Selection and testing of protein target focused compound libraries using chemoinformatics methods integrated with MTS biological assay

PhD Thesis

Krisztina Dobi

Supervisors

Dr. György Dormán

Prof. Dr. Ferenc Fülöp

Institute of Pharmaceutical Chemistry

University of Szeged

2017

Contents

Publication list

Papers related to the thesis

1. **Dobi K**, Hajdú I,Flachner B, Fabó G, Szaszkó M, Bognár M, Magyar Cs, Simon I, Szisz D, Lőrincz Zs, Cseh S, Dormán Gy. **(2014)**
Combination of 2D/3D ligand-based similarity search in rapid virtual screening from multimillion compound repositories. Selection and biological evaluation of potential PDE-4 and PDE-5 inhibitors.
Molecules, 19(6):7008-39 ; **IF: 2.095**

2. **Dobi K**, Flachner B, Pukáncsik M, Máthé E, Bognár M, Szaszkó M, Magyar Cs, Hajdú I, Lőrincz Zs, Simon I, Fülöp F, Cseh S, Dormán Gy. **(2015)**
Combination of pharmacophore matching, 2D similarity search and *in vitro* biological assays in the selection of potential 5-HT_6 antagonists from large commercial repositories
Chem Biol Drug Des;86(4):864-80 **IF:2.396**

Other publications

3. Flachner B, Hajdú I, **Dobi K**, Lőrincz Zs, Cseh S, Dormán Gy, **(2013)**
Melanin koncentráló hormon receptor-1 (MCHR1) antagonista fókuszált könyvtár kiválasztása és *in vitro* biológiai szűrése AequosScreen esszével
Acta Pharmaceutica Hungarica 83:(3) pp. 71-87.

4. Hajdú I, Flachner B, Bognár M, Végh B, **Dobi K**, Lőrincz Zs, Lázár J, Cseh S, Takács L, Kurucz I, **(2014)**
Monoclonal antibody proteomics: Use of antibody mimotope displaying phages and the relevant synthetic peptides for mAb scouting
Immunology Letters. 160(2):172-7; **IF:2.37**

5. Flachner B, Tömöri T, Hajdú I, **Dobi K**, Lőrincz Zs, Cseh S, Dormán Gy. **(2014)**
Rapid *in silico* selection of an MCHR1 antagonists' focused library from multi-million compounds' repositories. Biological evaluation
Medicinal Chemistry Research, Vol. 23, Issue 3, 1234-1247; **IF:1.61**

6. Szaszkó M, Hajdú I, Flachner B, **Dobi K**, Magyar C, Simon I, Lőrincz Z, Kapui Z, Pázmány T, Cseh S, Dormán G
Identification of potential glutaminyl cyclase inhibitors from lead-like libraries by in silico and in vitro fragment-based screening
Molecular Diversity 21:(1) pp. 175-186. **(2017)** **IF:1,752**

Posters

Beáta Flachner, **Krisztina Dobi**, János Varga, Zsolt Lőrincz, Sándor Cseh:
Epitope mapping of mAbs recognizing protein markers of obesity: a phage display study
Cecon II, Budapest 2009.

Krisztina Dobi, Mária Pukáncsik, Beáta Flachner, István Hajdú, Zsolt Lőrincz, Sándor Cseh and György Dormán:
DEVELOPMENT OF MT ASSAYS TO DISCOVER POTENTIAL 5-HT6 ANTAGONISTS FROM FOCUSED LIBRARIES
Hungarian Molecular Life Sciences, Siófok 2013.

