Supplementary Data

AlloFinder: a strategy for allosteric modulator discovery and allosterome analyses

Min Huang[1,‖], Kun Song[1,‖], Xinyi Liu[1,‖], Shaoyong Lu[1,‖], Qiancheng Shen[1,2], Renxiao Wang[3], Jingze Gao[2], Yuanyuan Hong[2], Qian Li[2], Duan Ni[1], Guoqiang Chen[1], Jianrong Xu[4], Jian Zhang[1,2,*]

[1] Department of Pathophysiology, Key Laboratory of Cell Differentiation and Apoptosis of Chinese Ministry of Education, Shanghai Jiao-Tong University School of Medicine (SJTU-SM), Shanghai, 200025, China.

[2] Medicinal Bioinformatics Center, Shanghai Jiao-Tong University School of Medicine (SJTU-SM), Shanghai, 200025, China.

[3] Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, Shanghai 200032, China

[4] Department of Pharmacology, Shanghai Jiao-Tong University School of Medicine (SJTU-SM), Shanghai, 200025, China.

Key Words: AlloFinder, allosteric modulator, allosteric mechanism, allosteric metabolites

[‖]The authors wish it to be known that, in their opinion, the first four authors should be regarded as joint First Authors

***To whom correspondence should be addressed:**
Dr. Jian Zhang, PhD
Tel: +86-21-63846590-776922
Fax: +86-21-64154900
E-mail: jian.zhang@sjtu.edu.cn

Materials and Methods

Ligand library construction

Ligand libraries used for virtual screening are provided, particularly for the availability of chemical compounds from commercial sources, including ChEMBL, SPECS, NCI Open Chemical Repository, and for-sale subset from ZINC. These pre-built ligand libraries were generated by the following procedures. First, both drug-like and allo-like filters were performed to rule out molecules with undesired physicochemical and allosteric properties. The drug-like parameters are used as follows: H-bond donors < 5, H-bond acceptors < 10, molecular weight < 650, LogP < 5, and rotatable bonds ≤ 7. The allo-like parameter is number of rings in the largest ring system = 1 or 2. In addition to the four libraries, Endogenous Ligands from ZINC are also provided without the filters. To guarantee chemical diversity, the molecules for each library were then clustered using the K-medoids algorithm and the Morgan fingerprint with a maximum distance of Tanimoto coefficient of 0.3 in the clusters. The cluster centroids were retained to constitute the diverse ligand library. Afterwards, to obtain the preferred conformation for each compound, the RDKit program and the UFF force field were used to generate up to 50 three-dimensional (3D) conformers for each compound. The conformer with the minimum energy for each compound was put into the ligand library. Finally, the Endogenous Ligands, ChEMBL Diversity, ZINC Diversity, SPECS Diversity, and NCI Diversity library included 10512, 40398, 48713, 10070, and 20302 compounds, respectively.

Ligand library selection for users

ZINC Diversity library contains 48713 diverse compounds, which was constructed from the “for-sale” subset of ZINC (version 15)[1]. All compounds in the library are for sale, a quarter of which are available for immediate delivery. In addition, the compounds in the library are annotated with chemical properties and available commercial vendors. Thus it is very convenient for users to obtain any compound to perform experimental validation when using ZINC Diversity for allosteric modulator screening.

Endogenous Ligands library has 10512 diverse endogenous metabolites, which was built from the “endogenous” subset of ZINC (version 15)[1]. The library focuses on the curation of primary metabolites observed in humans. Characterizing the potential

allosteric metabolite-protein interactions by screening the library may thus deepen our understanding of the molecular underpinnings for metabolites in feedback modulation of pathways and has profound importance in human health as a consequence of the strong interplay between metabolites and the pathogenesis of various diseases.

ChEMBL Diversity library contains 40398 diverse compounds, which was derived from ChEMBL (version 23)[2], an open large-scale bioactivity database. Compounds in the library were manually curated from the medicinal chemistry literature. As the result, the biological and pharmaceutical annotations of the compounds are rather rich and full of believable, such as structure-activity relationship, selectivity, and ADMET information. Users may unravel novel allosteric mechanism for these bioactive compounds when using the ChEMBL Diversity library for screening in the AlloFinder server.

SPECS Diversity library has 10070 diverse compounds, which was created from SPECS in-house compound database. SPECS (http://www.specs.net) is world's leading provider of compound management services since 1987. Their chemically diverse in-house collection consists of single synthesized, well-characterized and drug-like small molecules. Users can find novel structures using the library during the allosteric modulator screening.

NCI Diversity library contains 20302 diverse compounds from the Open National Cancer Institute (NCI) database[3]. Compounds in the library have been evaluated as potential anticancer and anti-HIV agents. The library is recommended for users to screen allosteric modulators on cancer-related and HIV-related targets.

Allosteric site preparation

Allosteric modulator discovery with the AlloFinder server employs the Allosite algorithm to generate all potential allosteric sites using feature-based regression combined with NMA-based perturbation[4,5]. We built the algorithm in the AlloFinder server and a parameter of ten thousand iterations in pocket detection is used to obtain more accurate pockets throughout a query protein. In addition, other parameters are set by default. Redundant pockets are removed. The scores of both feature-based logistic regression $P_{feature}$ and NMA-based perturbation P_{NMA} are calculated on each pocket, and potential allosteric sites are outputted with $P_{Allosite}$ >0.2 in the server:

$$P_{Allosite} = 0.8 * P_{feature} + 0.2 * P_{NMA}$$

Users submit the structure of proteins either with a PDB ID or by uploading PDB-formatted coordinate file. The size of the input structures should not exceed 2000 amino acids, considering normal mode analysis is time-consuming in large systems. The running time is depended on the size and the complexity of the query protein. Normally, a representative procedure of allosteric site preparation for a 500-residues protein will take 20 seconds. After all the computations are completed, a group of potential allosteric sites are well prepared and shown. Users may select one favorable site to start allosteric modulator discovery in the AlloFinder server.

Ligands preparation

To accelerate the large-scale screening of allosteric modulators at a selected allosteric site, we used a pocket-generated pharmacophore model automatically extracted from the allosteric site as a filter to quickly rule out compounds in a library that have no key pharmacophore groups for binding to the site. The pocket-based pharmacophore algorithm was from LigBuilder software developed by us[6] and transplanted into the AlloFinder server. Compounds in the library filtered through the pharmacophore model (Feature Score of pharmacophore > 0.5) are prepared for allosteric modulator discovery in the AlloFinder server.

Allosteric modulator screening of AlloFinder

An efficient two-step protocol for allosteric modulators screening was designed by the combination of conformational sampling and allosteric scoring procedures in the AlloFinder server. The protocol used the Python programming language (http://www.python.org) to bind two separate steps together into an automatic workflow.

Conformational sampling procedure. The AutoDock Vina algorithm (v1.1.2) was built in the AlloFinder server to generate an ensemble of diverse sampling conformations for each ligand at a predicted allosteric site of query protein as before[7] and the procedure is briefly described below. Each genetic algorithm (GA) run outputs 10 docked conformations as the result. Since a conformational ensemble is desired, 10 individual GA runs are performed to generate 100 docked conformations for each ligand. The parameters used for each GA are exhaustiveness arg = 7 and energy_range arg = 3. During sampling, all the rotatable single bonds in the ligand are allowed to rotate except those whose rotations do not result in different conformations, such as the ones connecting a terminal -CH3 group. Flexibility in cyclic parts of the

ligand is neglected. The size of the docking box is 30 Å × 30 Å × 30 Å, which is centered at the predicted allosteric site. Initial conformations of the ligand are generated randomly in the box. Other miscellaneous parameters are assigned the default values given by the AutoDock Vina algorithm. The protein structure is kept fixed during the sampling.

Allosteric scoring procedure. The conformational ensemble of each ligand generated from the conformational sampling procedure is directly evaluated for binding affinity of the ligand at the site by the Alloscore algorithm (built in the AlloFinder server), which is specifically designed for accurate allosteric modulator-protein interaction evaluation and outperforms other methods in the prediction of allosteric modulator-protein binding affinities[8]. The binding affinity of a given allosteric modulator-protein complex, as expressed in logarithm units ($\log K_a$), is calculated by summing all of six energy terms, with the following resulting functions:

$$pK_d = c_0 + c_1 M_{VDW} + c_2 M_{H-bond} + c_3 M_{RT} + c_4 M_{hydrohobic} + c_5 M_{\Delta \mathrm{VOL}} + c_6 M_{\Delta \mathrm{PSAS}}$$

where M_{VDW}, $M_{H\text{-}bond}$ M_{RT}, $M_{hydrohobic}$, $M_{\Delta VOL}$, and $M_{\Delta PSAS}$ represent the binding energy terms from VDW interaction, hydrogen bond, rotatable bond, hydrophobic interaction, buried molecular volume, and buried polar surface area, respectively. The weight parameters of the Alloscore for M_{VDW}, $M_{H\text{-}bond}$ M_{RT}, $M_{hydrohobic}$, $M_{\Delta VOL}$, and $M_{\Delta PSAS}$ used in the AlloFinder platform are set by default. A conformation with the best binding affinity score (or the lowest binding affinity) from the ensemble is chosen as the final "Alloscore Score" for each ligand. All ligands of a library in screening are ranked by the score in descending order and the top 100 docked ligands are outputted in the AlloFinder server.

Allosterome mapping of AlloFinder

Allosterome analysis contains two items, mapping of the predicted allosteric sites and modulators to the known allosteric sites and modulators, respectively.

Allosteric site mapping. The 3D structure of each predicted allosteric site is aligned to 636 known human allosteric sites by using the APOC algorithm[9]. The parameter "-fa" is set to 0 to perform the alignment of pocket structure only, and the normalized Pocket Similarity score (PS-score) between the predicted allosteric site and each known allosteric site is calculated for ranking. The 20 known allosteric sites that are

most similar with the predicted allosteric site are outputted. Besides, a structural homologous analysis of the predicted allosteric site with human known allosteric sites is performed to reveal evolutionary neighbors. The standard APE algorithm[10] in the AlloFinder server is used for generating phylogenetic tree.

Allosteric modulator mapping. The 2D structure of each predicted allosteric modulator is searched with the 263 endogenous and 70390 exogenous allosteric modulator libraries, respectively, using the MACCS Fingerprint of Open babel Fastsearch algorithm[11]. For similarity searching, 50 allosteric modulators in each library that are most similar to the predicted allosteric modulator are produced. The physicochemical properties of modulators such as TPSA, logP, and MW are calculated by the Pybel method[12]. In addition, the server also aligns the 2D structure of each predicted allosteric modulator with the 17323 allosteric metabolites engaged in pathways/networks annotated in AlloFinder and KEGG database[13] by using the same searching method. The pathways and networks of the predicted allosteric metabolite or its most similar metabolite are provided, revealing the potential metabolic mechanism engaged in by the modulator.

Protein expression and purification

E. coli cells were grown at 37°C and induced at 25°C with 0.5 mM IPTG. Ni Sepharose 6 Fast Flow (Qiagen) was used to purify STAT3. Lysis buffer: 20 mM Tris, 0.3 M NaCl, 10% glycerol, 20 mM imidazole, β-ME, pH 8.0. Wash buffer: 20 mM Tris, 0.1 M NaCl, 50 mM imidazole, β-ME, pH 8.0. Elution buffer: 20 mM Tris, 0.1 M NaCl, 400 mM imidazole, β-ME, pH 8.0. Gel filtration chromatography was used to purify STAT3 by AKTA explorer. We used PD-10 columns (GE Healthcare) to change the protein buffer to 20 mM HEPES, 0.1 M NaCl, pH 7.5. STAT3 CCD is the construction of STAT3 amino acid from 127 to 320. Mutants of STAT3 CCD (N175G/K177G, Q212A/M213G) and mutants of STAT3 (R595G, K591A/S611G, I634S/Q635G, K591A) are used the same purify assay.

Fluorescence polarization assay

The detailed fluorescence polarization assay was described previously[14]. We used the fluorescent probe peptide Ac-pYLPQTV-NH_2 (protected from light) as a negative control and the non-phosphorylated Ac-pYLPQTV-NH_2 peptide as a positive control.

