Journal of Molecular Graphics and Modelling 38 (2012) 324–333

Contents lists available at SciVerse ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Toward understanding the molecular basis for chemical allosteric modulator design

Qi Wang[a,b,1], Mingyue Zheng[c,1], Zhimin Huang[b], Xinyi Liu[b], Huchen Zhou[a], Yingyi Chen[b], Ting Shi[b], Jian Zhang[b,*]

[a] School of Pharmacy, Shanghai Jiao Tong University, Shanghai, 200240, China
[b] Department of Pathophysiology, Key Laboratory of Cell Differentiation and Apoptosis of Chinese Ministry of Education, Shanghai JiaoTong University, School of Medicine, Shanghai 200025, China
[c] Drug Discovery and Design Center, State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China

ARTICLE INFO

Article history:
Accepted 27 July 2012
Available online 8 August 2012

Keywords:
Allosteric regulation
Allosteric modulators
Structural rigidity
"Allosteric-like" rule

ABSTRACT

Among the regulation mechanisms of cellular function, allosteric regulation is the most direct, rapid and efficient. Due to the wider receptor selectivity and lower target-based toxicity, compared with orthosteric ligands, allosteric modulators are expected to play a larger role in pharmaceutical research and development. However, current difficulties, such as a low affinity and unknown structural features of potential allosteric small-molecules, usually obstruct the discovery of allosteric modulators. In this study, we compared known allosteric modulators with various compounds from different databases to unveil the structural and qualitative characteristics of allosteric modulators. The results show that allosteric modulators generally contain more hydrophobic scaffolds and have a higher structural rigidity, i.e., less rotatable bonds and more rings. Based on this analysis, an empirical rule was defined to determine the structural requirements for an allosteric modulator. It was found that a large proportion of allosteric modulators (80%) can be successfully retrieved by this "allosteric-like" filter, which shows good discriminatory power in identifying allosteric modulators. Therefore, the study provides deeper insight into the chemical properties of allosteric modulators and has a good potential for the design or optimization of allosteric compounds.

1. Introduction

With the advances in the techniques of molecular biology, our knowledge of the regulation mechanisms of cellular functions has expanded enormously in the past few decades [1]. Among these mechanisms, allostery, namely allosteric regulation, is regarded as the most direct, rapid and efficient regulatory mechanism [1]. It is defined as the regulation of protein function, structure and/or flexibility induced by the binding of a ligand or another protein at a site away from the orthosteric site [2]. This remote regulation effect is caused by the binding of the regulators to protein regions with a high conformational flexibility or by promoting the association or dissociation of oligomeric enzymes [3]. Allosteric regulation provides an effective mechanism for proteins to directly sense the alteration of the cellular milieu and to respond to maintain the homeostasis of the cell without a cellular energy cost [1]. In such regulation behaviors, the site other than the orthosteric site is defined as an allosteric site and is predominantly involved in the binding of metal ions and small molecules that are known as allosteric modulators.

The binding of an allosteric modulator can lead to the redistribution of protein conformation ensembles and can alter the rates of their interconversion, leading to positive or negative effects on protein function. Thus, allosteric binding sites have drawn increasing attention as novel targets for new drug development [2]. Chemical allosteric modulators provide several theoretical advantages over orthosteric ligands as potential therapeutic agents. For example, allosteric modulators have the potential to maintain their activity dependence and both temporal and spatial aspects of endogenous physiological signaling because of their quiescence in the absence of endogenous orthosteric activity. In addition, allosteric modulators can achieve a better selectivity due to the higher sequence divergence in allosteric sites, and the limited positive or negative cooperativity imposes a 'ceiling' on the magnitude of their allosteric effect. Therefore, allosteric modulators could be administered with a lower propensity toward target-based toxicity than orthosteric ligands [4].

Because of the potential advantages of allosteric modulators, allosteric sites are considered excellent drug targets, and the

* Corresponding author.
E-mail address: jian.zhang@sjtu.edu.cn (J. Zhang).
[1] Q.W. and M.Z. contributed equally to this work.

1093-3263/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmgm.2012.07.006

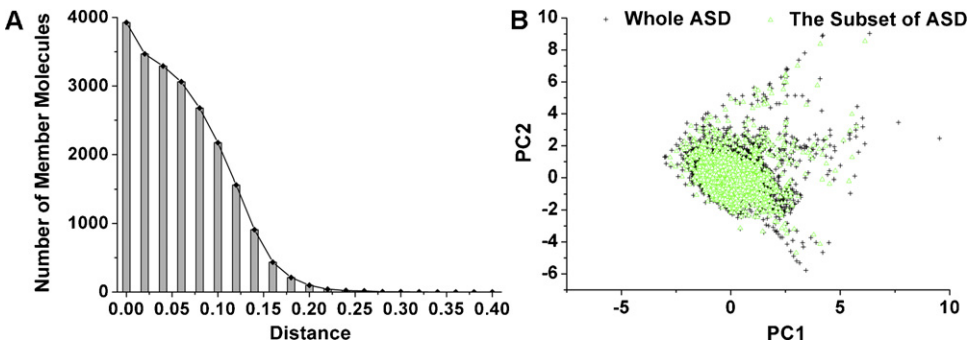


Fig. 1. (A) The number of compounds in the corresponding subset of the original ASD versus the distance parameter in the clustering. The distance represents a maximum dissimilarity calculated by "1 – Tanimoto similarity of molecular fingerprint FCFP4". (B) The distribution of compounds from the original ASD (7851 compounds, black plus sign) and ASD dataset (3916 compounds, green triangle) in the plot of the first two principal components extracted with the MOE descriptors.

discovery of more allosteric modulators could advance the development of novel drugs toward certain human diseases. The control of oxygen binding to hemoglobin by protons and the AMP-induced activation of glycogen phosphorylase are two of the earliest discovered cases of allosteric regulation [5,6]. As the investigation of this phenomenon moved forward, allosteric regulation aroused extensive interest and was defined by Jacob and Monod [7]. By the 1970s, approximately one dozen allosteric modulators had been discovered [8–10], and in recent years, a series of allosteric drugs targeting kinases, GPCRs and ion channels have been discovered, such as the development of Gleevec (allosteric inhibitor of Abl) [11], Cinacalcet (allosteric activator of a calcium-sensing receptor) [4] and Maraviroc (allosteric inhibitor of a chemokine receptor) [4]. Although much progress has been made in the development of allosteric modulators, we have not yet discovered a general method that can be widely used to deliberately discover allosteric modulators, other than the systematic discovery of genes and proteins that could be targeted. In the past, allosteric modulators were primarily found either by chance or by intuition followed by experimental verification [1]. Recently, a few allosteric modulators have been discovered by high-throughput screening (HTS). However, the allosteric small molecule "hits" typically have low affinities and unknown chemical compositions unsuitable for the discovery of allosteric modulators [1]. These problems could be alleviated by screening the potential for allosteric modulators in an "allosteric-like" small-molecule library and optimizing the structure of the hit according to the "allosteric-like" rule. Unfortunately, we have little knowledge about such rational design methods for chemical allosteric modulators.