List of abbreviations

μM, nM	micromol, nanomol
2D/3D	2 dimensional/3 dimensional
5-HT	5-hydroxytryptamine
AC	adenylyl cyclase
AD	Alzheimer disease
ADMET	absorption, distribution, metabolism, excretion and toxicity
AR	aromatic-ring hydrophobic site
cAMP, cGMP	cyclic adenosine monophosphate, cyclic guanosine monophosphate
cLogP	(calculated) logarithm of the octanol-water partitioning coefficient
CNS	central nervous system
Da	Dalton
E	epinephrine
EDI	explicit diversity index
GABA	gamma-aminobutyric acid
GDP	guanosine diphosphate
HBA	hydrogen bond acceptor group
HBD	hydrogen bond donor group
HTS	high throughput screening
HYD	hydrophobic site
LE	ligand efficiency index

LLE	lipophilic ligand efficiency index
LLE_{AT}	LLE adjusted for heavy atom count
MPO	multiparameter optimization
Mwt	molecular weight
NE	norepinephrine
NMR	Nuclear Magnetic Resonance
PAINS	Pan Assay Interference Compounds
PD	Parkinson’s disease
PDE	phosphodiesterase
PI	positive ionizable atom
PKA	protein kinase A
R&D	research and development
SAR	structure-activity relationship
T	Tanimoto coefficient
T2D	Tanimoto coefficient of 2D similarity search
T3D	Tanimoto coefficient of 3D similarity search
VS	virtual screening

1. Literature overview

1.1. Contemporary drug discovery and development

The present dissertation represents a major direction in modern early phase drug discovery combining chemoinformatics, virtual screening and biological screening.

In order to demonstrate the relevance and position of this integrated approach the major phases of drug discovery are shown in Figure 1.

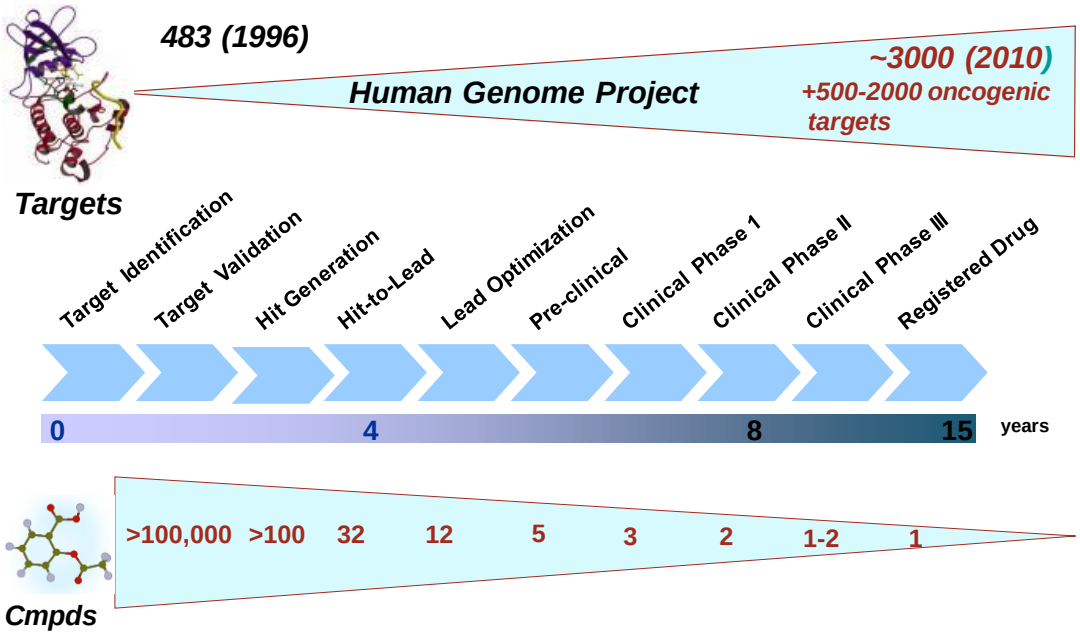


Figure 1. The overall pipeline of the contemporary drug discovery and development. (partly based on http://www.biolog.com/products-static/phenotype_microbial_cells_drug_discovery.php)