The small compounds were purchased from SPECS library. Varying concentrations of small molecular compounds were prepared in black 96-well round-bottom plates (Corning 3560), and STAT3 protein (0.5 μM) was incubated with these compounds for 1.5 h on a shaker. Next, the fluorescent tracer was added to the sample wells at 1 nM. The assay buffer contained 10 mM HEPES, 50 mM NaCl, 1 mM EDTA, 0.1% NP-40, and 2 mM DTT. An additional 1.5 h was used for smaller compounds and probes to combine with STAT3 protein. The fluorescence polarization value was measured on a BioTek Plate reader with excitation at 485 nm and emission at 520 nm. Data were processed using GraphPad Prism 5.

Surface plasmon resonance (SPR)

SPR experiments were performed at 25 °C using a Biacore T200 instrument. First, the baseline RU of the chip, when bathed in PBS buffer, was determined. Second, K116 were injected into the flow cell and flowed for 90 s. The association phase showed increasing resonance units (RUs) during the flow of K116. Finally, the flow of K116 was terminated, and PBS buffer flow was initiated for 90 s. The concentration of K116 was tested in twofold dilutions from 25μM to 0.39μM. This is known as the dissociation phase. Subsequently, we obtained the exact binding affinity of K116 with STAT3 CCD and mutants. The kinetic parameter K_d for each interaction was determined by globally fitting the experimental data with the Biacore T200 Evaluation Software 4.1 (GE Healthcare).

Cell growth assay

MDA-MB-468 and DU145 cells were plated at equal confluence in 96-well plates and were left either untreated or treated of K116 with concentrations ranging from 1 mM to 0.18 mM continually during the indicated time for 24 h. Cells from individual wells were supplied with 10 μl of CCK8 for 4 h at 37°C. Absorption was measured on a BioTek Plate reader at 450 nm. Data were processed using GraphPad Prism 5. Graphs represent the total cell numbers at each time interval, and the error bars represent the s.e.m. Half-maximal inhibitory concentrations (IC_{50}) were calculated using GraphPad Prism 5.

Cell apoptosis assay

MDA-MB-468 cells were seeded in 6-well plates at a density of 5×10^5 cells per well and grown with L15 at 37 °C in a humidified incubator without 5% CO_2. Cells were

treated with K116 (1 μM, 10 μM, 20 μM, 30 μM, and 50 μM) or 0.1% DMSO for 24 h. Cell morphology was visualized using light microscopy.

Immunoblotting assay

Protein samples were separated using SDS-PAGE and transferred onto a NC membrane. After blockade with 5% nonfat milk, the proteins were analyzed by standard procedures using antibodies to STAT3 (Cell Signaling Technology), phospho-STAT3 (pY705STAT3, Cell Signaling Technology), Src (Cell Signaling Technology), phospho-Src (p-Src, Cell Signaling Technology), Cyclin D1, the anti-apoptotic gene Bcl-xL, and C-Myc. Alternatively, secondary antibodies of peroxidase-labeled mouse-specific or rabbit HRP-specific antibodies were detected.

Figures

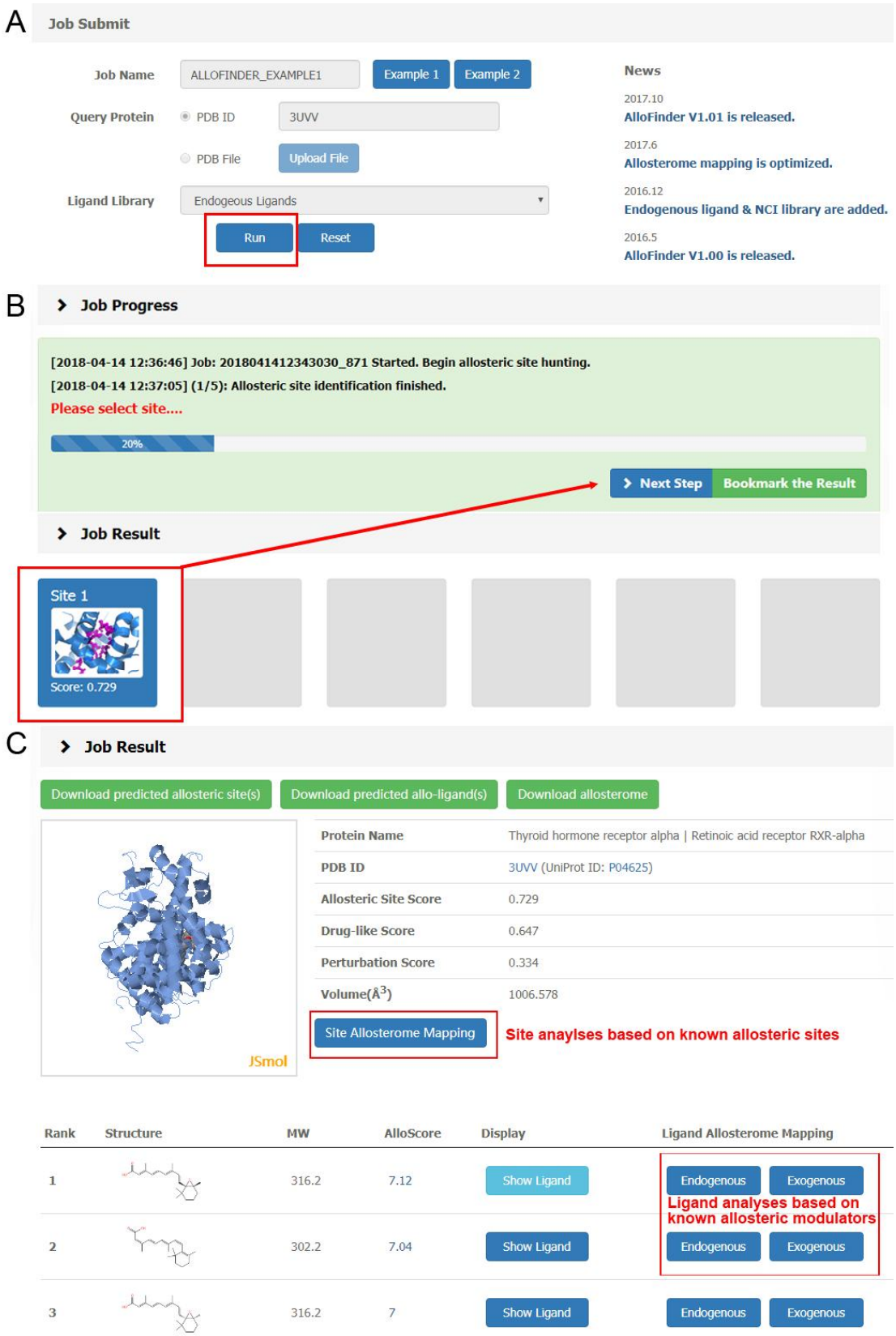


Figure S1. Screenshots of the procedure from Example 1 application of the AlloFinder webserver to the discovery of potential allosteric metabolites of TR:RXR

heterodimers (PDB ID: 3UVV)[15]. (A) Input parameters for the screening. (B) Selection of predicted allosteric site "Site 1" for the screening. (C) Job result of the screening as output.

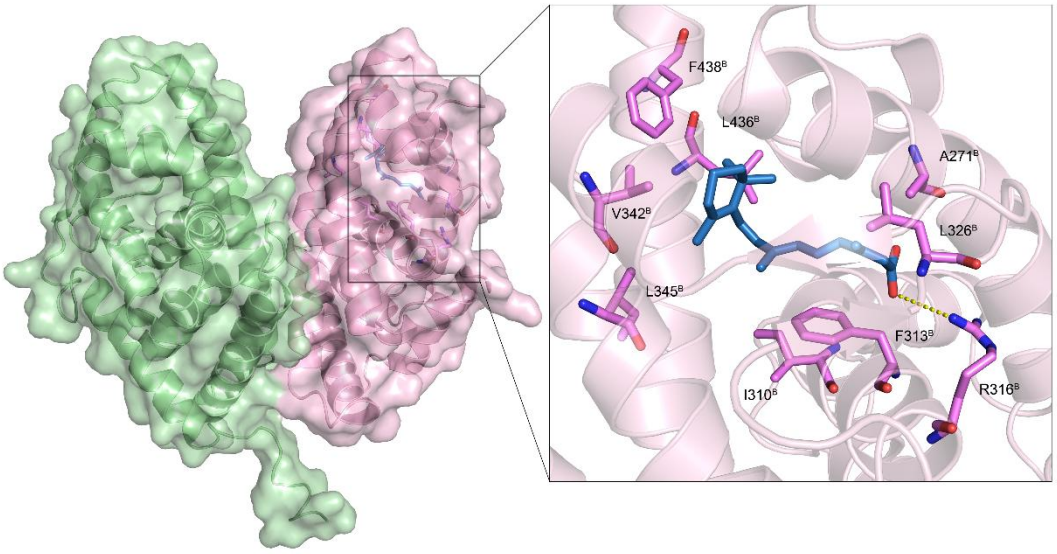


Figure S2. The binding result of 9-cis retinoic acid at the predicted allosteric site in Example 1. Left: The TR (palegreen): RXR (lightpink) heterodimer is represented as surface and cartoon, with the docked 9-cis retinoic acid and site residues shown as stick. Right: The zoom-in version of the binding site of 9-cis retinoic acid. As shown at the site, the 2,6,6-trimethylcyclohexen group of 9-cis retinoic acid is embedded into a hydrophobic subpocket formed by $F438^{B}$, $V342^{B}$, $L345^{B}$, $L436^{B}$, and the 3,7-dimethyl group makes hydrophobic contacts with residues $A271^{B}$, $L326^{B}$, $I310^{B}$ and $F313^{B}$. More importantly, the terminal carboxyl group forms a salt bridge with the side chain of $R316^{B}$. All of the atomic interactions together contribute to the high-ranked binding affinity of 9-cis retinoic acid at the predicted allosteric site of TR: RXR.

