatic moieties

Fig. 1. Statistical view of a number of functional groups and their major biological activities, (1) Saturated cyclic, (2) Unsaturated cyclic, (3) Heterocyclic compound unsaturated (4) Heterocyclic compound saturated (5) Acid (6) Amide, (7) Aldehydes the carbonyl group is polar and (8) Neutral aromatic, Polar Deactivated

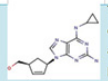
Generic Name & coc	The mechanism of actio	The main disease Categories	Structure	Description Cell target	SMILES	(FG)Cycloh	(FG)cyclopropane
		Anti-HIV Agents					
Abacavir DB01048	Anti-HIV Agents, Nucleoside and Nucleotide Reverse Transcriptase Inhibitors			This compound belongs to the class of organic compounds known as 1,3-substituted cyclopentyl purine nucleosides. These are nucleoside analogues with a	[(1S,4R)-4-[2-amino-6-(cyclopropylamino)-9H-purin-9-		
Amprenavir DB00701 (APRD00605)	Antibiotics, Antitubercular, HIV Protease Inhibitors Anti-HIV Agents			This compound belongs to the class of organic compounds known as aminobenzenesulfonamides. These are organic compounds containing a benzenesulfonamide	CC(C)CN(C[C@@H](O)[C@H](CC1=CC=CC=C1)NC(=O)O[C@H]1CCOC1)S(=O)(=O)C1=CC=C(N)C=C1		
Atazanavir DB01072 (APRD00804)	AgentsReverse Transcriptase Inhibitors			class of organic compounds known as phenylpyridines. These are polycyclic aromatic compounds	O)N[C@@H](CC1=CC=CC=C1)[C@@H](O)CN(CC1=CC=C(C=C1		
				This compound belongs to the			

Fig. 2. A view of the assembled database used in this work

ketone, acyl Halide, Carboxylic Acid, Ester, Acid Anhydrie, carbonate, carbonate, amid, carbamate, Imide, Thioether (Sulfide), urea, sulfonamide, thiol, Sulfoxide, sulfone, DISULFIDE, thione, thial, PHOSPHINE, borane, borinic acid, Borinic ester, boronic Acid, boronic Ester, silane, Silanol, silyl ether, azanium, Which are distinguished in blue in (Fig 1).

Group 8:Functional Groups that have Neutral aromatic, Polar Deactivated, Ether properties include fragmentation inden, heptadec, tetracyclo, hexahydrotetracene Which are distinguished by the gray color in (Fig 1).

In this study we aim to find a correlated between identified functional group to pharmaceutical structure. Using the SMILES code and their biological qualities corresponding to functional groups, available in each drug code.

The most frequent functional groups among the drug categories

Categuries	Number of drugs in this category	Total number of functional groups in this category
Anti-Bacterial Agents	88	338
Antineoplastic Agents	77	178
Antihypertensive Agents	74	177
Prostaglandins	54	96
Anti-Arrhythmia Agents	53	145
Dietary supplement	50	75
Acetohydroxamic Acid	47	123
Anti depressive Agents	45	133
Enzyme Inhibitors	40	96
Hypnotics and Sedatives	40	116
Penicillins	40	175
Antibiotics	40	142
Micronutrients	40	54
Anticonvulsants	39	78
Analgesics	37	97
Anti-inflammatory	37	81
Histamine H1 Antagonists	35	98
Benzodiazepines	35	82
Cephalosporins	35	162
Corticosteroids	33	80
Antifungal Agents	32	82
Platelet Aggregation Inhibitors	30	57
Sympathomimetics	30	54
Vasodilator Agents	26	71
Antipsychotic Agents	25	55
Antiviral Agents	24	52
Anti-Allergic Agents	24	65
Antipsychotic Agents: Dopamine	23	64
Antagonists¡ Sedatives and Hypnotics		
Bone Density Conservation Agents	23	48
Antidyskinetics	22	47
Central Nervous System Stimulants,Stimulants	22	53
Narcotics	21	69
Analgesics, Opioid	20	61
Antiemetics	20	58
Anti-anxiety Agents	20	22
GABA Modulators	20	31
Sympatholytics	20	42
Antiparkinson Agents	20	43
Bronchodilator Agents	20	38
Muscarinic Antagonists	20	38
Immunosuppressive Agents	20	46

Referrals and information sources that have been extracted from the data collected in this article, first, the Canada drug bank version 4.2, the pubchem site is a data base of small molecules, ChemSpider is a database of chemicals., from Wikipedia, the free encyclopedia . This study will help us to detect elucidate more biological features and properties in the drugs discovery. correlated between identified functional group to pharmaceutical structure will enable Researchers and pharmaceutical designers are more likely to design drugs based on the fragment codes contained in each drug straightener and the properties that appear in each treatment group. Researchers and pharmaceutical designers are more likely to design drugs by considering the fragment codes contained in each drug straightener and the properties appearing in each treatment group. At the first stage, the Canada drug bank version 4.2 was downloaded (http:// drugbank.ca). This dataset consists of nearly 9,000 unique DRUGCARD identification having the chemical, pharmacological and pharmaceutical features of drugs. At the next stage, In order to select the desired features and have a better and convenient tabular format, At the next stage,In order to select the desired features and have a better and convenient tabular format, All information extracted from these resources is categorized and this tabular dataset was imported into Microsoft Excel to extract the desired columns (Fig. 2).