To understand the inherent nature of allosteric modulators, the structural and physicochemical characteristics of allosteric modulators were systematically analyzed by comparing with various compounds from other chemical resources. Based on the analysis, an empirical rule was defined to determine the structural requirements for an allosteric modulator. Taken together, the study provides useful information for understanding the chemical properties of allosteric modulators and is a good guideline for the design or optimization of allosteric modulators.

2. Materials and methods

2.1. Databases

Six databases of compounds were collected for this research. As the primary research target, allosteric modulators were obtained from the AlloSteric Database (ASD), which is a recent database that provides a central resource for allosteric molecules [12]. In the ASD, 7851 organic allosteric modulators were manually curated from the literature. The number of reported allosteric modulators can differ significantly between allosteric proteins due to the different amount of research effort for different targets [12]. To avoid the potential bias caused by the inhomogeneous distribution of modulators, a refined subset with 3916 diverse allosteric modulators is provided to users of the ASD by the clustering method in Scitegic Pipeline Pilot v7.5 [13], as described below. The clustering analysis was performed with the "Cluster Molecules" component in the "Analysis and Statistics" module of Pipeline Pilot 7.5. In the component, the algorithm is a partitioning method, and a number of representative compounds are chosen as cluster centers to derive a subset from the original dataset by the maximum dissimilarity method. In each cluster, the distance between the center compound and member compounds is calculated by "one minus the Tanimoto similarity of molecular fingerprint FCFP4". The clustering begins by randomly choosing a compound as the first cluster center. The compound maximally distant from the first point is selected as the next cluster center. The compound maximally distant from both current points is selected after that. The non-selected objects are then assigned to the nearest cluster center to determine the cluster membership. The process repeats itself until there are the largest number of cluster centers and at least 2 compounds in each cluster. From this process, 3916 clusters were built, and the center compounds were extracted as the ASD subset. In every cluster, the maximum distance between the center and the members was 0.34 (Fig. 1A). To evaluate the final subset, descriptors of all compounds were calculated with the MOE 2008 modeling suite [14]. Twenty-seven common descriptors in drug discovery were calculated by MOE, including *weight* (molecular weight), *SlogP* (log of the octanol/water partition coefficient), *TPSA* (polar surface area), *a_aro* (number of aromatic atoms), *a_count* (number of atoms), *a_heavy* (number of heavy atoms), *a_nC* (number of carbon atoms), *a_nN* (number of nitrogen atoms), *a_nO* (number of oxygen atoms), *a_nP* (number of phosphorus atoms), *a_nS* (number of sulfur atoms), *a_nF* (number of fluorine atoms), *a_nCl* (number of chlorine atoms), *a_nBr* (number of bromine atoms), *a_nI* (number of iodine atoms), *b_ar* (number of aromatic bonds), *b_count* (number of bonds), *b_single* (number of single bonds), b_double (number of double bonds), *b_triple* (number of triple bonds), *b_heavy* (number of bonds between heavy atoms), *b_rotN* (number of rotatable bonds), *b_rotR* (fraction of rotatable bonds), *chiral* (number of chiral centers), *rings* (number of rings), *lip_acc* (number of O and N atoms) and *lip_don* (number of OH and NH atoms). In addition, principle component analysis (PCA) was used to analyze the relationship between the final subset and the original dataset, and the total variance represented within the first two principal components was 95%. As shown in Fig. 1B, the distribution of the subset is highly overlapped with the whole dataset, indicating that it is well balanced and is representative of the allosteric modulators. Therefore, the final subset (hereinafter referred to as the ASD dataset)

Table 1
Summary of the databases used in this study.

Name	Date	Size	Source
ASD (original)	2010.8	7851	http://mdl.shsmu.edu.cn/ASD
ASD dataset	2010.10	3926	http://mdl.shsmu.edu.cn/ASD
ACD	2005.9	582,301	http://accelrys.com
CMC	2009	9098	http://accelrys.com
CNPD	2005.6	54,276	http://www.neotrident.com
DrugBank	2010.1	4886	http://www.DrugBank.ca
MDDR	2009	197,549	http://accelrys.com
NCI	2003.9	260,071	http://cactus.nci.nih.gov

was selected to explore the features of allosteric modulators in this study.

For comparison, other compounds were collected from the following databases: Accelrys Available Chemicals Directory (ACD) [15], Accelrys Comprehensive Medicinal Chemistry (CMC) [16], Chinese Natural Product Database (CNPD) [17], DrugBank [18–20], MDDR [21], and NCI Open Database [22]. The general information of the databases above is shown in Table 1. Accelrys ACD is one of the largest structure-searchable collections of commercially available chemicals in the world. Accelrys CMC accumulates compounds used or studied as medicinal agents in humans. CNPD is the largest repository for natural products in China. DrugBank is a comprehensive database for drugs approved by the FDA. MDDR is a database covering biologically relevant compounds and well-defined derivatives. The NCI open database has collected samples of compounds from the National Cancer Institute from both organic synthesis and natural source extracts for testing as drugs for cancer and AIDS. The allosteric modulators, peptide-like compounds with more than three residues and metal ions were removed from the databases. After that, all compounds in these databases were converted to SDF format for calculation, and some SDF files in the databases lacking the necessary tags were fixed by in-house Python scripts.

2.2. Molecular descriptors

Eighteen molecular descriptors widely used in drug discovery were calculated to evaluate compounds in all datasets: molecular weight (MW), average molecular weight (AMW, molecular weight per atom), number of atoms (nAT), number of bonds (nBT), number of rotatable bonds (RBN), rotatable bond fraction (RBF), unsaturation index (Ui), number of rings (nR), number of circuits (nCIR), number of ring systems (nRS), number of rings in the largest ring system (nRIS), complexity of ring systems (CRS), number of hydrogen bond acceptors (nHAcc), number of hydrogen bond donors (nHDon), hydrogen bond acceptors fraction (HAccF), hydrogen bond donors fraction (HDonF), SlogP and topological polar surface area (TPSA).

AMW is calculated by dividing the molecular weight by the number of atoms; RBF is the number of rotatable bonds divided by the number of bonds in a molecule; and Ui is defined as the degree of unsaturation in a molecule as Eq. (1) [23]:

$$\mathrm{Ui} = \log_2(1 + \mathrm{nDB} + \mathrm{nTB} + \mathrm{nAB}) \quad (1)$$

where nDB, nTB and nAB are the number of double, triple and aromatic bonds, respectively.