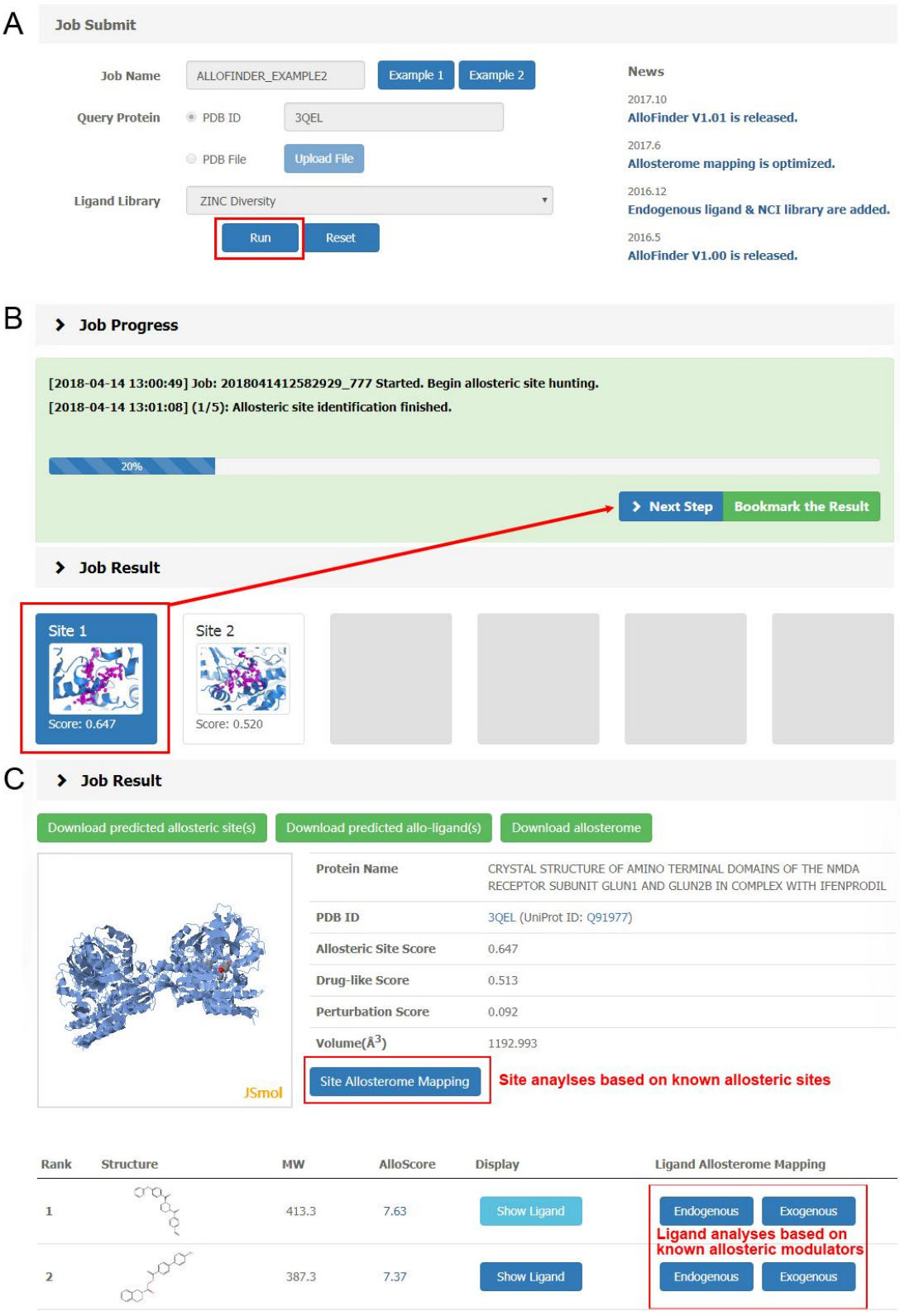


Figure S3. Screenshots of the procedure from Example 2 application of the AlloFinder webserver to the discovery of potential allosteric modulators of GluN1b:GluN2B NMDA receptors (PDB ID: 3QEL)[16]. (A) Input parameters for the screening. (B) Selection of predicted allosteric site "Site 1" for the screening. (C) Job

result of the screening as output.

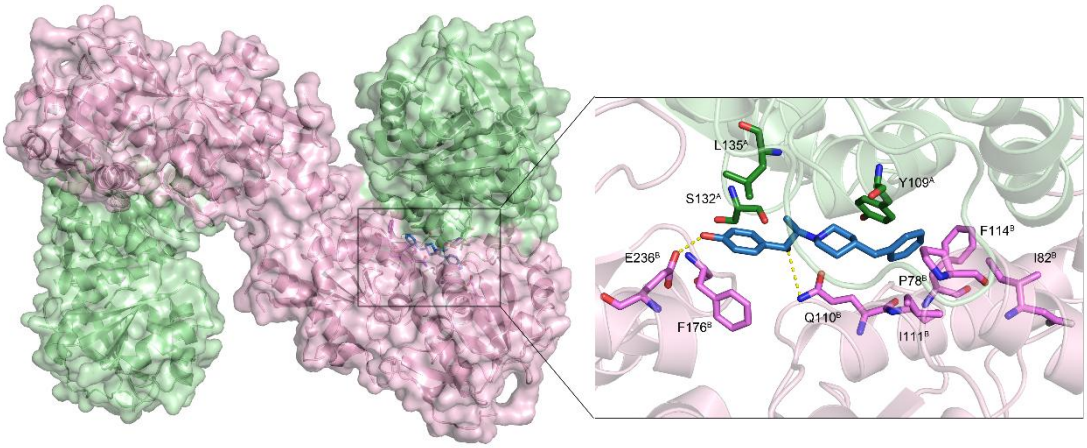


Figure S4. The binding result of ifenprodil at the predicted allosteric site in Example 2. Left: The GluN1b (palegreen): GluN2B (lightpink) NMDA receptor is represented as surface and cartoon, with the docked ifenprodil and pocket residues shown as stick. Right: The zoom-in version of the binding site of ifenprodil. According to the result of AlloFinder, two hydrogen bonds are formed between ifenprodil and the side chains of $E236^{B}$ and $Q110^{B}$. In addition, the benzyl group of ifenprodil forms hydrophobic interactions with $L135^{A}$ and $F176^{B}$, and the terminal benzene group creates hydrophobic interactions with $Y109^{A}$, $F114^{B}$, $I82^{B}$, $I111^{B}$ and $P78^{B}$. The rich interactions between ifenprodil and GluN1b: GluN2B explain the better binding affinity at the top 2 of ranked benchmark dataset.

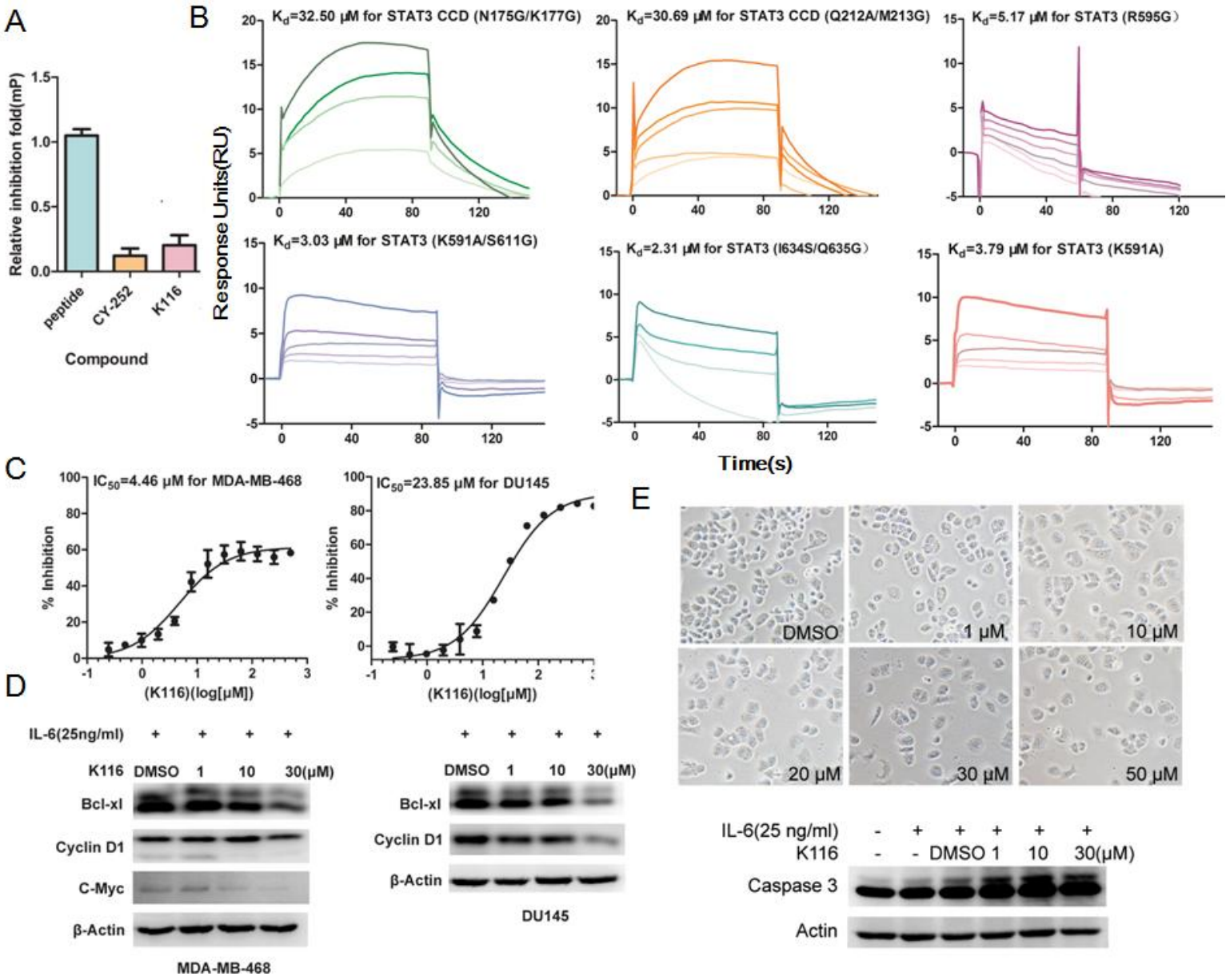


Figure S5. Biological evaluation of K116 against STAT3. (A) Identification of the inhibition of K116 by FP. Ac-pYLPQTV-NH2 as a negative control (peptide). Positive control (CY-252) is non-phosphorylated Ac-YLPQTV-NH2 peptide. (B) The binding affinity of K116 with mutants (N175G/K177G, Q212A/M213G, R595G, K591A/S611G, I634S/Q635G, K591A) of STAT3 by SPR. (C) K116 inhibits the growth of MDA-MB-468 and DU145 cells. (D) Down-regulation of STAT3 target genes (CyclinD1, Bcl-xL, and C-Myc) in MDA-MB-468 and DU145 cells treated with K116 for 24 h. (E) K116 induced MDA-MB-468 cell apoptosis. Cell morphology features were visualized using light microscopy, and increasing Caspase-3 levels were detected after treatment with DMSO or 1 μM, 10 μM, 20 μM, 30 μM K116 for 24 h. Quantitative analysis of cells from three independent experiments was performed.

Tables

Table S1. Enrichment factors (EFs) of known allosteric modulators for 60 proteins in the benchmark by AlloFinder.

Protein class	Protein(site)[a]	PDB ID	Resolution (Å)	Number of actives	EF1[b]	EF5	EF20
Kinase	ABL1	3PYY	1.9	34	20.6	8.2	3.1
	AKT1	3O96	2.7	34	29.4	17.1	4.7
	Chk1	3F9N	1.9	35	28.6	18.9	4.9
	CDK2	3PXF	1.8	32	28.1	15.0	4.7
	FAK	4EBW	2.7	5	100.0	20.0	5.0
	GK	3H1V	2.1	26	30.8	10.0	3.3
	IGF-1R	3LW0	1.8	18	38.9	15.6	4.4
	JAK2	5UT6	1.7	1	0.0	20.0	5.0
	LIMK2	4TPT	2.6	34	29.4	19.4	5.0
	MAPK7	4ZSJ	2.5	1	100.0	20.0	5.0
	MAPK8	3O2M	2.7	41	22.0	9.8	3.8
	PAK1	4ZLO	2.5	3	66.7	20.0	5.0
	WNK1	5WDY	2.5	3	66.7	13.3	5.0
Protease	CASP6	4NBN	1.8	6	33.3	10.0	3.3
	ClpP	5DL1	3.0	1	0.0	20.0	5.0
	CTSK	4LEG	2.2	1	0.0	0.0	5.0
	IN	4NYF	1.9	48	18.8	8.8	3.6
	KSHV_Pr	3NJQ	2.0	1	100.0	20.0	5.0
	MALT1	4I1R	2.7	1	0.0	20.0	5.0
Transferase	Bcy1	3OF1	2.2	30	33.3	15.3	4.0
	DAHPS	1KFL	2.8	4	0.0	10.0	2.5
	Ep	1MP3	2.2	21	23.8	8.6	3.6
	FPPS	3N1V	2.2	50	20.0	12.4	3.5
	GP	3CEH	2.8	1	100.0	20.0	5.0
	UAP	4BQH	1.8	10	60.0	16.0	4.0
	UPPs	4Q9M	2.1	1	100.0	20.0	5.0
Hydrolase	BRR2	5URK	3.0	1	100.0	20.0	5.0
	FBP1	2JJK	2.0	18	33.3	8.9	3.6
	FBP2	3IFA	1.9	16	12.5	8.8	4.1
	GLS	3UO9	2.3	39	17.9	8.7	3.8
	SHP2	5EHP	1.9	2	100.0	20.0	5.0
	Sirt1	5BTR	3.2	1	100.0	20.0	5.0
	Sirt3	4C7B	2.1	5	40.0	16.0	4.0
Oxidoreductase	IDH1	5DE1	2.3	1	100.0	20.0	5.0
	IDH2	4JA8	1.6	11	63.6	18.2	4.5
	NAD-IDH	5GRE	2.7	1	100.0	20.0	5.0
	RNR	1XJF	2.4	7	0.0	5.7	2.9