Figure 1. displays both the chemical and biological space. Compounds („Cmpds", most frequently small molecules, < 500 Da molecular weight) covers the so called chemical space. A typical drug discovery project starts with huge compound collections (small molecule banks, libraries, commercial resources) and during the multiple phases it leads to one single registered drug. This huge attrition rate is responsible for the high expenditure of the drug discovery pipeline which now exceeds 1000 million USD.[1] The biologically relevant chemical space (or biological space) contains the potential drug targets that could interact with small molecules. The number of the drug targets of the registered drugs is slowly increasing[2] (483 in 1996) due to the expensive target validation process. However, novel targets often allow more effective therapies.[3]

1.2. Trends and deficits of drug design

One of the most important tasks of the pharmaceutical industry is the rapid production of new and safe drugs. The failures increase primarly the costs of clinical investigations, therefore, any process that terminates as early as possible the development has become a major interest. Some estimates suggest that the expenditure of a new drug development has been exceeded the $1000 million. Increasing attrition rate and decreasing number of New Chemical Entities were observed at the end of the 20th century.[4,5] Combinatorial chemistry and HTS as emerging technologies were expected to reverse these trends in the early 2000's.

After a decade it has become clear that these approaches did not increase the success rate of the drug discovery inspite of the high expectation.[6,7] Even though combinatorial chemistry contributed to generate multi-million compound libraries it does not improve the overall performance of the drug discovery process due to the following reasons:

1. Structural redundancy and lack of novelty: most of the compound providers produced compounds with easy, parallizable chemistry. Difficult chemistries were neglected, therefore, the large commercial libraries contain relatively similar molecules and miss important chemotypes.
2. Since all the polar functional groups were substituted (capped) in order to achieve diversity the library members have become lipophilic (having high LogP values) compared to the existing drugs.
3. Scaffold diversity was rather low, the structural diversity of the libraries was mostly due to large substituent diversity ('decoration of the core/scaffold structure').

In the post-combinatorial era new approaches emerged:

1. Fragment-based or lead like libraries[8] such as GDB-13, „the chemical universe database", which enumerates 977 million organic molecules up to 13 atoms of C, N, O, S and Cl following simple chemical stability and synthetic feasibility rules.[9]
2. Small focused libraries with large scaffold diversity[10]
3. Novel scaffold chemistries including sp^3-rich core structures[11]

The first approach introduced the principle of the lead-likeness, the rule of three (LogP=3, Mwt=300)[12] by the analogy of drug-likeness (Lipinski's Rule-of-5)[13] (Figure 2).

Properties	Leadlikeness	Druglikeness
Molecular weight (MW)	≤350	≤500
Lipophilicity (clog *P*)	≤3.0	≤5.0
H-bond donor (sum of NH and OH)	≤3	≤5
H-bond acceptor (sum of N and O)	≤8	≤10
Polar surface area (PSA)	≤120 Å^2	≤150 Å^2
Number of rotatable bonds	≤8	≤10
Structural filters	Reactive groups Warhead-containing agents Frequent hitters Promiscuous inhibitors	

Figure 2. Comparison of properties used for leadlikeness and druglikeness criteria. (reproduced from Böhm M. Virtual Screening of Chemical Space: From Generic Compound Collections to Tailored Screening Libraries. pp. 3-26. in: Virtual screening: principles, challenges, and practical guidelines Vol. 48.,2011).

The small, less lipophilic molecules are easier to optimize with the attachment of substituents. Fragment screening and linking are also preferred methods in our days.[14]

In summary, the pharma industry has to overcome the challenges and difficulties of low productivity, rising R&D costs, divided proprietary products.[15,16] The need for developing safe and innovative drugs, under the increasing pressure of speed and cost reduction, has shifted the focus toward improving the early discovery phase of lead identification and optimization. „Fail early, fail fast, and fail cheap" has become the key principle to increase the efficiency in drug discovery.[17]

The earlier method for discovering novel hits using high-throughput screening (HTS) of large compound libraries has coupled with the extensive use of virtual (*in silico*) screening (VS). The advantages of the integration of the virtual screening with HTS was early recognized[18] and since then many successful applications of virtual screeing using huge compound libraries were reported.[19]

1.3. The emergence of virtual screening (VS)

Over the past two decades huge compound repositories were built exploiting historical collections as well as compound libraries. At the same time the chemoinformatics methods have developed rapidly, the computational power increased allowing fast or real time calculations. In addition, deeper knowledge has gathered about the small molecule - protein interactions using state-of-the-art X-ray crystallography, docking and 3D modelling.[20]

Virtual screening has become a popular technique[21] since it was expected to reduce the synthesis and biological screening cost and shortens the life cycles of the discovery phases. From its emergence it combined 2D/3D ligand-based and structure-based approaches.[22] The major elements and the most important VS approaches are shown on Figure 3.