Rings here refer to the smallest set of smallest rings (SSSR) in a molecule, and a circuit means a self-returning cyclic path starting and ending on the same vertex (atom). A ring is an independent circuit because there are no repeated vertices (atoms) other than the first and last atoms. A ring system is a set of rings with fused or spiro connections that are linked to other ring systems with only single and double bonds (not with ring bonds) [24]. Then, nRS derived from the Euler formula is defined as Eq. (2):

$$\mathrm{nRS} = (e_t - e_r) - (v_t - v_r) + 1 \quad (2)$$

where e_t and v_t are the total number of edges (bonds) and vertices (atoms); e_r and v_r are the number of edges and vertices within rings in a molecule; nRIS is the number of rings in the ring system that contains the most rings in a molecule; and CRS is obtained by dividing the number of rings by the number of ring systems in a molecule.

The definition of hydrogen bond acceptors and donors can differ in the literature. Here, we used the classical definitions as described by Lipinski [25]. Then, nHAcc is obtained as the sum of the N and O atoms, and nHDon is expressed as the sum of the —NH and —OH groups. HAccF is calculated by dividing the number of hydrogen bond acceptors by the number of atoms. Similarly, HDonF is obtained by dividing the number of hydrogen bond donors by the number of atoms.

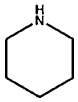
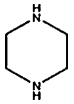
2.3. Statistical analysis

The normality of the distributions was evaluated by the "Normality test (Shapiro–Wilk)" method in the "Statistics" module of Origin 8.0. To evaluate the different distributions of the descriptors between the allosteric dataset and the other databases, "Two sample *t*-test" and the "Mann–Whitney test" in the "Statistics" module of Origin 8.0 were performed, and the statistical results (*P*-values) and standard error bars are shown in each figure.

Table 2
Mean values of the 18 molecular descriptors of the compounds from the seven databases in the study.

Descriptors	ASD	ACD	CMC	CNPD	DrugBank	MDDR	NCI
Molecular weight (MW)	391	307	374	400	346	462	311
Average molecular weight (AMW)	8.35	8.39	7.84	7.14	8.10	7.83	8.30
Number of atoms (nAT)	48.1	38.0	49.8	57.3	44.8	60.6	39.1
Number of bonds (nBT)	50.6	39.2	51.6	59.6	46.0	63.2	40.5
Number of rotatable bonds (RBN)	7.47	7.07	9.33	10.7	9.63	11.0	6.84
Rotatable bond fraction (RBF)	0.15	0.17	0.18	0.18	0.21	0.16	0.16
Unsaturation index (Ui)	3.81	3.25	3.17	2.75	2.83	3.78	3.13
Number of rings (nR)	3.50	2.23	2.80	3.29	2.29	3.67	2.38
Number of circuits (nCIR)	7.94	2.80	4.82	8.48	4.37	5.36	3.87
Number of ring systems (nRS)	2.73	1.83	1.84	1.53	1.58	2.41	1.56
Number of rings in the largest RS (nRIS)	1.70	1.33	1.75	2.41	1.49	1.81	1.58
Complexity of ring systems (CRS)	1.39	1.20	1.55	2.26	1.32	1.56	1.43
Number of H-bond acceptors (nHAcc)	6.34	5.02	6.17	6.50	7.09	7.78	4.93
Number of H-bond donors (nHDon)	1.00	1.16	1.87	3.54	2.29	1.54	1.25
H-bond acceptors fraction (HAccF)	0.137	0.137	0.131	0.116	0.171	0.131	0.136
H-bond Donors Fraction (HDonF)	0.022	0.033	0.038	0.061	0.057	0.025	0.033
SlogP	3.98	2.87	2.32	2.77	0.91	3.54	2.76
Topological polar surface area (TPSA)	75.4	70.1	96.6	100	113.3	110.2	70.1

Table 3
The top ten ring systems and unique ring systems in the 'single ring' type in the ASD dataset.

Top ring systems			Unique ring systems[a]		
Code	Structure	Quantity	Code	Structure	Quantity
ASa1		4588	pASa1		6
ASa2		722	pASa2		2
ASa3		373	pASa3		1
ASa4		224			
ASa5		208			
ASa6		202			
ASa7		190			
ASa8		182			
ASa9		147			
ASa10		143			

[a] Structures are ring scaffolds frequently found in allosteric modulators but not in other reference databases.

2.4. Ring system analysis

All ring systems in both the ASD and ACD datasets were extracted and counted in descending order with a Python script with the openbabel module [26] (http://openbabel.org). Three types of ring systems (single ring and fused ring including two or more than two single rings) were summarized from the collection, and the top ten are shown in Tables 3–5.

2.5. Software

Scitegic Pipeline Pilot v7.5 [13] was used to unify the SDF format of the compounds in each database. For descriptors, SlogP and TPSA were calculated with the MOE 2008 modeling suite [14]; nRS, nRIS, CRS, HAccF and HDonF were calculated by Python scripts with the openbabel module [26]; and others were calculated by Dragon software [23]. Ring systems were parsed by Python scripts with the openbabel module [26], and statistical analyses were performed with Origin 8.0 software [27].

3. Results and discussion

The mean values of the eighteen molecular descriptors for all of the databases are shown in Table 2. The distributions of the properties for the ASD dataset and the comparison between the ASD dataset and the other databases are described in the figures in Supporting information.

3.1. Molecular weight, number of atoms and bonds

The mean value of the MW of the ASD dataset is located in the middle of these six databases (see Table 2 and Fig. 2A), and approximately 80% of the allosteric modulators have a MW in the range of 200–500 (see Fig. S1). As a whole, the MW of the allosteric modulators is apparently only similar to that of the natural products from the CNPD. The AMW of the allosteric modulators is obviously higher (Fig. 2B), and the nAT and nBT of the allosteric modulators are less than those of the compounds in the CNPD (see Fig. 2C and D). The differences in the properties are mainly derived from the constituent elements of the compounds. The occurrence of heavy atoms, such as sulfur and halogen atoms (Cl, Br, I), in the ASD dataset (50.61%) is much higher than in the natural products (CNPD 6.69%), drugs (DrugBank 29.32%, CMC 35.03%) and biologically relevant compounds (MDDR 44.57%), indicating the potential impact of hydrophobic heteroatoms in the design of allosteric compounds. In addition, the MW and nAT of the allosteric modulators are slightly higher than those of the current drugs (DrugBank), and this might be considered in the allosteric drug discovery.

3.2. Rotatable bonds, rotatable bond fraction and unsaturation index

The number of rotatable bonds serves as one avenue for describing the rigidity of a chemical molecule. As shown in Table 2 and Fig. 3A, the average RBN of allosteric modulators is less than 8, which is significantly less than the compounds in the known drug or bioactive databases, i.e., CNPD, CMC, DrugBank and MDDR (mean ranges from 9 to 11). Given the distribution of the RBN in allosteric modulators (Fig. S3), nearly half of the allosteric modulators are covered by an RBN range from 4 to 8, and no more than 15% of the modulators have an RBN above 11. To better elaborate the rigidity of a molecule, the ratio of the rotatable bonds in the structure is evaluated as RBF, which is generated by dividing the RBN by the number of bonds (see Fig. 3B). Allosteric modulators show the lowest RBF (15%) among the drug and bioactive databases, confirming their high structural rigidity.