Distribution:122 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 1 sheet 1).This database holds the 8 column numberal, Generic Name and code, The mechanism of action, The main disease (Categories), Drug Structure, Description, Cell target, SMILES code , Generic Name, IUPAC Name of each drug and 114 column Functional groups and is ready for future processing.

In the next step, we downloaded the image of each structure from the Sources mentioned above. 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.

In this study, we aim to find the predominant functional group is dominant in every category of treatment. We are interested in using the valuable association of functional groups and their biological activity with the types of categories in which the drug code is categorized. And the optimal use of this valuable information in drug design.

In the database prepared for all predominant functional drug codes, we set the dominant group, according to SMILES code and drug structure.

DISCUSSION

Our analysis shown that benezene functional group in Dopamine Agonists category has higher frequency between other functional groups ($22/23 \times 100 = 95\%$) and in Amphetamines, Cyclooxygenase Inhibitors, Estrogen Antagonists and Dopamine Agents categuries, frequencies are 57%, 51%, 50% and 50%, respectively. Dopamine Agonists is a major drug group in treatment of diseases that affects the nerve cells, for instance monotherapy resulted in an 87% lower risk for dyskinesia effect of levodopa in Parkinson's disease (PD) patients (12-14). A meta analysis explored current knowledge about the organelle-targeting features of small molecules, in this study approved drug molecules derived from the DrugBank database and random organic molecules derived from the PubChem database therefore meta analysis data shown that weight and hemical structure of small moleculars are important parameter affecting transport properties and drug-likelihood (15). A chemical ontology tool based on chemical functional groups drived from PubChem and the small molecule interaction database automatically was developed to categorize small molecules, it applicable for searching chemical databases and identifying key functional groups responsible for biological activities (16).

The high-throughput screening (HTS) systems in the drug discovery process is used for reliable analysis of massive amounts of data. A new automated multidomain clustering method to NCI Anti-HIV-1 database is used to examine both the active and the inactive compounds. Large and small compound sets that defined both chemical families and potential pharmacophore points were discovered and structure-activity relationships aided by the unique classification method(17).

REFERENCES

1. Ertl P. An algorithm to identify functional groups in organic molecules. *Journal of Cheminformatics.* 2017; **9**(1):36.
2. Villanueva-Rosales N, Dumontier M, editors. Describing Chemical Functional Groups in OWL-DL for the Classification of Chemical Compounds. OWLED; 2007.
3. Braslavsky SE. Glossary of terms used in photochemistry, (IUPAC Recommendations 2006). *Pure and Applied Chemistry.* 2007; **79**(3):293-465.
4. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, *et al*. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic acids research.* 2006; **34**(suppl_1):D668-D72.
5. Feldman HJ, Dumontier M, Ling S, Haider N, Hogue CW. CO: A chemical ontology for identification of functional groups and semantic comparison of small molecules. *FEBS letters.* 2005; **579**(21):4685-91.
6. Mao F, Ni W, Xu X, Wang H, Wang J, Ji M, *et al*. Chemical structure-related drug-like criteria of global approved drugs. *Molecules.* 2016; **21**(1):75.
7. Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, *et al*. DrugBank 4.0: shedding new light on drug metabolism. *Nucleic acids research.* 2013; **42**(D1):D1091-D7.
8. Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A, *et al*. DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic acids research*. 2010; **39**(suppl_1):D1035-D41.
9. Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, *et al*. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic acids research.* 2007; **36**(suppl_1):D901-D6.
10. Awale M, Reymond J-L. Cluster analysis of the DrugBank chemical space using molecular quantum numbers. *Bioorganic & medicinal chemistry.* 2012; **20**(18):5372-8.
11. Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of chemical information and computer sciences.* 1988; **28**(1):31-6.
12. Chondrogiorgi M, Tatsioni A, Reichmann H, Konitsiotis S. Dopamine agonist monotherapy in Parkinson's disease and potential risk factors for dyskinesia: a meta analysis of levodopa controlled trials. *European journal of neurology.* 2014; **21**(3):433-40.
13. Stowe R, Ives N, Clarke C, Van Hilten J, Ferreira J, Hawker R, *et al*. Dopamine agonist therapy in early Parkinson's disease. *Cochrane Database of Systematic Reviews.* 2008(2).
14. Thorlund K, Wu P, Druyts E, Eapen S, Mills EJ. Nonergot dopamine-receptor agonists for treating Parkinson's disease–a network meta-analysis. Neuropsychiatric disease and treatment. 2014; **10**:767.
15. Zheng N, Tsai HN, Zhang X, Shedden K, Rosania GR. The subcellular distribution of small molecules: a meta-analysis. *Molecular pharmaceutics*. 2011; **8**(5):1611-8.
16. Feldman HJ, Dumontier M, Ling S, Haider N, Hogue CW. CO: A chemical ontology for identification of functional groups and semantic comparison of small molecules. *FEBS Lett.* 2005; **579**(21):4685-91.
17. Tamura SY, Bacha PA, Gruver HS, Nutt RF. Data analysis of high-throughput screening results: application of multidomain clustering to the NCI anti-HIV data set. *Journal of medicinal chemistry.* 2002; **45**(14):3082-93.