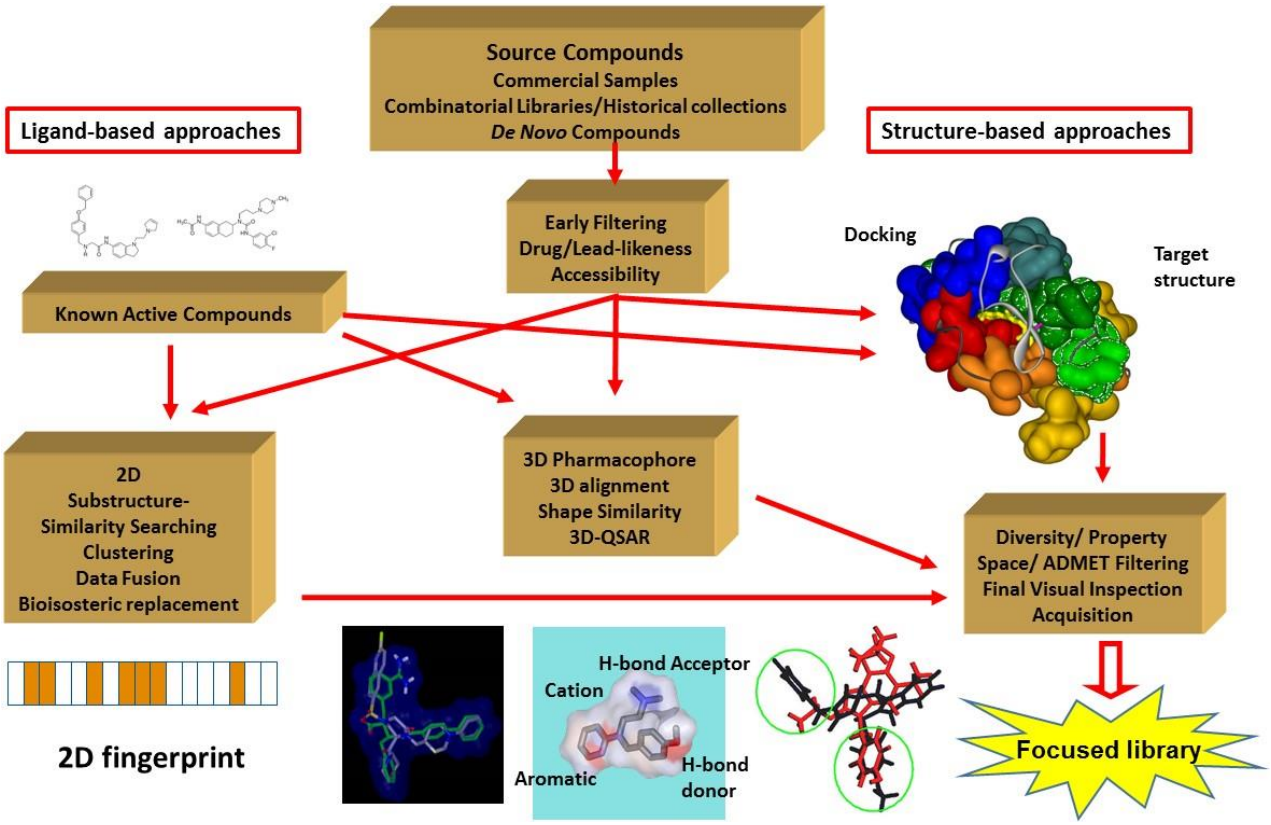


Figure 3. Overview of the Virtual Screening approaches

1.3.1. The chemical space

The „source compounds", the „known active compounds" and the „protein target structures" are the most important elements in VS and these elements constitute the chemical space.