The unsaturation index is another common descriptor for characterizing the rigidity of chemicals. The degree of unsaturation strongly depends on the number of double, triple and aromatic bonds, which are included in the non-rotatable bonds. As expected, the difference in the Ui in all of the databases is in accordance with the result drawn from the RBF (see Fig. 3C). The allosteric modulators show a higher Ui than the bioactive compounds in the CMC, DrugBank, and CNPD; thus, the modulators could suffer from less conformational changes in the process of binding to their target.

3.3. Ring topology

Various rings are essential components in chemical molecules and significantly contribute to both spatial characteristics and physicochemical properties of the molecule. To investigate the features in allosteric modulators, ring topology was analyzed on the ASD and six other datasets. The results show that more than 98% of the allosteric modulators have ring structures according to the SSSR analysis, and the average nR in the allosteric modulators is greater than 3.5 (Table 2 and Fig. 4A), which is far more numerous

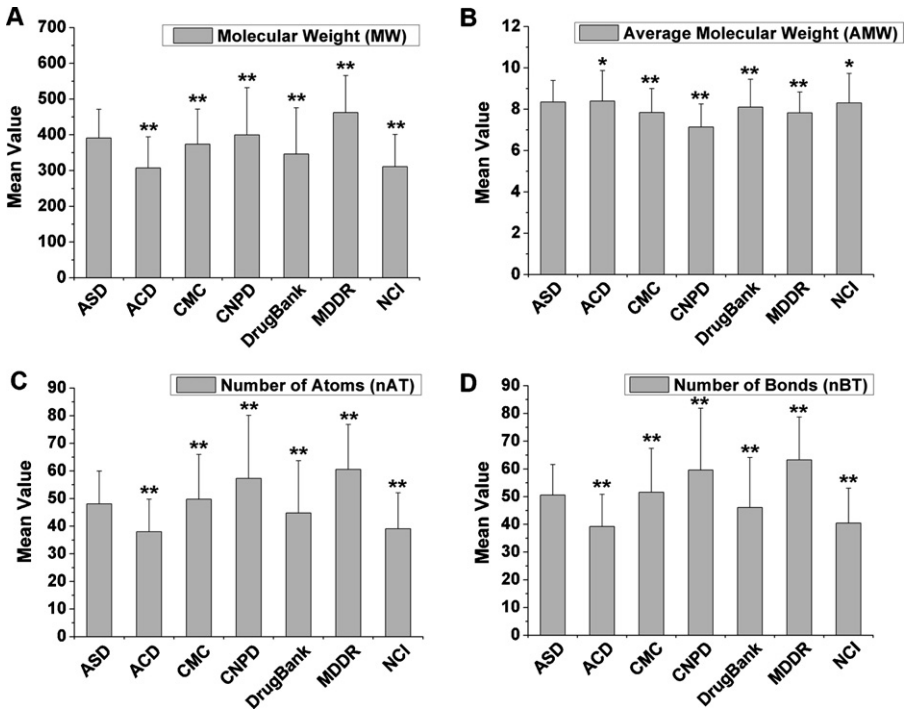


Fig. 2. The mean values of (A) MW, (B) AMW, (C) nAT and (D) nBT of the compounds in the seven databases used in this study. * and ** indicate that the property in the database is statistically different from that in the ASD dataset at the level of 0.05 and 0.01, respectively.

than that of drugs in the DrugBank (mean: 2.3) and is comparable to that of the natural products in the CNPD and the bioactive compounds of the MDDR. Nevertheless, only the natural products among all of the datasets have a similar number of circuits to the allosteric modulators (nCIR in Table 2 and Fig. 4B). Rings or circuits are basic constituent elements of a molecular scaffold, and many natural products are constructed on the basis of a rigid scaffold by the combination of complex rings or circuits. Thus, the existence of comparable features in the ASD dataset highlights the frequent occurrence of rings and circuits in the allosteric scaffold.

Due to the high diversity of rings in chemical structures, SSSR alone cannot fully reflect the global view of rings for allosteric modulators, and the analysis of the ring systems could be more appropriate to characterize the ring effect. It should be noted that in this study, a ring is defined as the smallest single circuit, and a ring system is defined as sets of rings with fused and spiro internal

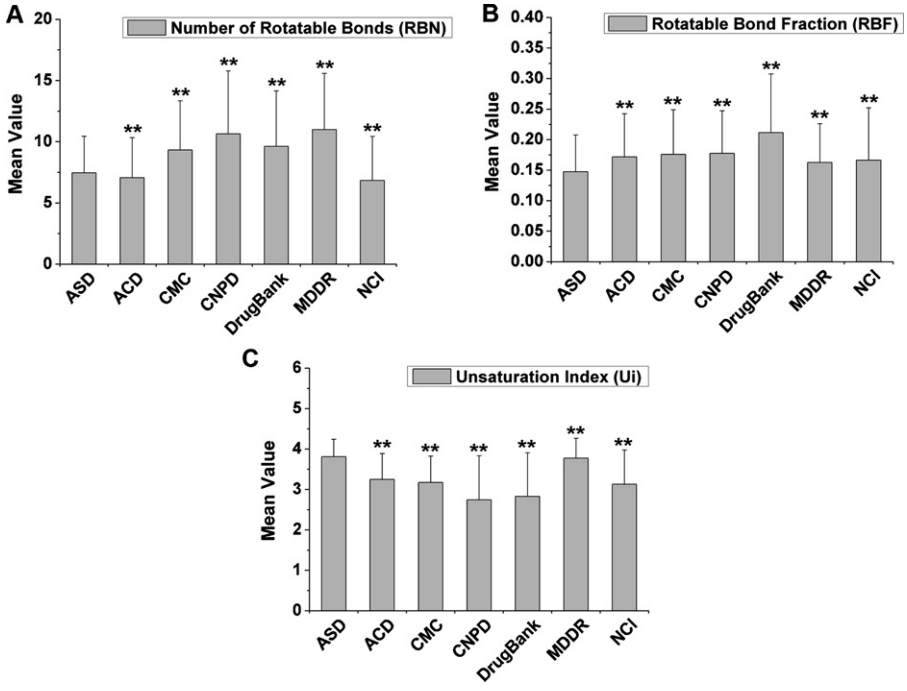


Fig. 3. The mean values of (A) RBN, (B) RBF and (C) Ui of the compounds in the seven databases used in this study. ** Indicates that the property in the database is statistically different from that in the ASD dataset at the level of 0.01.