Isomerase	MurI(site 1)	2W4I	1.9	1	100.0	20.0	5.0
	MurI(site 2)	4B1F	2.1	1	100.0	20.0	5.0
	Ugd	3PJG	2.7	3	33.3	20.0	5.0
GPCR	β2AR	5X7D	2.7	1	100.0	20.0	5.0
	CCR5	4MBS	2.7	24	41.7	18.3	5.0
	mGlu5	4OO9	2.6	1	0.0	20.0	5.0
Ligase	ACC	1W96	1.8	32	25.0	8.1	2.7
	ProRS	4WI1	1.7	27	25.9	14.1	4.8
Nuclear receptor	RORγt	5C4T	1.8	4	100.0	20.0	5.0
	TR	3UVV	3.0	1	100.0	20.0	5.0
Channel	nAChR	5AFJ	2.2	1	100.0	20.0	5.0
	GluN2B	3QEL	2.6	18	16.7	10.0	3.9
Cell cycle	Cdc4	3MKS	2.6	1	100.0	20.0	5.0
	KSP	4BBG	2.8	1	100.0	20.0	5.0
Transport	HAS	3LU6	2.7	1	100.0	20.0	5.0
	HBA	2D60	1.7	1	100.0	20.0	5.0
Motor protein	Myo(site 1)	2JHR	2.8	1	100.0	20.0	5.0
	Myo(site 2)	3BZ7	2.0	1	100.0	20.0	5.0
Transcription Factor	TetR	2VPR	2.5	1	100.0	20.0	5.0
Lyase	AC	4OYA	2.0	9	22.2	8.9	3.9
Cell adhesion	LFA-1	3M6F	1.9	28	35.7	20.0	5.0
Other Proteins	NS5B	4NLD	2.8	3	66.7	20.0	5.0
	TcdA	3HO6	1.6	2	50.0	20.0	5.0

[a]Abbreviations: ABL1, Tyrosine-protein kinase ABL1; AKT1, RAC-alpha serine/threonine-protein kinase; Chk1, Serine/threonine-protein kinase Chk1; CDK2, Cyclin-dependent kinase 2; FAK, Focal adhesion kinase 1; GK, Glucokinase; IGF-1R, Insulin-like growth factor 1 receptor; JAK2, Tyrosine-protein kinase JAK2; LIMK2, LIM domain kinase 2; MAPK7, Mitogen-activated protein kinase 7; MAPK8, Mitogen-activated protein kinase 8; PAK1, Serine/threonine-protein kinase PAK 1; WNK1, Serine/threonine-protein kinase WNK1; CASP6, Caspase-6; ClpP, ATP-dependent Clp protease proteolytic subunit; CTSK, Cathepsin K; IN, HIV-1 integrase; KSHV_Pr, Capsid scaffolding protein; MALT1, Mucosa-associated lymphoid tissue lymphoma translocation protein 1; Bcy1, cAMP-dependent protein kinase regulatory subunit; DAHPS, Phospho-2-dehydro-3-deoxyheptonate aldolase, Phe-sensitive; Ep, Glucose-1-phosphate thymidylyltransferase; FPPS, Farnesyl pyrophosphate synthase; GP, Glycogen phosphorylase, liver form; UAP, UDP-N-acetylglucosamine pyrophosphorylase; UPPs, Isoprenyl transferase; BRR2, U5 small nuclear ribonucleoprotein 200 kDa helicase; FBP1, Fructose-1,6-bisphosphatase 1; FBP2, Fructose-1,6-bisphosphatase isozyme 2; GLS, Glutaminase kidney isoform, mitochondrial; SHP2, Tyrosine-protein phosphatase non-receptor type 11; Sirt1, NAD-dependent protein deacetylase sirtuin-1; Sirt3, NAD-dependent protein deacetylase sirtuin-3, mitochondrial; IDH1, Isocitrate dehydrogenase [NADP] cytoplasmic; IDH2, Isocitrate dehydrogenase [NADP], mitochondrial; NAD-IDH, Isocitrate dehydrogenase [NAD] subunit alpha, mitochondrial; RNR, Ribonucleotide reductase; MurI, Glutamate racemase; Ugd, UDP-glucose 6-dehydrogenase; ACC, Acetyl-CoA carboxylase; ProRS, Proline--tRNA ligase; AC, Adenylate cyclase type 10; β2AR, Beta-2 adrenergic receptor; CCR5, C-C chemokine receptor type 5; mGlu5, Metabotropic glutamate receptor; RORγt, Nuclear receptor ROR-gamma; TR, Thyroid hormone receptor alpha; GluN2B, Glutamate receptor ionotropic, NMDA 2B; nAChR, Neuronal acetylcholine receptor subunit alpha-7; Cdc4, Cell division control protein 4; KSP, Kinesin-like protein KIF11; HAS, Serum albumin; HBA, Hemoglobin subunit beta; Myo, Myosin-2 heavy chain; TetR, Tetracycline repressor protein class H; LFA-1, Integrin alpha-L; NS5B, HCV NS5B polymerase; TcdA, Toxin A.

[b]EF(n) is the concentration of the known ligands among n% top-scoring hits compared to their concentration throughout the entire library.

Supplementary Table S2: Hit Rate (HR) of known allosteric modulators for 60 proteins in the benchmark by AlloFinder.[a]

Protein class	Protein(site)	PDB ID	Resolution (Å)	Number of actives	HR[b]
Kinase	ABL1	3PYY	1.9	34	0.62
	AKT1	3O96	2.7	34	0.94
	Chk1	3F9N	1.9	35	0.97
	CDK2	3PXF	1.8	32	0.94
	FAK	4EBW	2.7	5	1.00
	GK	3H1V	2.1	26	0.65
	IGF-1R	3LW0	1.8	18	0.89
	JAK2	5UT6	1.7	1	1.00
	LIMK2	4TPT	2.6	34	1.00
	MAPK7	4ZSJ	2.5	1	1.00
	MAPK8	3O2M	2.7	41	0.76
	PAK1	4ZLO	2.5	3	1.00
	WNK1	5WDY	2.5	3	1.00
Protease	CASP6	4NBN	1.8	6	0.67
	ClpP	5DL1	3.0	1	1.00
	CTSK	4LEG	2.2	1	1.00
	IN	4NYF	1.9	48	0.73
	KSHV_Pr	3NJQ	2.0	1	1.00
	MALT1	4I1R	2.7	1	1.00
Transferase	Bcy1	3OF1	2.2	30	0.80
	DAHPS	1KFL	2.8	4	0.50
	Ep	1MP3	2.2	21	0.71
	FPPS	3N1V	2.2	50	0.70
	GP	3CEH	2.8	1	1.00
	UAP	4BQH	1.8	10	0.80
	UPPs	4Q9M	2.1	1	1.00
Hydrolase	BRR2	5URK	3.0	1	1.00
	FBP1	2JJK	2.0	18	0.72
	FBP2	3IFA	1.9	16	0.81
	GLS	3UO9	2.3	39	0.77
	SHP2	5EHP	1.9	2	1.00
	Sirt1	5BTR	3.2	1	1.00
	Sirt3	4C7B	2.1	5	0.80
Oxidoreductase	IDH1	5DE1	2.3	1	1.00
	IDH2	4JA8	1.6	11	0.91
	NAD-IDH	5GRE	2.7	1	1.00
	RNR	1XJF	2.4	7	0.57
Isomerase	MurI(site 1)	2W4I	1.9	1	1.00
	MurI(site 2)	4B1F	2.1	1	1.00
	Ugd	3PJG	2.7	3	1.00
GPCR	β2AR	5X7D	2.7	1	1.00
	CCR5	4MBS	2.7	24	1.00
	mGlu5	4OO9	2.6	1	1.00
Ligase	ACC	1W96	1.8	32	0.53
	ProRS	4WI1	1.7	27	0.96
Nuclear receptor	RORγt	5C4T	1.8	4	1.00
	TR	3UVV	3.0	1	1.00
Channel	nAChR	5AFJ	2.2	1	1.00
	GluN2B	3QEL	2.6	18	0.78
Cell cycle	Cdc4	3MKS	2.6	1	1.00
	KSP	4BBG	2.8	1	1.00

Transport	HAS	3LU6	2.7	1	1.00
	HBA	2D60	1.7	1	1.00
Motor protein	Myo(site 1)	2JHR	2.8	1	1.00
	Myo(site 2)	3BZ7	2.0	1	1.00
Transcription Factor	TetR	2VPR	2.5	1	1.00
Lyase	AC	4OYA	2.0	9	0.78
Cell adhesion	LFA-1	3M6F	1.9	28	1.00
Other Proteins	NS5B	4NLD	2.8	3	1.00
	TcdA	3HO6	1.6	2	1.00

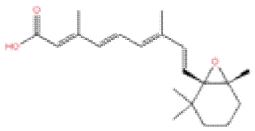
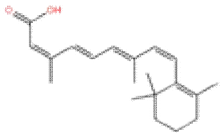
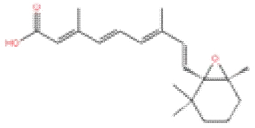
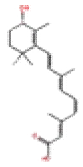
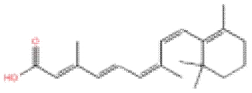
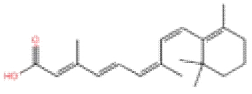
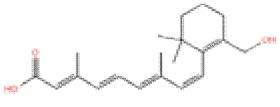
[a]Abbreviations of proteins are the same as those in Table S1.
[b]HR value represents the ratio of the EF20 (in Table S1) to ideal EF20 (here it is the number of known allosteric modulators) of the ranked benchmark dataset as described before[17].

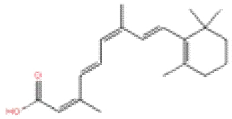
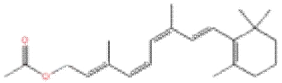
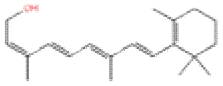
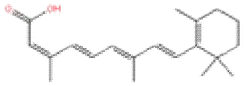
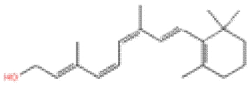
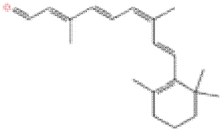
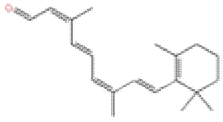
Supplementary Table S3. Area Under Curve (AUC) values on proteins with ≥10 known allosteric modulators in the benchmark by AlloFinder.[a]

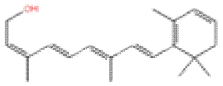
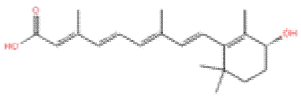
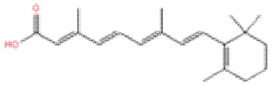
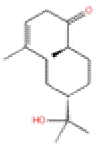
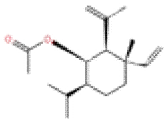
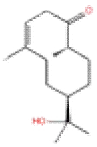
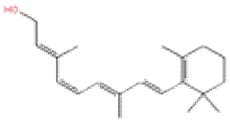
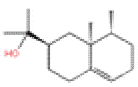
Protein class	Protein	PDB ID	Resolution (Å)	Number of actives	AUC values[b]
Kinase	ABL1	3PYY	1.9	34	0.839
	AKT1	3O96	2.7	34	0.979
	Chk1	3F9N	1.9	35	0.987
	CDK2	3PXF	1.8	32	0.958
	GK	3H1V	2.1	26	0.773
	IGF-1R	3LW0	1.8	18	0.946
	LIMK2	4TPT	2.6	34	0.999
	MAPK8	3O2M	2.7	41	0.853
Protease	IN	4NYF	1.9	48	0.847
Transferase	Bcy1	3OF1	2.2	30	0.910
	Ep	1MP3	2.2	21	0.866
	FPPS	3N1V	2.2	50	0.869
	UAP	4BQH	1.8	10	0.949
Hydrolase	FBP1	2JJK	2.0	18	0.850
	FBP2	3IFA	1.9	16	0.900
	GLS	3UO9	2.3	39	0.877
Oxidoreductase	IDH2	4JA8	1.6	11	0.937
GPCR	CCR5	4MBS	2.7	24	0.992
Ligase	ACC	1W96	1.8	32	0.706
	ProRS	4WI1	1.7	27	0.957
Channel	GluN2B	3QEL	2.6	18	0.889
Cell adhesion	LFA-1	3M6F	1.9	28	0.999

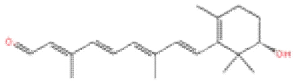
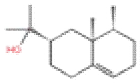
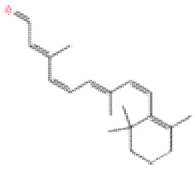
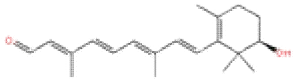
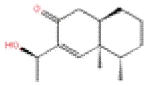
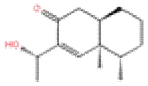
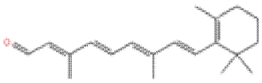
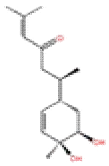
[a]Abbreviations of proteins are the same as those in Table S1.
[b]AUC value is the area under an ROC curve for the screening.