The full size of the chemical (virtual) space is in theory between 10^{13} to 10^{180} molecules.[23] It was estimated that with less than 30 heavy atoms more than 10^{63} molecules with a molecular weight of less than 500 can be generated, predicted to be stable at room temperature and stable toward oxygen and water (this collection represents the medicinally relevant chemical space, Figure 4). [24] The overall size of the existing compound collection is estimated over 35M based on the Zinc library.[25,26] ChemSpider[27] contains 58M structures, while eMolecules[28] has 5.9M commercially available, unique structures.

The biologically relevant chemical space (Figure 4) contains about 100.000 peptides (while only a small portion of that space is druggable, which means can be modulated with small molecules).[29] Several databases provide access to biological activity data for small molecules that can serve as input data for model building and virtual screening. For example, PubChem[30] contains 91,371,681 reported compounds together with 232,760,104 bioactivities; ChEMBL[31] collects 2,036,512 compounds with 14,371,197 bioactivities for 11,224 targets; Binding_DB[32] reports 590,985 compounds with 1,328,228 binding data for 6,922 protein targets as public databases. Wombat[33] is a commercial Bioactivity Database for Lead and Drug Discovery databases and contains 300k structures with biological activities. Pharmaprojects[34] monitors the progress of drug candidates with searchable profiles for over 50,000 drugs and drug candidates; Integrity database[35], which is a multidisciplinary database, and supports the drug research and development. The indicated subportion (dotted circle) of the chemical space could be exploited by virtual screening. (Figure 4).

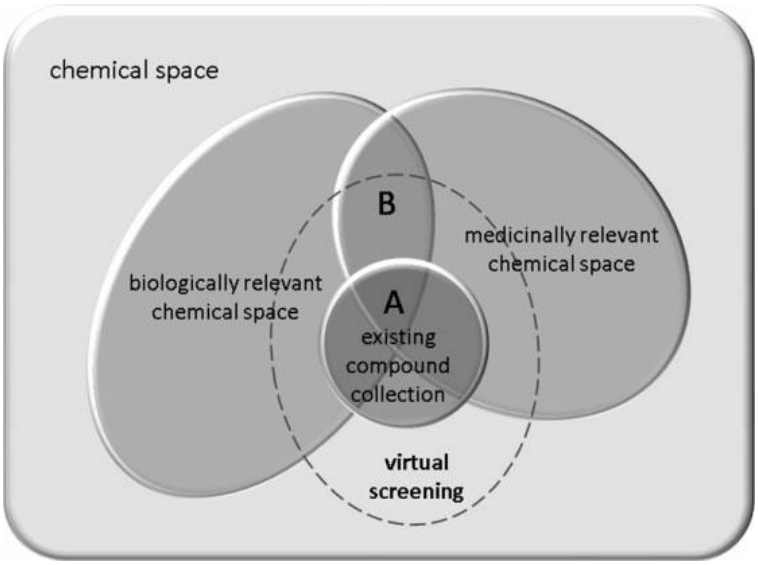


Figure 4. Regions of biologically and medicinally relevant chemical space within the continuum of chemical space. Only a small portion of chemical space has been sampled by existing compound collections, which led to the discovery of drugs (A). Virtual screening has the unique opportunity to expand into unexplored chemical space to find new pockets of space where drugs are likely to be discovered (B).
The Chemical space (reproduced from Böhm M. Virtual Screening of Chemical Space: From Generic Compound Collections to Tailored Screening Libraries. pp. 3-26. in: Virtual screening: principles, challenges, and practical guidelines Vol. 48.,2011).

VS methods can be used through the entire process of drug development, starting with the search of lead molecules to the simulation of clinical trials. According to predictions

in silico approach will reach 10-15 % of the research and development expenditures in the next 5-10 years.