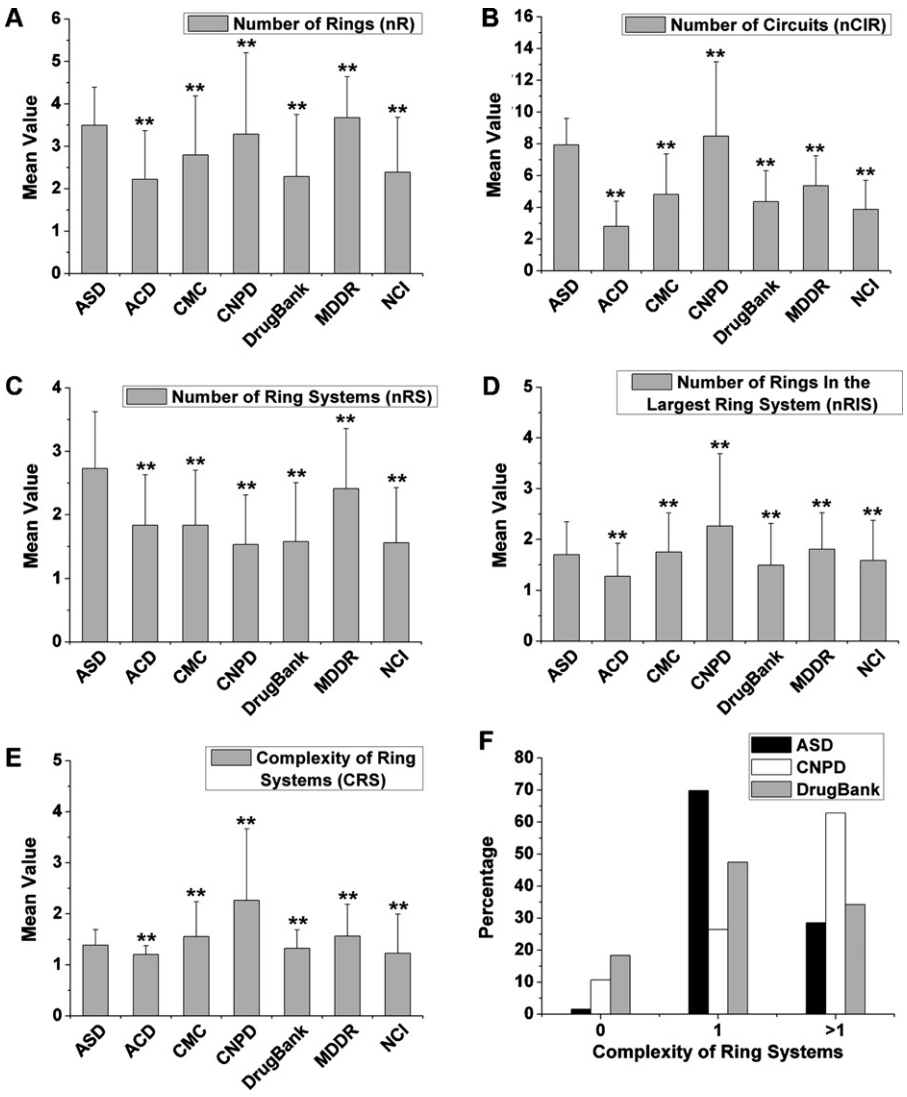


Fig. 4. The mean values of (A) nR, (B) nCIR, (C) nRS, (D) nRIS and (E) CRS of the compounds in the seven databases used in this study. (F) CRS comparison of the compounds among the ASD dataset, CNPD and DrugBank. ** indicates that the property in the database is statistically different from that in the ASD dataset at the level of 0.01.

connections. Principally, any two-ring system is linked only by single or double bonds, and three types of ring systems are described in Section 2.4 of Section 2. Three advanced descriptors of ring systems (nRS, nRIS and CRS) were calculated for all seven datasets. The highest nRS was found in the ASD dataset (Fig. 4C), and the nR values of the ASD and CNPD are similar, implying a lower ring system complexity for allosteric modulators than for natural products. Furthermore, on average, the nRIS is less than 2 in the ASD dataset, demonstrating the uncomplicated composition of the ring systems (Fig. 4D). The CRS index in the ASD dataset reveals a moderate complexity of the ring systems as a whole (Fig. 4E). Remarkably, these two descriptors in the CNPD are the highest of all of the databases, suggesting that they may be used to differentiate allosteric modulators from natural products.

By comparing these descriptors, the importance of the ring profile of allosteric modulators emerged. Although the number of rings between allosteric modulators and natural products is not different, the number of ring systems in allosteric modulators is on average more than that in natural products, indicating that complexity of ring systems in allosteric modulators could be lower than that in natural products. Actually, most of the allosteric modulators were developed in the pipeline of drug discovery, and the trends of compounds in the ASD dataset were compared to the drugs in the DrugBank. Unlike the monotonic increasing CRS in the CNPD, allosteric modulators have a parabolic distribution of CRS similar to drugs (Fig. 4F). The proportions of molecules with fused and spiro rings (CRS > 1) in the ASD and DrugBank datasets are almost the same (approximately 30%) and much less than that in the CNPD (60%). In addition, 70% of allosteric modulators only have single ring systems (CRS = 1), in comparison with 47% of drugs, and few allosteric modulators are linear (CRS = 0), while 23% of drugs lack rings. Thus, this analysis reveals the necessity of a rigid ring as well as the quantity and complexity of rings in allosteric modulators.

Bioactive compounds with few flexible conformations have been found to play an important role in several fields. For example, amantadine is an inhibitor of M2 protein of the influenza A virus [28–30], glucocorticoids are modulators of the cytosolic glucocorticoid receptor [31] and fluorouracil is an inhibitor of thymidylate synthase [32]. All such compounds are rather rigid and have an effect on their targets due to steric hindrance. Based on the finding in this study, less rotatable bonds and more rings from multiple moderate ring systems are two prominent features for allosteric modulators. Accordingly, it is further inferred, from the known cases listed above, that allosteric modulators regulate the conformational changes of their targets by the effect of steric hindrance in allosteric sites.