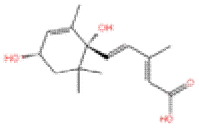
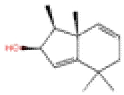
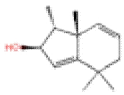
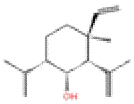
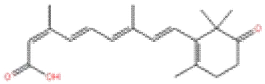
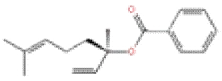
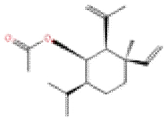
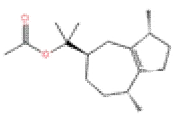
Supplementary Table S4. The whole 100 ranked allosteric metabolites as the result of Example 1 by AlloFinder.

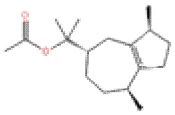
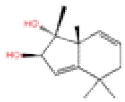
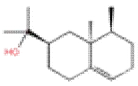
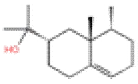
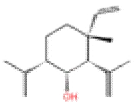
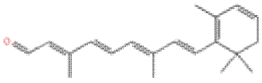
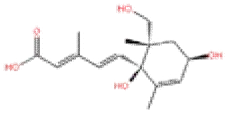
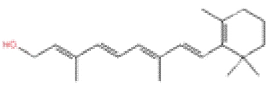
Rank	Compound Id[a]	Allosteric metabolites	Molecule Weight (MW)	Alloscore Score[b]
1	ZINC000030731390		316.2	7.12
2	**ZINC000012661824**		302.2	7.04
3	ZINC000030731384		316.2	7
4	ZINC000012496833		316.2	6.98
5	ZINC000036766734		302.2	6.95
6	ZINC000036766734		302.2	6.95
7	ZINC000030731379		318.2	6.91

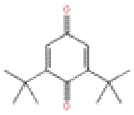
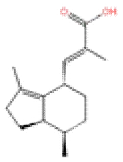
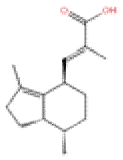
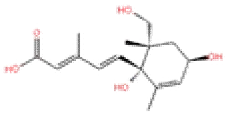
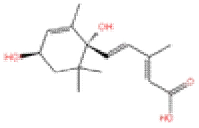
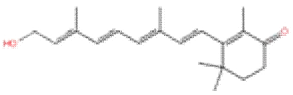
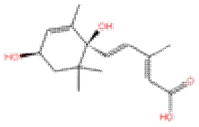
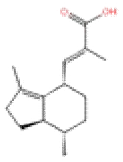
8	ZINC000001697340		258.2	6.69
9	ZINC000096014710		300.2	6.68
10	ZINC000014685555		328.2	6.68
11	ZINC000012496774		286.2	6.65
12	ZINC000033943508		300.2	6.62
13	ZINC000012496767		286.2	6.61
14	ZINC000022066345		284.2	6.6
15	ZINC000012953146		284.2	6.58

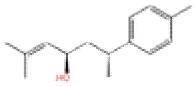
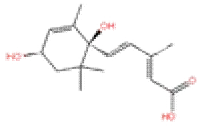
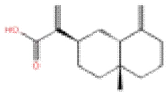
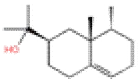
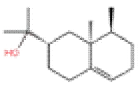
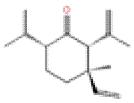
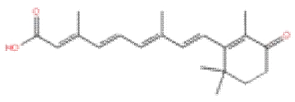
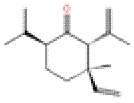
16	ZINC000004215527		284.2	6.56
17	ZINC000012496837		316.2	6.56
18	ZINC000012358651		300.2	6.55
19	ZINC000014453726		238.1	6.53
20	ZINC000085664079		264.2	6.53
21	ZINC000014453725		238.1	6.52
22	ZINC000012496764		286.2	6.52
23	ZINC000006030794		222.1	6.49

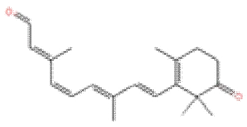
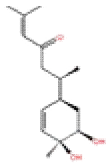
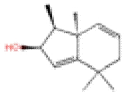
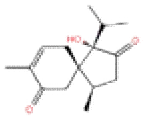
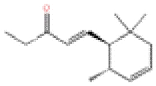
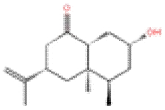
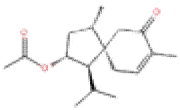
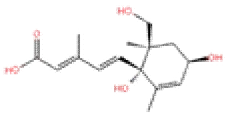
24	ZINC000034206309		300.2	6.48
25	ZINC000015201058		222.1	6.47
26	ZINC000003870195		284.2	6.46
27	ZINC000034206308		300.2	6.46
28	ZINC000014493252		222.1	6.4
29	ZINC000014493255		222.1	6.38
30	ZINC000004228262		284.2	6.37
31	ZINC000015048171		252.1	6.36

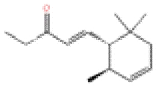
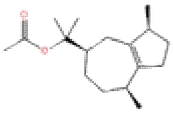
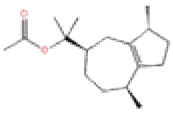
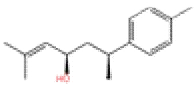
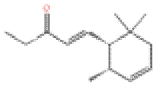
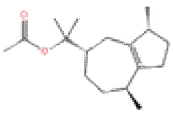
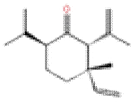
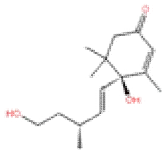
32	ZINC000013303509		266.1	6.36
33	ZINC000014724752		194.1	6.35
34	ZINC000014724754		194.1	6.34
35	ZINC000085664071		222.1	6.34
36	ZINC000004801872		314.2	6.32
37	ZINC000002015523		258.2	6.32
38	ZINC000085664083		264.2	6.31
39	ZINC000000394327		262.2	6.31

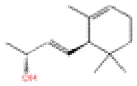
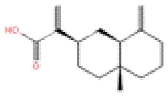
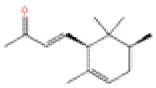
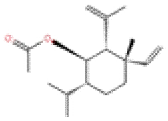
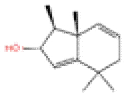
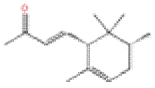
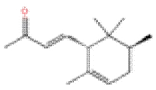
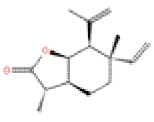
40	ZINC000001319187		262.2	6.3
41	ZINC000095619540		210.1	6.3
42	ZINC000006037475		222.1	6.3
43	ZINC000230080207		222.1	6.3
44	ZINC000085664074		222.1	6.3
45	ZINC000004096952		282.2	6.3
46	ZINC000014411632		282.1	6.29
47	ZINC000003831417		286.2	6.29

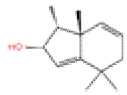
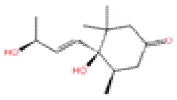
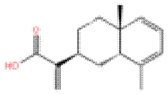
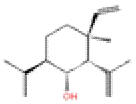
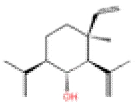
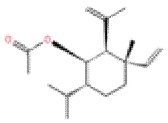
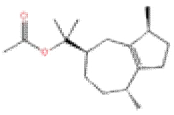
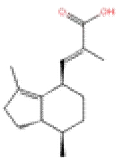
48	ZINC000000132915		220.1	6.28
49	ZINC000015657741		232.1	6.28
50	ZINC000013783213		232.1	6.28
51	ZINC000013303571		282.1	6.28
52	ZINC000013303506		266.1	6.28
53	ZINC000012502479		300.2	6.28
54	ZINC000013303512		266.1	6.26
55	ZINC000140984829		232.1	6.25

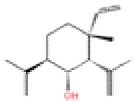
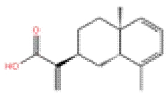
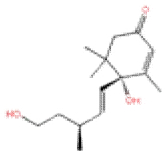
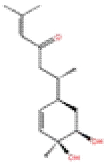
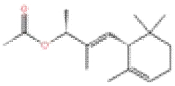
56	ZINC000013377435		218.1	6.25
57	ZINC000013303515		266.1	6.25
58	ZINC000085757038		234.1	6.24
59	ZINC000006030793		222.1	6.23
60	ZINC000015201060		222.1	6.22
61	ZINC000100778697		220.1	6.22
62	ZINC000013415989		314.2	6.22
63	ZINC000100778696		220.1	6.21

64	ZINC000085589170		298.2	6.18
65	ZINC000015048177		252.1	6.17
66	ZINC000085925336		194.1	6.16
67	ZINC000014442515		250.1	6.15
68	ZINC000095618230		206.1	6.14
69	ZINC000085848467		236.1	6.13
70	ZINC000085874223		278.2	6.13
71	ZINC000013303578		282.1	6.12

72	ZINC000038611118		206.1	6.12
73	ZINC000004578920		264.2	6.12
74	ZINC000012417829		264.2	6.1
75	ZINC000013377434		218.1	6.09
76	ZINC000095618231		206.1	6.09
77	ZINC000100781957		264.2	6.09
78	ZINC000100778694		220.1	6.08
79	ZINC000013303627		252.1	6.08

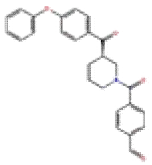
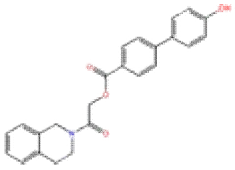
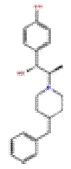
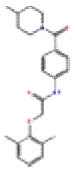
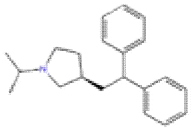
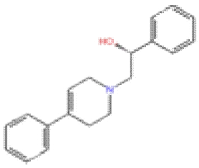
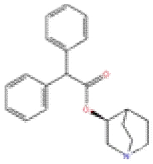
80	ZINC000005766979		194.1	6.08
81	ZINC000033985877		234.1	6.08
82	ZINC000001531537		206.1	6.08
83	ZINC000015149987		264.2	6.07
84	ZINC000014724749		194.1	6.07
85	ZINC000003875521		206.1	6.07
86	ZINC000002584370		206.1	6.07
87	ZINC000085664116		234.1	6.07