1.3.2. Ligand-based virtual screening

The key concept of the ligand-based VS approaches[36,37] is the Similarity Property Principle, which states that similar molecules should have similar biological properties.[38,39] In other words, a molecule that has not been tested for biological activity but that is structurally similar to an active molecule (reference or seed molecule) is also likely to be active.[40] If determining the similarity between the biologically active reference compound and each molecule in a database, followed by ranking the database molecules according to the similarities would lead to potentially active, target-focused libraries.

The similarity search uses 2D fingerprints, i.e., binary strings encoding the presence or absence of a substructure within the molecules.[41] Applying simple 2D fingerprints is often the method of choice[42] particularly when numerous reference compounds and multimillion compound databases are available. Its computational efficiency is coupled with demonstrated effectiveness in many comparative studies.[43]

Rapid 2D similarity search can be performed on multimillion compound's databases if structures of active molecules are available therefore, it is a „real time" method compared with the other ligand-based approaches (including shape-based screening - 3D flexible alignment or pharmacophore building). It was also reported that 2D methods could often provide better VS performance than 3D shape-based approaches even though 3D molecular shape and conformation is crucial for ligand binding.[44,45]

Focused library screening often results in a many fold increase in the hit rate compared with random screening of a commercial libraries.[46,47] In many cases virtual screening methods and *in vitro* HTS were combined[48] and found complementary to each other.[49,50,51]

The 2D similarity search can be completed with physico-chemical parameter and diversity filtering. The empirical physico-chemical parameter ranges stand for drug likeness. If most of the parameters of the drug candidate fall into the pre-defined ranges the concerning molecule could be administered orally.[52] The major physico-chemical properties are Molecular weight, cLogP, H-bond donors, H-bond acceptors, rotatable bonds and Topological polar surface area.

The first 4 parameters are included in the Lipinski's Rule of 5[53], while the other 2 are noted as Veber rules.[54] These days heavy atom count often replaces molecular weight since it characterizes more accurately the size of the molecule. Heavy atom count is also an important

term when calculating the Ligand Efficiency index (LE)[55], which is an important indicator of the compound developability. LE values are expressed as LE = (1.37/ HA(heavy atom)) x p(Activity), where p(Activity) is pKi, pIC_{50}, or pEC50 (LE is preferred if >0.3). Similar measure is the Lipophilic ligand efficiency index (LLE), which is simply the difference between p(Activity) and lipophilicity (cLogP or LogD) and is an estimate of the specificity of a molecule in binding to the target relative to partitioning into 1-octanol. LLE adjusted for heavy atom count (LLE_{AT}) is scaled to be comparable to LE (LLE is preferred if >5, while LLE_{AT} is preferred if >0.3, similar to LE).

Additional filters. There are filters to refine the focused library obtained by 2D ligand-based similarity seaerch. ADMET prediction[56] is related to the physico-chemical parameter calculation and aiming at removing compounds that are not suitable for oral drug discovery. ADMET filters are more specific than the simple drug-likeness indicators. Another two filters remove compounds that contain reactive functional groups[57], and so-called frequent hitter compounds (PAINS)[58] that could react or bind many different protein targets. Finally, there are prediction systems that could assess the synthetic accessibility[59], which is particularly important in the developability of the compounds if identified active.

Property space filtering. Virtual focused library generation is often linked to target families that represent a distinct chemical, biological and property space.[60] Several target families have a different property range, therefore, defining a target-specific parameter space is often more practical for focused library filtering than the simple Lipinski's Rule of 5 based filtering. Scientists at pharma companies also identified empirically the favorable ranges of the critical physico-chemical parameters for CNS compatibility or desirability and introduced a multiparameter optimization (MPO) score.[61]

Those rules propose general limits (cut-off values and ranges) for library filtering and represent a parameter "window" favorable for the specific target area. Furthermore, such property space successfully differentiates drugs–non-drugs or actives from non-actives by using annotated datasets[62] within target families.

A typical work-flow of the 2D similarity search is shown in Figure 5.