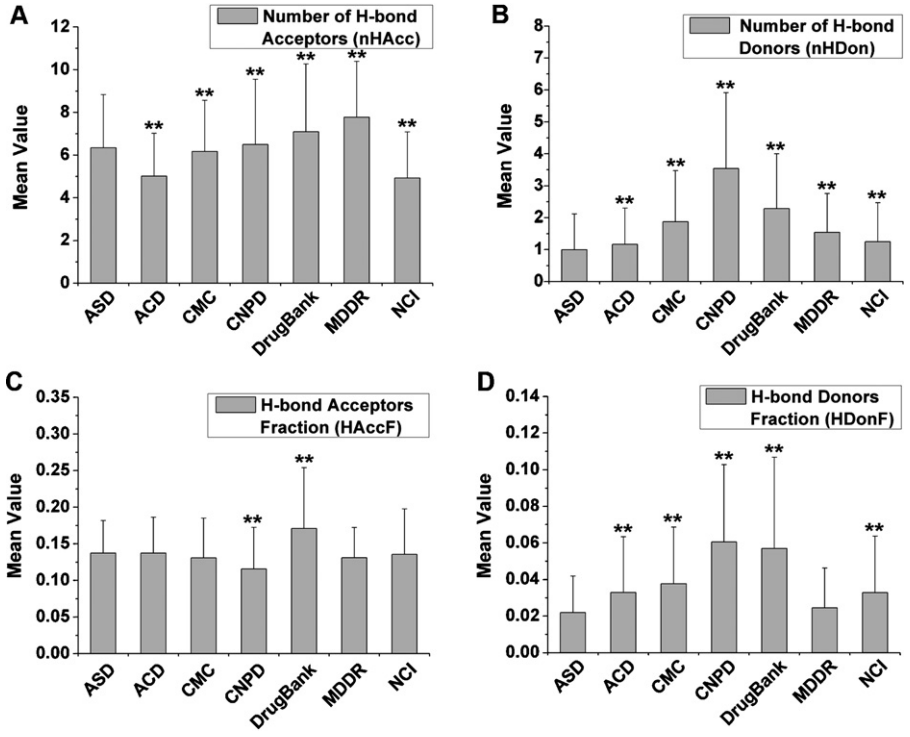


Fig. 5. The mean values of (A) nHAcc, (B) nHDon, (C) HAccF and (D) HDonF of the compounds in the seven databases used in this study. * and ** indicate that the property in the database is statistically different from that in the ASD dataset at the level of 0.05 and 0.01, respectively.

3.4. Hydrogen bond acceptors and donors

Considering the broad prospects of allosteric modulators in drug discovery, an understanding of the interactions with their binding targets will provide us with deeper insight into the design of allosteric molecules [24]. Hydrogen bonding is one of the most important factors for the specific binding of a ligand to its receptor. The attractive interaction is formed between a hydrogen atom from a molecular fragment X–H (donor) in which X is more electronegative than H, and an electronegative atom such as fluorine, oxygen or nitrogen (acceptor). As shown in Fig. 5, the nHAcc and HAccF of allosteric modulators are both located at a moderate level among the databases, and the nHDon and HDonF of allosteric modulators are even lower than those of the other bioactive compounds (CMC, CNPD, DrugBank and MDDR). The results indicate that fewer —OH and —NH groups have been found in allosteric modulators and may thus confine hydrogen bond donors to a relatively lower contribution to the binding affinities of allosteric interactions in comparison with other bioactive compounds.

3.5. SlogP and topological polar surface area (TPSA)

As mentioned above, a high ratio of hydrophobic heteroatoms to polar atoms was found in the ASD dataset, implying the importance of hydrophobicity to the structure of allosteric modulators. To further examine the effect, SlogP and TPSA, which are measures of how hydrophilic or hydrophobic a chemical substance is, were evaluated, and the results are shown in Fig. 6. In accordance with the analysis above, the highest SlogP and the lowest TPSA were found in the allosteric modulators compared to the other bioactive compounds (DrugBank, CMC, CNPD and MDDR), which again demonstrates the hydrophobic bias in allosteric modulators and

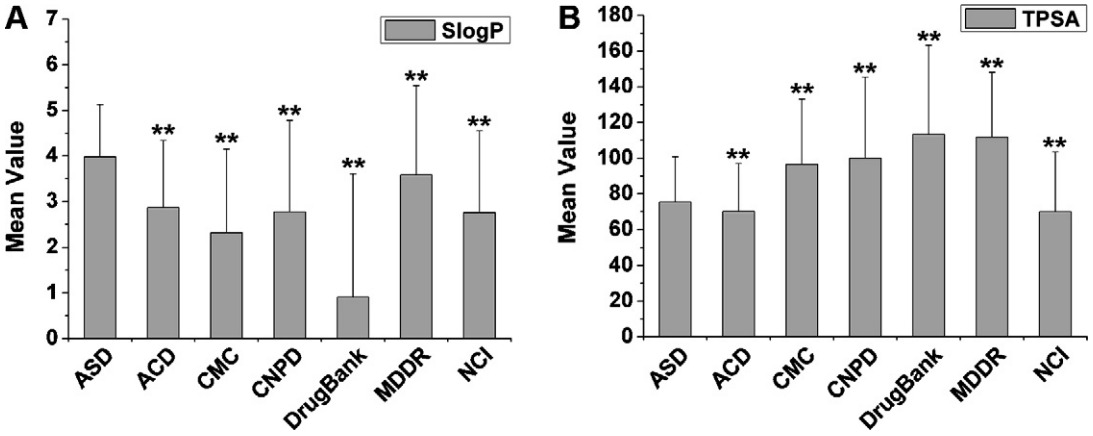
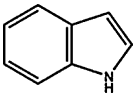
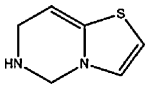
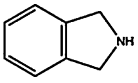
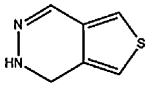
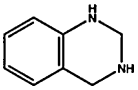
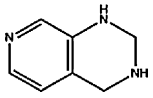
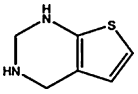
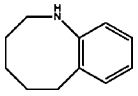
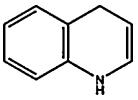
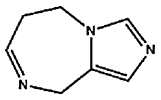
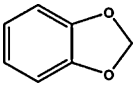
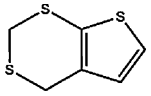
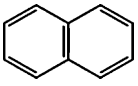
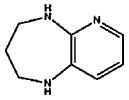
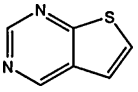
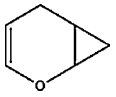
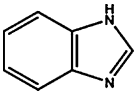
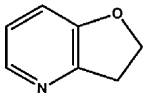
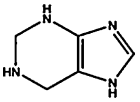
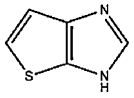


Fig. 6. The mean values of (A) SlogP and (B) TPSA of the compounds in the seven databases used in this study. ** Indicates that the property in the database is statistically different from the ASD dataset at the level of 0.01.

Table 4
The top ten ring systems and unique ring systems in the 'fused rings including two rings' type in the ASD dataset.

Top ring systems			Unique ring systems[a]		
Code	Structure	Quantity	Code	Structure	Quantity
ASb1		207	pASb1		33
ASb2		99	pASb2		7
ASb3		97	pASb3		4
ASb4		85	pASb4		4
ASb5		83	pASb5		3
ASb6		65	pASb6		3
ASb7		65	pASb7		3
ASb8		62	pASb8		3
ASb9		58	pASb9		3
ASb10		57	pASb10		2

[a] Structures are ring scaffolds frequently found in allosteric modulators but not in other reference databases.

indicates that the ligand-receptor binding of allosteric modulators could be deeply affected by hydrophobic interactions.