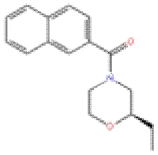
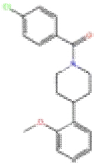
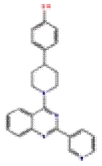
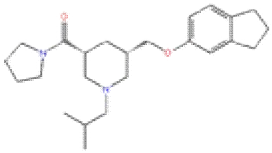
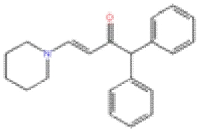
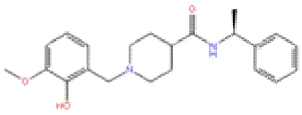
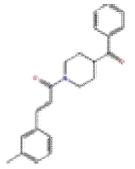
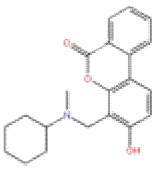
88	ZINC000014724750		194.1	6.06
89	ZINC000140122154		226.1	6.06
90	ZINC000015207976		232.1	6.06
91	ZINC000085664072		222.1	6.06
92	ZINC000015149979		222.1	6.06
93	ZINC000085664081		264.2	6.06
94	ZINC000004578919		264.2	6.06
95	ZINC000013783212		232.1	6.05

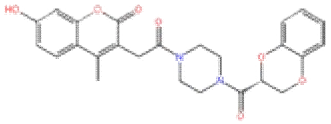
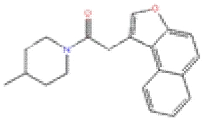
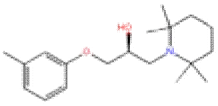
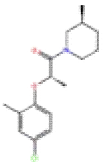
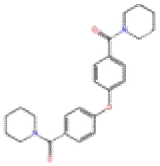
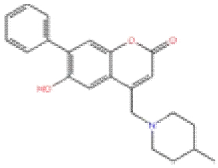
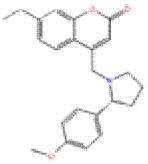
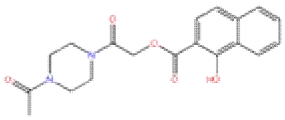
96	ZINC000085664077		222.1	6.05
97	ZINC000015207978		232.1	6.05
98	ZINC000013303615		252.1	6.05
99	ZINC000015048174		252.1	6.05
100	ZINC000033953588		250.2	6.04

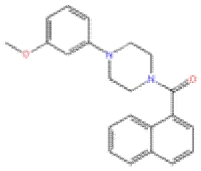
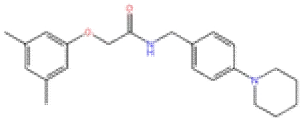
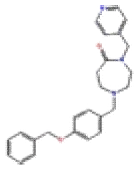
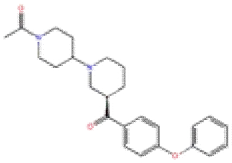
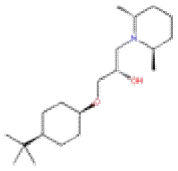
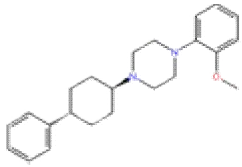
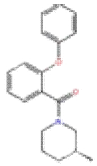
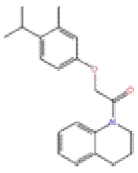
[a]Compound ID of each modulator is the same as the ID of the compound in ZINC (version 15). The known allosteric metabolite is highlighted in blue.
[b]Alloscore Score is calculated by the Alloscore algorithm (8).

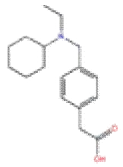
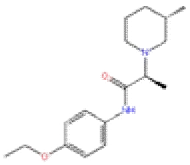
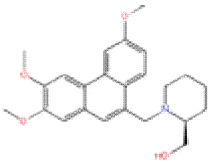
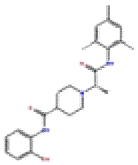
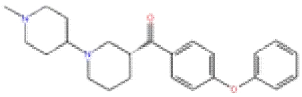
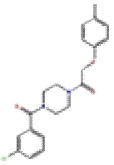
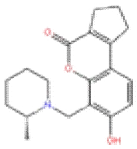
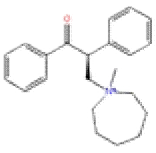
Supplementary Table S5. The whole 100 ranked allosteric modulators as the result of Example 2 by AlloFinder.

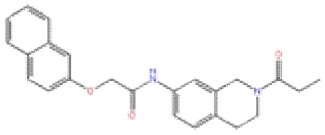
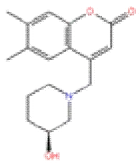
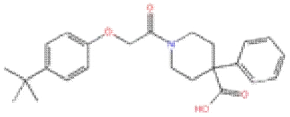
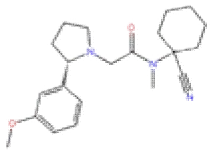
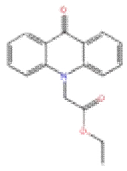
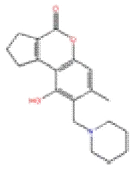
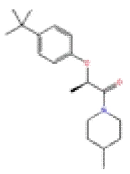
Rank	Compound Id[a]	Allosteric modualtor	Molecule Weight (MW)	Alloscore Score[b]
1	ZINC000012454396		413.3	7.63
2	ZINC000003484557		387.3	7.37
3	**ZINC000000014551**		325.2	7.2
4	ZINC000010460579		380.3	7.15
5	ZINC000001672573		293.2	7.13
6	ZINC000002651813		279.2	7.07
7	ZINC000005515780		321.2	6.97

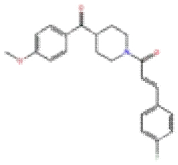
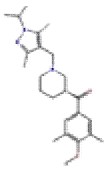
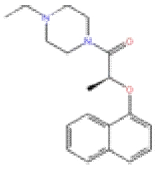
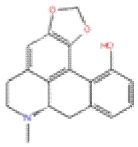
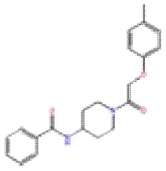
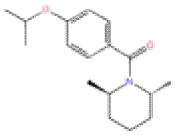
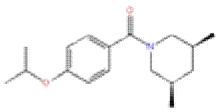
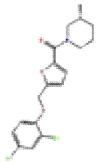
8	ZINC000012907490		269.2	6.95
9	ZINC000000135251		329.6	6.93
10	ZINC000012769146		383.3	6.87
11	ZINC000014755282		382.3	6.83
12	ZINC000002020697		305.2	6.81
13	ZINC000004725584		368.2	6.76
14	ZINC000009987912		333.2	6.75
15	ZINC000000142645		337.2	6.7

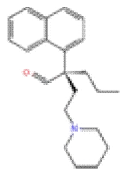
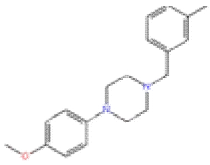
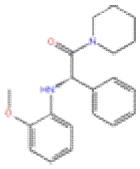
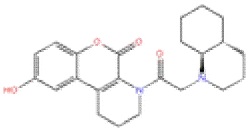
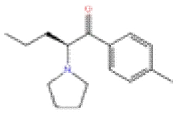
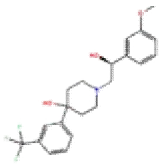
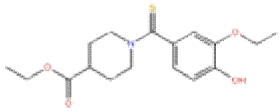
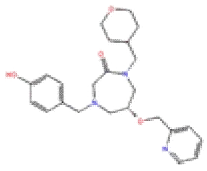
16	ZINC000009410579		464.3	6.69
17	ZINC000006777537		307.2	6.69
18	ZINC000002696119		305.2	6.63
19	ZINC000005536045		295.6	6.62
20	ZINC000001446457		392.3	6.62
21	ZINC000002122938		349.2	6.61
22	ZINC000012810542		363.2	6.6
23	ZINC000006121450		356.2	6.59

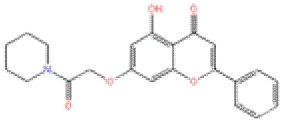
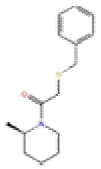
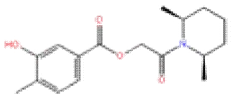
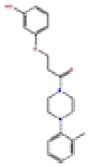
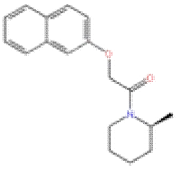
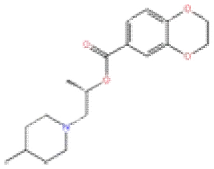
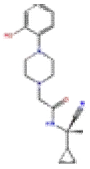
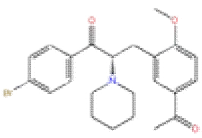
24	ZINC000006976394		346.2	6.58
25	ZINC000013669625		352.2	6.58
26	ZINC000014882802		401.3	6.57
27	ZINC000012200626		406.3	6.57
28	ZINC000013154395		325.2	6.55
29	ZINC000000599027		350.2	6.54
30	ZINC000005672601		295.2	6.54
31	ZINC000005697782		323.2	6.53

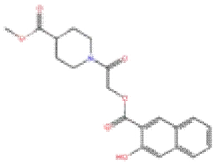
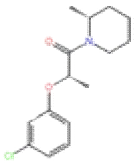
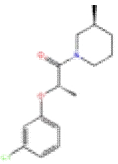
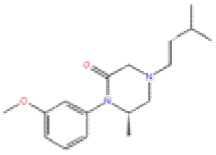
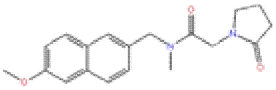
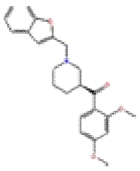
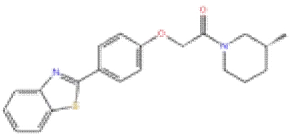
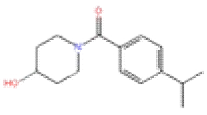
32	ZINC000011861546		275.2	6.51
33	ZINC000008142595		290.2	6.51
34	ZINC000001619789		395.2	6.51
35	ZINC000015045432		409.3	6.51
36	ZINC000011753571		378.3	6.49
37	ZINC000002760672		372.7	6.49
38	ZINC000000122731		311.2	6.46
39	ZINC000004752027		322.2	6.46

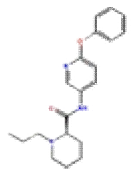
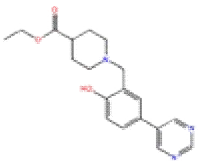
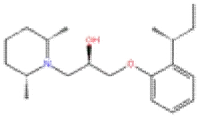
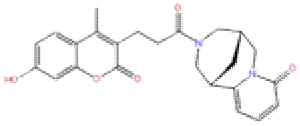
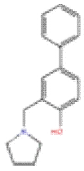
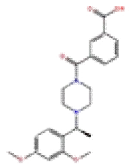
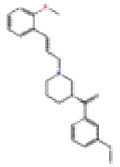
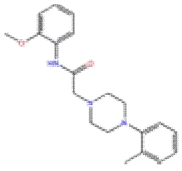
40	ZINC000009035184		388.3	6.46
41	ZINC000009497767		287.2	6.46
42	ZINC000013185723		395.3	6.45
43	ZINC000010227811		355.2	6.44
44	ZINC000003880343		281.2	6.43
45	ZINC000005235524		311.2	6.43
46	ZINC000005102075		347.7	6.42
47	ZINC000014396620		303.2	6.41

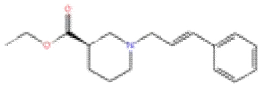
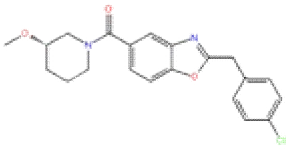
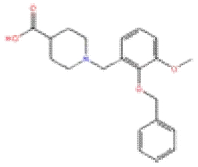
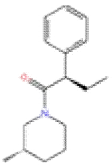
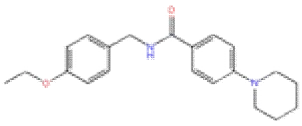
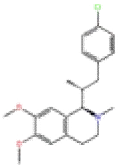
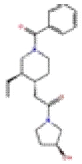
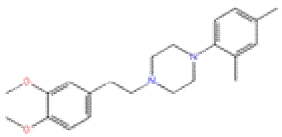
48	ZINC000009983384		367.2	6.4
49	ZINC000015063239		397.3	6.38
50	ZINC000013709298		312.2	6.37
51	ZINC000000898765		295.2	6.37
52	ZINC000010775710		352.2	6.37
53	ZINC000008570308		275.2	6.36
54	ZINC000006468367		275.2	6.36
55	ZINC000001083454		368.1	6.36