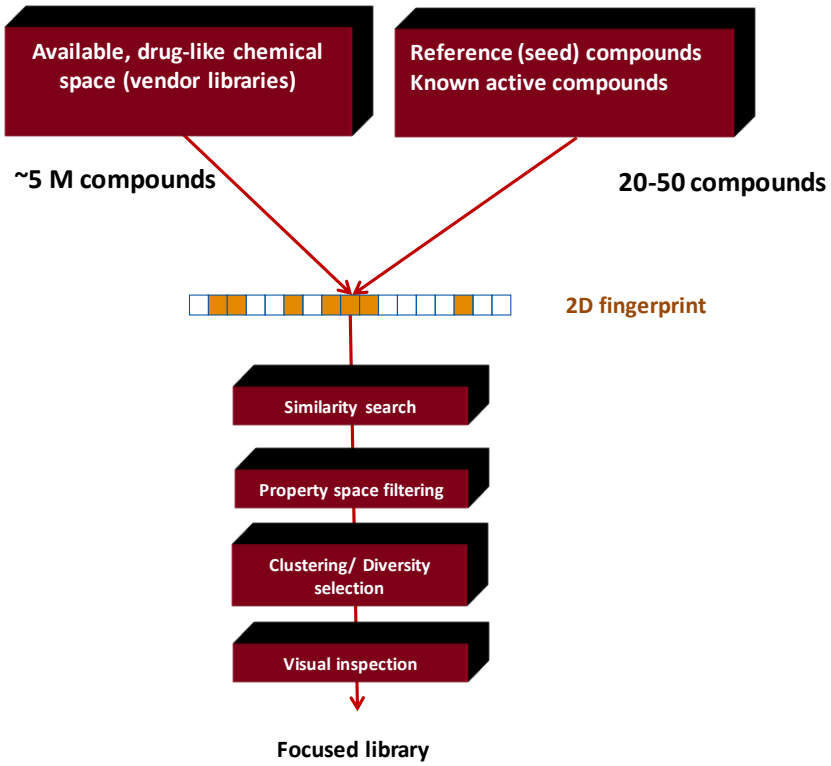


Figure 5. Typical work-flow for 2D similarity search

Diversity and related selections. If the focused library is still huge, diversity selection, clustering or scaffold analysis are the method of choice to reduce the number of compounds for biological screening. Diversity selection[63] could follow the *in silico* property-based filtering, which results in a more diverse molecule library, where the structural clusters or molecular scaffolds are evenly represented. Clustering is usually based on the Maximum Common Substructure search[64] algorithm. Scaffold analysis could also be combined with the simple chemical diversity selection and a combined index could reflect both (EDI, Explicit Diversity Index).[65]

Similarity measures.[66] Most frequently Tanimoto coefficient[67] is used for measuring similarity[68] in spite of its marked size-dependency.[69] It typically yields low similarity values when the reference molecules are relatively small and just lead to few bits set in the fingerprint. There are multiple similarity measures. There are so called data fusion methods[70] that combine the results of the similarity searches. After combining (fusing) the similarity scores the compounds are ranked according to the accumulated score. Group fusion method applies multiple reference compounds and combines the individual similarity measures.[71] In this approach the similarity scores are first calculated to 'n' reference compounds for each compound within a medium sized searchable compound library and then the 'n' similarity

measures obtained for each compound of the library are accumulated ('fused'). The calculation is followed by ranking the compounds according to the 'fused' similarity measures. If several structurally diverse reference structures are available, novel analogues can be selected where the structural motifs coming from different reference compounds are joined in one single compound.[92] Turbo similarity searching[72] is used when relatively small number of reference compounds is available. Then the nearest neighbors of the reference structures (analogues with highest similarity score) are also involved in the similarity search on the same database. Finally, involvement of both actives and non-actives could also be involved into the similarity search. Since most of the available drug databases collect only active compounds (except for some rather expensive annotated libraries like Wombat[33]), normally non-actives can be involved into the similarity search only after the first *in vitro* screening round.