3.6. Ring systems analysis

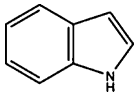
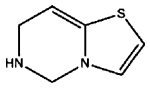
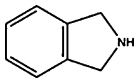
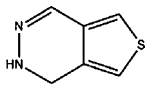
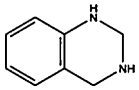
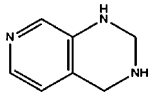
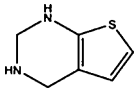
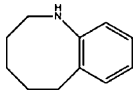
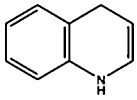
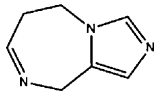
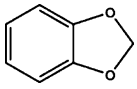
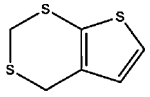
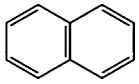
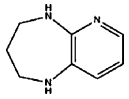
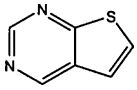
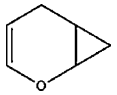
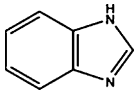
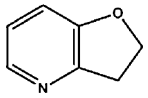
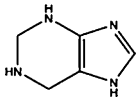
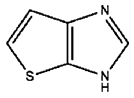
The scaffold in drug design is the fixed part of a molecule on which the functional groups are substituted or exchanged. Ring systems are the most represented scaffold and have been found to influence target selectivity [33]. To elucidate the preference for ring systems in allosteric modulators, all of the ring systems in compounds from the ASD dataset were extracted and divided into three categories: (i) single rings, (ii) fused rings including two rings, and (iii) fused rings including more than two rings. In addition, the top 10 scaffolds in the ASD not present within the other reference datasets are also listed in the three categories. In each category, all of the ring scaffolds were sorted in descending order of frequency, and the top ten scaffolds with their structures and quantities are presented (Tables 3–5). This identification provides a full profile of the allosteric frame, and the rational design of allosteric modulators could therefore really benefit from these known scaffolds.

3.7. A rule for guiding the design of chemical allosteric modulators

Considering the lack of a systematic method for allosteric modulator discovery, chemical screening is still the major strategy for finding allosteric hits [1]. In drug discovery, it is desirable to screen a focused chemical library against a drug target, which could result in an improved "hit" rate from the screening [34]. Therefore, the allosteric features of the chemical library are crucial for the screening of allosteric modulators. Today, more than 100 libraries covering 20 million compounds are available from academia and commercial vendors for screening purposes [35]. For example, two libraries, the NCI and ACD, were selected to explore the bias of their chemical space for allosteric modulators (Table 1). The NCI and ACD compounds had a smaller MW (Fig. 2A and C), fewer rings and a lower hydrophobicity (Figs. 4A and 6A) in comparison with allosteric chemical space.

To facilitate the understanding of the chemical characteristics of allosteric modulators, an empirical rule was defined based on the analyses of allosteric modulators in this study. All 18 descriptors and all of the compounds were imported into the "association

Table 5
The top ten ring systems and unique ring systems in the 'fused rings including more than two rings' type in the ASD dataset.

Top ring systems			Unique ring systems[a]		
Code	Structure	Quantity	Code	Structure	Quantity
ASc1		37	pASc1		32
ASc2		32	pASc2		16
ASc3		22	pASc3		11
ASc4		19	pASc4		10
ASc5		16	pASc5		9
ASc6		11	pASc6		6
ASc7		10	pASc7		5
ASc8		10	pASc8		3
ASc9		10	pASc9		2
ASc10		9	pASc10		2

[a] Structures are ring scaffolds frequently found in allosteric modulators but not in other reference databases.

rules learning method" module of WEKA [36] to find a combination of the most associated descriptors for retrieving allosteric modulators. After automatic optimization in the module, the combination of five descriptors (MW, RBN, nR, nRIS and SlogP) had the best scores. This process of optimization was independently repeated five times by different initial random seeds, and the results were the same. According to the five descriptors and their value ranges, the rule for allosteric modulators was built as follows: (i) $MW \leq 600$; (ii) $RBN \leq 6$; (iii) $2 \leq nR \leq 5$; (iv) $nRIS = 1$ or 2; and (v) $3 \leq SlogP \leq 7$. Following the rule, more than 80% of the allosteric modulators in the ASD dataset are successfully retrieved, and the different bias of compounds toward allosteric properties was compared to the other databases (Fig. 7). We have analyzed the distribution of violated descriptors in the remaining 20% compounds as following: 74.67% in the RBN, 72.56% in the SlogP (8.06% < 3 and 64.5% > 7), 41.56% in the nR, 34.98% in the nRIS, and 14.95% in the MW. Due to the inherent hydrophobicity of allosteric modulators, about 92% of allsoteric modulators can pass the SlogP filter if upper limit of the SlogP is not set in the rule. Actually, allsoteric modulators are mainly from three sources: allosteric drugs, allosteric tools for chemical biology and allosteric switches for bioengineering based on the curated literature. Upper limit of the SlogP should carefully reduce to avoid poor pharmacokinetics profile if the rule is used in the screening of allosteric drug discovery. On the other hand, the RBN filter blocks most of the remaining compounds, indicating that the item could be relaxed in the design of allosteric modulators if the "allosteric-like" rule does not work well in specific biological system. To further verify the perception, we have treated all reference datasets into subsets in a similar way and evaluated the chemical features between ASD and the reference subsets (Figs. S13–S17). The result shows that all distributions of the reference subsets are similar to those of their original datasets, and the "allosteric-like" rule can even differentiate ASD with the reference subsets better than the original databases (Fig. S18). Therefore, the rule summarized in this study depicts the chemical space of allosteric modulators and may provide a good guideline in the design or optimization of allosteric modulators.

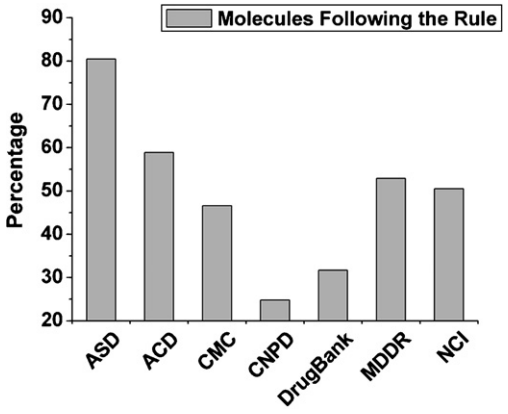


Fig. 7. The proportion of molecules following the "allosteric-like" rule in each database.

4. Conclusions

In this study, the chemical properties of allosteric modulators were analyzed to better understand the allosteric characteristics in medicinal chemistry. Altogether, eighteen molecular descriptors frequently used in drug discovery were evaluated for allosteric modulators and compounds from reference databases. The results show that less rotatable bonds and more rings from multiple moderate ring systems are two prominent features for allosteric modulators, and both of these features lead to rigid frames in the chemical structures. Higher hydrophobicity is also found for allosteric modulators in comparison with drugs and bioactive compounds from other databases. In addition, the most frequent ring scaffolds in allosteric modulators were identified. According to the analysis of allosteric modulators, a rule for simply deciphering the chemical characteristics of allosteric modulators was developed below: (i) $MW \leq 600$; (ii) $RBN \leq 6$; (iii) $2 \leq nR \leq 5$; (iv) $nRIS = 1$ or 2; and (v) $3 \leq SlogP \leq 7$. Therefore, the "allosteric-like" rule provides a preliminary filter for the identification of allosteric modulators, which could be used for the construction of focused libraries in the screening of allosteric modulators and as a guide for rational design and optimization of allosteric hits. The data generated in this study are available upon request.