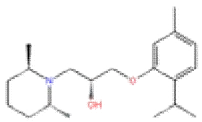
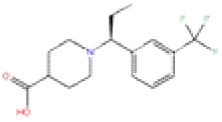
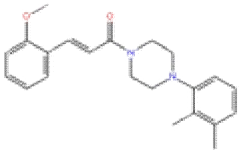
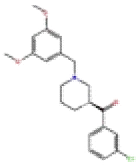
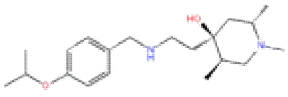
56	ZINC000005843999		323.2	6.36
57	ZINC000000202023		296.2	6.36
58	ZINC000007115860		324.2	6.35
59	ZINC000004689526		396.3	6.35
60	ZINC000000000681		245.2	6.33
61	ZINC000012292520		395.2	6.33
62	ZINC000009188298		337.2	6.3
63	ZINC000014746818		425.3	6.3

64	ZINC000009547646		379.2	6.3
65	ZINC000006471396		263.2	6.29
66	ZINC000012567989		305.2	6.28
67	ZINC000008415961		340.2	6.27
68	ZINC000006175576		283.2	6.26
69	ZINC000000808190		319.2	6.26
70	ZINC000006982754		328.2	6.26
71	ZINC000003355892		444.1	6.26

72	ZINC000005079221		371.2	6.26
73	ZINC000001480343		281.6	6.23
74	ZINC000000473873		281.6	6.23
75	ZINC000012034317		290.2	6.22
76	ZINC000008360937		324.2	6.22
77	ZINC000012207068		379.2	6.22
78	ZINC000003237214		367.3	6.21
79	ZINC000012974265		247.2	6.21

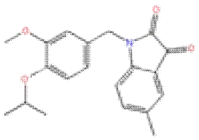
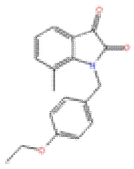
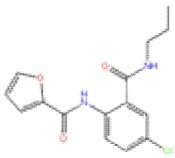
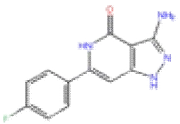
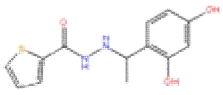
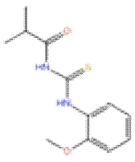
80	ZINC000011817411		339.2	6.2
81	ZINC000014992436		341.2	6.2
82	ZINC000002694403		319.2	6.19
83	ZINC000006624264		420.3	6.18
84	ZINC000001844767		253.2	6.17
85	ZINC000002456988		398.2	6.17
86	ZINC000012227333		365.2	6.16
87	ZINC000007620349		339.2	6.16

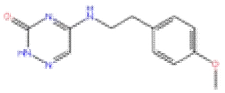
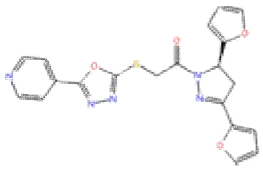
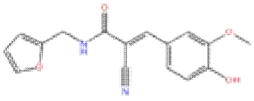
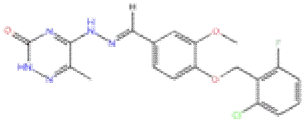
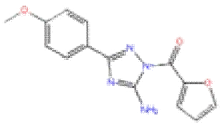
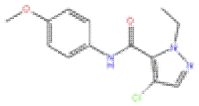
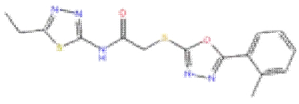
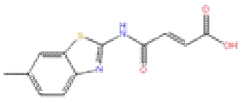
88	ZINC000005967438		273.2	6.15
89	ZINC000012588866		385.7	6.15
90	ZINC000003747794		355.2	6.15
91	ZINC000001466537		245.2	6.14
92	ZINC000010113682		338.2	6.14
93	ZINC000005735275		359.7	6.14
94	ZINC000008299601		344.2	6.14
95	ZINC000005160488		354.2	6.14

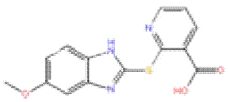
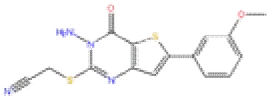
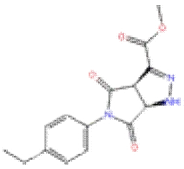
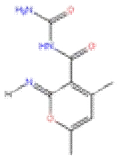
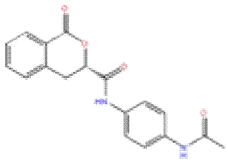
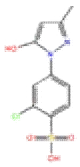
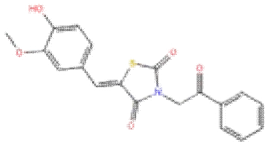
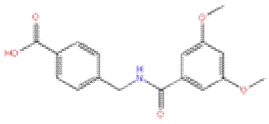
96	ZINC000000438151		319.2	6.13
97	ZINC000003760644		315.2	6.13
98	ZINC000008025580		350.2	6.12
99	ZINC000011955293		373.7	6.12
100	ZINC000002277989		334.2	6.12

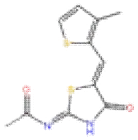
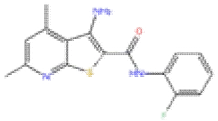
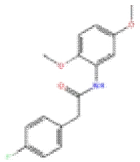
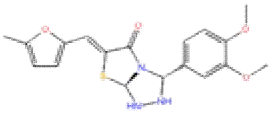
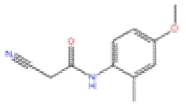
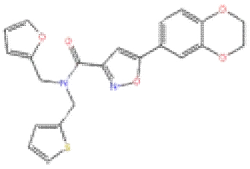
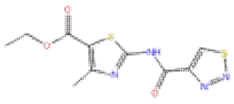
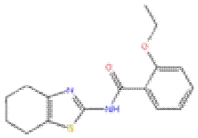
[a]Compound ID of each modulator is the same as the ID of the compound in ZINC (version 15). The known allosteric modulator is highlighted in blue.
[b]Alloscore Score is calculated by the Alloscore algorithm (8).

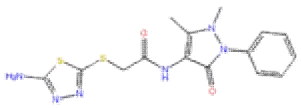
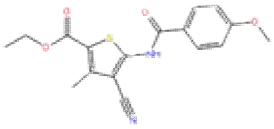
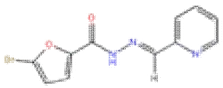
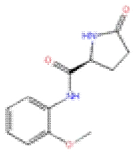
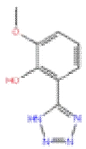
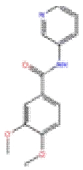
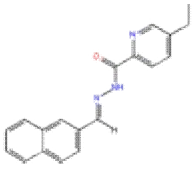
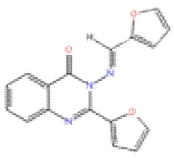
Supplementary Table S6. The whole 100 ranked allosteric modulators as the result of STAT3 by AlloFinder.

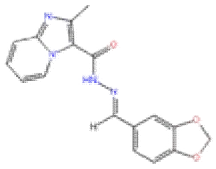
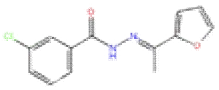
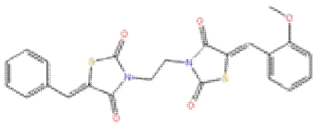
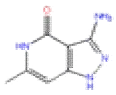
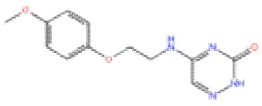
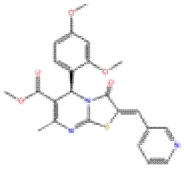
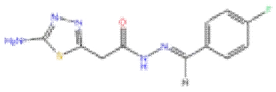
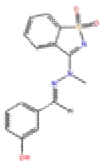
Rank	Compound Id[a]	Allosteric modualtor	Molecule Weight (MW)	Alloscore Score[b]
1	Specs_AT-057_42304490		339.2	5.28
2	Specs_AT-057_42304548		295.2	5.25
3	Specs_AP-853_41269897		382.3	5.09
4	Specs_AP-064_41684622		306.6	4.85
5	Specs_AR-270_43409614		246.2	4.79
6	**Specs_AH-034_11963955**		278.2	4.11
7	Specs_AN-329_40275188		252.2	4.1

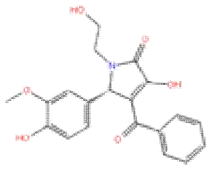
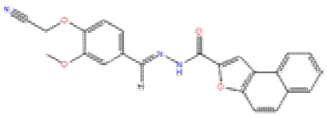
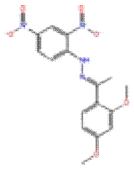
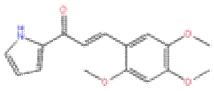
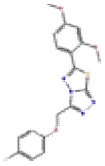
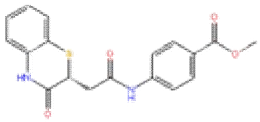
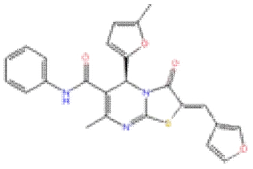
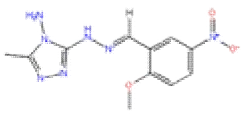
8	Specs_AS-871_43475754		248.2	4.1
9	Specs_AO-476_43250228		427.3	4.1
10	Specs_AS-871_40340909		298.2	4.1
11	Specs_AS-871_43239534		421.7	4.1
12	Specs_AE-848_42278486		288.2	4.09
13	Specs_AK-968_15605076		281.6	4.09
14	Specs_AG-690_40750303		369.3	4.09
15	Specs_AK-968_37156099		264.2	4.09

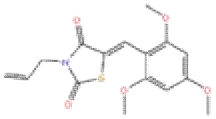
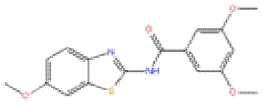
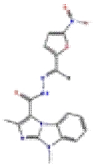
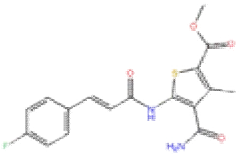
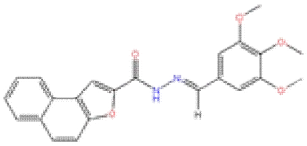
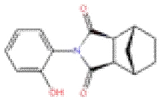
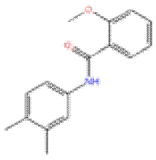
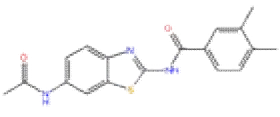
16	Specs_AI-204_43372123		305.2	4.09
17	Specs_AO-476_43417608		346.3	4.09
18	Specs_AT-057_43468141		303.2	4.09
19	Specs_AP-185_43377638		211.1	4.09
20	Specs_AS-871_43170215		324.2	4.09
21	Specs_AO-080_43441540		292.6	4.08
22	Specs_AT-057_40822043		369.3	4.08
23	Specs_AS-871_43477158		315.2	4.08

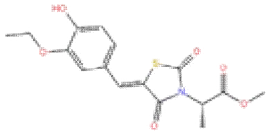
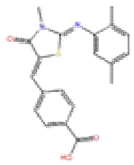
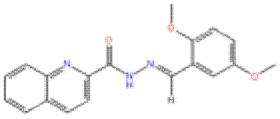
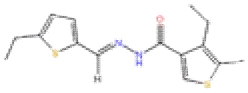
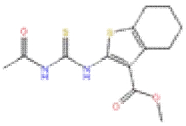
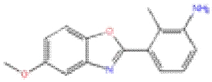
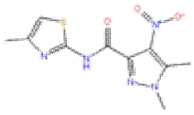
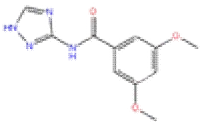
24	Specs_AN-648_15240030		268.3	4.08
25	Specs_AN-919_14790018		317.3	4.08
26	Specs_AK-968_10815012		289.2	4.08
27	Specs_AO-081_42298135		373.3	4.07
28	Specs_AP-970_43492187		204.1	4.07
29	Specs_AS-871_43475091		424.3	4.07
30	Specs_AS-871_43478445		302.3	4.06
31	Specs_AK-968_15605868		304.2	4.06