Acknowledgements

This work was supported in part by grants from a Foundation for the Author of National Excellent Doctoral Dissertation of PR China (201083), the National Natural Science Foundation of China (21002062 and 81001399), the Innovative Research Team of Shanghai Municipal Education Commission, the Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning and the Shanghai PuJiang Program (10PJ406800).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jmgm.2012.07.006.

References

[1] J.E. Lindsley, J. Rutter, Whence cometh the allosterome? Proceedings of the National Academy of Sciences of the United States of America 103 (2006) 10533–10535.
[2] N.M. Goodey, S.J. Benkovic, Allosteric regulation and catalysis emerge via a common route, Nature Chemical Biology 4 (2008) 474–482.
[3] A. Villaverde, Allosteric enzymes as biosensors for molecular diagnosis, FEBS Letters 554 (2003) 169–172.
[4] P.J. Conn, A. Christopoulos, C.W. Lindsley, Allosteric modulators of GPCRs: a novel approach for the treatment of CNS disorders, Nature Reviews Drug Discovery 8 (2009) 41–54.
[5] C. Bohr, Theoretische Behandlung der quantitativen Verhaltnisse bei der Sauerstoffaufnahme des Heamoglobins, Zentralbl. Physiol 17 (1903) 682–691.
[6] G.T. Cori, S.P. Colowick, C.F. Cori, The action of nucleotides in the disruptive phosphorylation of glycogen, Journal of Biological Chemistry 123 (1938) 381–389.
[7] J. Monod, J.P. Changeux, F. Jacob, Allosteric proteins and cellular control systems, Journal of Molecular Biology 6 (1963) 306–329.
[8] J. Monod, J. Wyman, J.P. Changeux, On the nature of allosteric transitions: a plausible model, Journal of Molecular Biology 12 (1965) 88–118.
[9] E.G. Krebs, E.H. Fischer, The phosphorylase b to a converting enzyme of rabbit skeletal muscle, Biochimica Biophysica Acta 20 (1956) 150–157.
[10] E.W. Sutherland Jr., W.D. Wosilait, Inactivation and activation of liver phosphorylase, Nature 175 (1955) 169–170.
[11] M.A. Bogoyevitch, D.P. Fairlie, A new paradigm for protein kinase inhibition: blocking phosphorylation without directly targeting ATP binding, Drug Discovery Today 12 (2007) 622–633.
[12] Z. Huang, L. Zhu, Y. Cao, G. Wu, X. Liu, Y. Chen, Q. Wang, T. Shi, Y. Zhao, Y. Wang, W. Li, Y. Li, H. Chen, G. Chen, J. Zhang, ASD: a comprehensive database of allosteric proteins and modulators, Nucleic Acids Research 39 (2011) 663–669.
[13] Pipeline Pilot, version 7.5, Accelrys, Inc., San Diego, USA, 2009.
[14] Molecular Operating Environment (MOE), version 2008, Chemical Computing Group, Inc., Montreal, Canada, 2008.
[15] Accelrys Available Chemicals Directory (ACD), Accelrys, Inc., San Diego, USA, 2005.
[16] Comprehensive Medicinal Chemistry (CMC), Accelrys, Inc., San Diego, USA, 2009.
[17] Chinese Natural Product Database (CNPD), NeoTrident Technology Ltd., Beijing, China, 2005.
[18] D.S. Wishart, C. Knox, A.C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, Z. Chang, J. Woolsey, DrugBank: a comprehensive resource for in silico drug discovery and exploration, Nucleic Acids Research 34 (2006) 668–672.
[19] D.S. Wishart, C. Knox, A.C. Guo, D. Cheng, S. Shrivastava, D. Tzur, B. Gautam, M. Hassanali, DrugBank: a knowledgebase for drugs, drug actions and drug targets, Nucleic Acids Research 36 (2008) 901–906.
[20] C. Knox, V. Law, T. Jewison, P. Liu, S. Ly, A. Frolkis, A. Pon, K. Banco, C. Mak, V. Neveu, Y. Djoumbou, R. Eisner, A.C. Guo, D.S. Wishart, DrugBank 3.0: a comprehensive resource for 'omics' research on drugs, Nucleic Acids Research 39 (2011) 1035–1041.
[21] MDDR, Accelrys, Inc., San Diego, USA, 2009.
[22] NCI Open Database, National Cancer, Ins., Bethesda, USA, 2003.
[23] Dragon, version 5.2, TALETE, srl., Milano, Italy, 2004.
[24] M. Feher, J.M. Schmidt, Property distributions: differences between drugs, natural products, and molecules from combinatorial chemistry, Journal of Chemical Information and Computer Sciences 43 (2003) 218–227.
[25] C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, Experimental computational approaches to estimate solubility and permeability in drug discovery and development settings, Advanced Drug Delivery Reviews 23 (1997) 3–25.
[26] N.M. O'Boyle, M. Banck, C.A. James, C. Morley, T. Vandermeersch, G.R. Hutchison Open Babel, An open chemical toolbox, Journal of Chemical Information 3 (2011) 33.
[27] Origin, version 8.0, OriginLab, Corporation, Northampton, USA, 2007.
[28] L.H. Pinto, L.J. Holsinger, R.A. Lamb, Influenza virus M2 protein has ion channel activity, Cell 69 (1992) 517–528.
[29] C. Wang, K. Takeuchi, L.H. Pinto, R.A. Lamb, Ion channel activity of influenza A virus M2 protein: characterization of the amantadine block, Journal of Virology 67 (1993) 5585–5594.
[30] R.A. Lamb, K.J. Holsinger, L.H. Pinto, in: E. Wemmer (Ed.), Cellular Receptors of Animal Viruses, Cold Spring Harbor Press, New York, 1994, pp. 303–321.
[31] M. Beato, M. Truss, S. Chavez, Control of transcription by steroid hormones, Annals of the New York Academy of Sciences 784 (1996) 93–123.
[32] D.B. Longley, D.P. Harkin, P.G. Johnston, 5-Fluorouracil: mechanisms of action and clinical strategies, Nature Reviews Cancer 3 (2003) 330–338.
[33] Y. Yang, H. Chen, I. Nilsson, S. Muresan, O. Engkvist, Investigation of the relationship between topology and selectivity for druglike molecules, Journal of Medicinal Chemistry 53 (2010) 7709–7714.
[34] G. Chen, S. Zheng, X. Luo, J. Shen, W. Zhu, H. Liu, C. Gui, J. Zhang, M. Zheng, C.M. Puah, K. Chen, H. Jiang, Focused combinatorial library design based on structural diversity, druglikeness and binding affinity score, Journal of Combinatorial Chemistry 7 (2005) 398–406.
[35] Y. Fukunishi, M. Lintuluoto, Development of chemical compound libraries for in silico drug screening, Current Computer-Aided Drug Design 6 (2010) 90–102.
[36] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, The WEKA data mining software: an update, SIGKDD Explorations 11 (2009) 10–18.