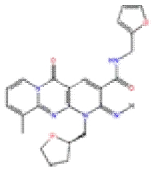
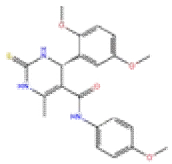
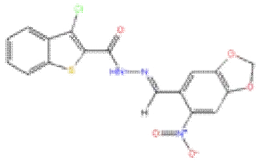
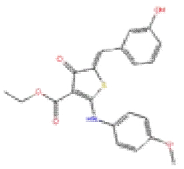
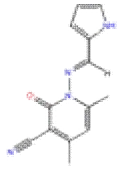
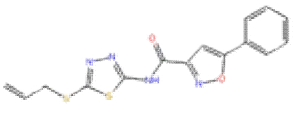
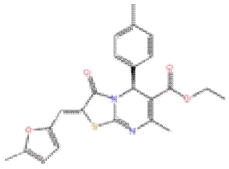
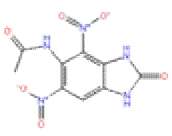
32	Specs_AG-205_32458033		380.4	4.06
33	Specs_AH-034_07286049		344.3	4.06
34	Specs_AK-968_37005275		298	4.06
35	Specs_AK-968_36977127		232.1	4.06
36	Specs_AN-655_15531018		194.1	4.06
37	Specs_AG-205_11957309		260.2	4.06
38	Specs_AG-690_40144611		307.2	4.06
39	Specs_AE-848_34078025		309.2	4.06

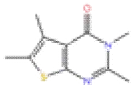
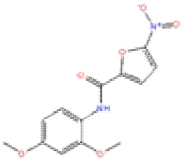
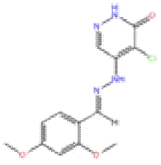
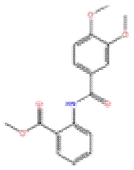
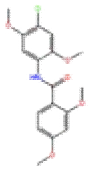
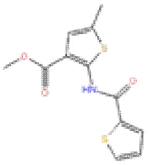
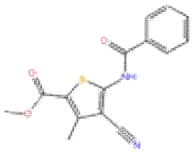
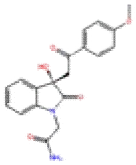
40	Specs_AH-487_34187016		326.2	4.06
41	Specs_AN-329_09187036		264.6	4.06
42	Specs_AE-848_13005012		466.4	4.06
43	Specs_AR-270_43409754		166.1	4.05
44	Specs_AS-871_43476174		264.2	4.05
45	Specs_AN-758_13664004		457.3	4.05
46	Specs_AM-900_15050008		285.2	4.05
47	Specs_AT-057_41349241		319.3	4.05

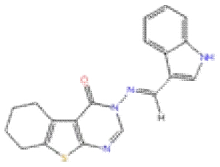
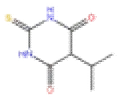
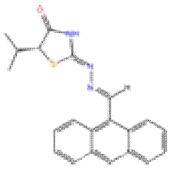
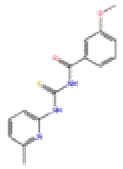
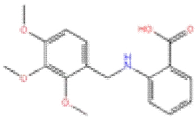
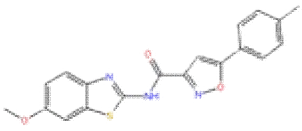
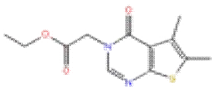
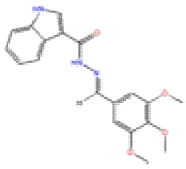
48	Specs_AT-057_42189162		369.2	4.05
49	Specs_AN-988_40679903		401.3	4.05
50	Specs_AG-690_33250063		362.2	4.05
51	Specs_AF-399_41753362		287.2	4.05
52	Specs_AO-365_43368513		392.3	4.04
53	Specs_AI-204_31720039		356.3	4.04
54	Specs_AK-968_41170900		449.3	4.04
55	Specs_AG-205_33159015		297.2	4.04

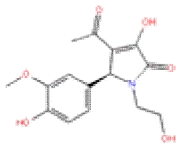
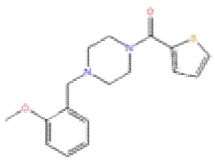
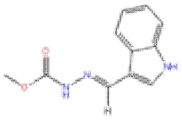
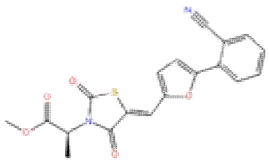
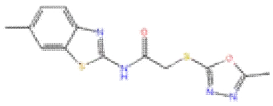
56	Specs_AH-487_41660228		335.2	4.03
57	Specs_AE-848_36255015		346.3	4.03
58	Specs_AG-690_11636562		370.2	4.03
59	Specs_AK-968_12117115		362.3	4.03
60	Specs_AN-988_40680531		406.3	4.03
61	Specs_AK-918_11909063		257.2	4.02
62	Specs_AN-652_11760408		255.2	4.02
63	Specs_AP-970_43012288		341.3	4.02

64	Specs_AH-487_41663592		351.2	4.02
65	Specs_AN-698_40718564		368.3	4.02
66	Specs_AG-690_03036047		339.2	4.02
67	Specs_AK-968_41922043		308.3	4.02
68	Specs_AN-329_42504558		312.3	4.02
69	Specs_AP-970_43374117		256.2	4.02
70	Specs_AK-968_41926710		285.2	4.02
71	Specs_AK-968_37171079		252.1	4.02

72	Specs_AO-022_41399309		437.3	4.02
73	Specs_AJ-292_12641038		413.3	4.02
74	Specs_AK-968_40340372		405.7	4.02
75	Specs_AK-968_41017619		397.3	4.01
76	Specs_AN-329_40602098		242.2	4.01
77	Specs_AS-871_43478439		350.3	4.01
78	Specs_AM-879_11808031		426.3	4.01
79	Specs_AG-690_36169011		281.1	4

80	Specs_AS-871_43477393		210.2	4
81	Specs_AG-690_11634414		292.1	4
82	Specs_AF-399_40714847		312.6	4
83	Specs_AN-652_11380105		315.2	4
84	Specs_AG-670_34988020		351.6	4
85	Specs_AK-968_40734154		281.3	4
86	Specs_AK-968_15256326		300.2	4
87	Specs_AK-778_41182446		354.2	4

88	Specs_AG-690_36808019		352.3	4
89	Specs_AJ-333_36115046		186.2	3.99
90	Specs_AE-848_11104122		365.3	3.99
91	Specs_AN-648_40682186		303.2	3.99
92	Specs_AN-465_43369870		317.2	3.99
93	Specs_AS-871_43475282		369.3	3.99
94	Specs_AP-906_40900953		268.2	3.98
95	Specs_AG-690_11171282		355.2	3.98

96	Specs_AT-057_43468823		307.2	3.98
97	Specs_AT-057_40126903		316.3	3.98
98	Specs_AN-329_40067267		219.1	3.98
99	Specs_AH-487_41657738		382.3	3.98
100	Specs_AP-853_42939736		326.3	3.98

[a]Compound ID of each modulator is the same as the ID of the compound in SPECS (http://www.specs.net). The active allosteric inhibitor of STAT3 found in the study is highlighted in blue.
[b]Alloscore Score is calculated by the Alloscore algorithm (8).

Reference

1. Sterling,T. and Irwin,J.J. (2015) ZINC 15—ligand discovery for everyone. *J. Chem. Inf. Model.*, **55**, 2324– 2337. DOI:10.1021/acs.jcim.5b0055
2. Gaulton,A., Hersey,A., Patr,A., Chambers,J., Mendez,D., Mutowo,P., Atkinson,F., Bellis,L.J., Cibri,E., Davies,M., *et al.* (2017) The ChEMBL database in 2017. *Nucleic Acids Res.*, **45**, D945–D954. DOI:10.1093/nar/gkw1074
3. Ihlenfeldt,W.D., Voigt,J.H., Bienfait,B., Oellien,F. and Nicklaus,M.C. (2002) Enhanced CACTVS browser of the Open NCI Database. *J. Chem. Inf. Comput. Sci.*, **42**, 46–57. DOI:10.1021/ci010056s
4. Huang,W., Lu,S., Huang,Z., Liu,X., Mou,L., Luo,Y., Zhao,Y., Liu,Y., Chen,Z., Hou,T. and Zhang,J. (2013) Allosite: A method for predicting allosteric sites. *Bioinformatics*, **29**, 2357–2359. DOI:10.1093/bioinformatics/btt399

5. Song,K., Liu,X., Huang,W., Lu,S., Shen,Q., Zhang,L. and Zhang,J. (2017) Improved Method for the Identification and Validation of Allosteric Sites. *J. Chem. Inf. Model.*, **57**, 2358–2363. DOI:10.1021/acs.jcim.7b00014
6. Chen,J. and Lai,L. (2006) Pocket v.2: Further developments on receptor-based pharmacophore modeling. *J. Chem. Inf. Model.*, **46**, 2684–2691. DOI:10.1021/ci600246s
7. Wang,R., Lu,Y. and Wang,S. (2003) Comparative evaluation of 11 scoring functions for molecular docking. *J. Med. Chem.*, **46**, 2287–2303. DOI:10.1021/jm0203783
8. Li,S., Shen,Q., Su,M., Liu,X., Lu,S., Chen,Z., Wang,R. and Zhang,J. (2016) Alloscore: A method for predicting allosteric ligand-protein interactions. *Bioinformatics*, **32**, 1574–1576. DOI:10.1093/bioinformatics/btw036
9. Gao,M. and Skolnick,J. (2013) APoc: large-scale identification of similar protein pockets. *Bioinformatics*, **29**, 597–604. DOI: 10.1093/bioinformatics/btt024
10. Paradis,E., Claude,J. and Strimmer,K. (2004) APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*, **20**, 289-290. DOI:10.1093/bioinformatics/btg412
11. O'Boyle,N.M., Banck,M., James,C.A., Morley,C., Vandermeersch,T. and Hutchison,G.R. (2011) Open Babel: an open chemical toolbox. *J. Cheminf.*, **3**, 33. DOI:10.1186/1758-2946-3-33
12. O'Boyle,N.M., Morley,C. and Hutchison,G.R. (2008) Pybel: a Python wrapper for the OpenBabel cheminformatics toolkit. *Chem. Cent. J.*, **2**, 5. DOI:10.1186/1752-153X-2-5
13. Kanehisa,M. and Goto,S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, **28**, 27–30. DOI:10.1093/nar/28.1.27
14. Huang,M., Chen,Z., Zhang,L., Huang,Z., Chen,Y., Xu,J., Zhang,J. and Shu,X. (2016) Screening and biological evaluation of a novel STAT3 signaling pathway inhibitor against cancer. *Bioorganic & Medicinal Chemistry Letters.*, **26**, 5172-5176. DOI:10.1016/j.bmcl.2016.09.073
15. Forouhar,F., Anderson,J.L.R., Mowat,C.G., Vorobiev,S.M., Hussain,A., Abashidze,M., Bruckmann,C., Thackray,S.J., Seetharaman,J., Tucker,T., *et al.*

(2007) Molecular insights into substrate recognition and catalysis by tryptophan 2,3-dioxygenase. *Proc. Natl. Acad. Sci.*, **104**, 473–478. DOI:10.1073/pnas.0610007104

16. Karakas,E., Simorowski,N. and Furukawa,H. (2011) Subunit arrangement and phenylethanolamine binding in GluN1/GluN2B NMDA receptors. *Nature*, **475**, 249–253. DOI:10.1038/nature10180
17. Roy,A., Srinivasan,B. and Skolnick,J. (2015) PoLi: a virtual screening pipeline based on template pocket and ligand similarity. *J. Chem. Inf. Model.*, **55**, 1757-1770. DOI: 10.1021/acs.jcim.5b00232