Similarity vs novelty.[73] This is one of the critical issues in 2D ligand-based similarity selection approaches. Structural analysis of the hits is a necessary task after biological evaluation of the focused library. It could include generating a structural evolution tree from the seed to the *in vitro* hit compounds as well and a novelty check compared with the seeds. Novelty could mean different core structures (chemotypes), a different arrangement of the various motifs available in the reference compounds or simply a different substitution pattern.[74]

The performance of the 2D similarity search relies on the diversity of the seeds and the searchable chemical space (vendor libraries), the computational methods (fingerprints) as well as the applied criteria (similarity cut-off values, and property space filters etc.). In order to make the seed compounds more diverse their bioisosteric analogues[11] can be generated and applied in the similarity search.

Various groups have analyzed large activity data sets and came to the conclusion that on average there is a 30% probability that two compounds share the same activity if the compound is similar within a certain cut-off value (2D Tanimoto coefficient – T2D $\geq$ 0.85). However, applying T2D $\geq$ 0.65 cut-off value the virtual hits still represent acceptable similarities and reasonable bioactivities. On the other hand, such higher structural similarity does not necessarily lead to novel chemotypes or scaffolds, which is one of the major challenge in the contemporary drug discovery. It is widely discussed that new chemotype discovery or scaffold hopping is possible if we go below T2D = 0.65 similarity value[75,76] but, of course, it also reduces the hit rate. In agreement with the above reports, in previous studies

our group found that lower similarity thresholds provide higher structural diversity which might lead to novel chemotypes with/or without scaffold hopping even though the hit rate is somewhat reduced.[77] Multiple *in silico* and *in vitro* screening rounds could normally provide a hit structure refinement and hit validation (enough structures for SAR) with increased biological activity.

In multiple screening rounds the evolution and distribution of the hits regarding the source seed (reference) compounds could be taken into consideration in the selection of the 2nd round focused library. In many cases the 2nd round screening is combined with additional ligand and/or structure-based methods with 2D/3D fragment-based approaches, 3D similarity and pharmacophore search methods as well as 3D docking in order to increase the novelty level.

Combination with 3D similarity methods. The possible binding features of small molecules can be assessed by their conformational flexibility and shape. Applying flexible alignment (such as Screen3D software), the statistically average conformations generated by the accumulated dynamic 3D structures can be compared and based on that the similarity between two compounds could be assessed and characterized by 3D similarity measures (3D Tanimoto = T3D).[78] For better and more efficient performance the 2D/3D similarity selection methods can be combined. Thus, we proposed and applied a fusion (2D/3D) score and a reasonable cut-off value.

Combination with Fragment-based methods. Fragments are small molecules typically with a molecular mass between 150–250 Da and rather hydrophilic, yet fragments represent virtually a 'combinatorial explosion' of the chemical space.[79] Fragments allow a more effective sampling of the available chemical space than is possible with more complex molecules and are expected to explore more binding motifs. Furthermore, although fragment hits are weak binders, at lower level of complexity, there is a higher probability that the compounds match the receptors' major interactions. More complex molecules could not fit perfectly or bind to the binding site, and thus, fragments are very "atom efficient" and serve as better starting points for optimization. Finally, by reducing the number of pharmacophoric elements in the initial fragment, only the necessary interactions can be built into the compound during optimization (fragment decoration).

Similarly to the drug-likeness indicator Rule-of-5[26], the Rule-of-3[25] was adapted for fragment molecules and it comprises the following 4 criteria: Mwt < 300 Da; number of hydrogen bond donors ≤ 3; hydrogen bond acceptors ≤ 3, and calculated logP (clogP) of ≤ 3.

The typical fragment-based approach starts with fragment library design followed by fragment screening. Finally, fragment merging, linking, or growing would lead to drug-like candidates.[80] Thus, fragments could be gained from the disconnection of bioactive drug-like compounds if available and the resulting fragments would serve as starting points for *de novo* fragment-based design. After carrying out virtual and *in vitro* screening of the fragments, active fragments could be remerged if their binding mode is complementary within the binding/active site.[81]

Combination with pharmacophore models and search. Pharmacophore-based virtual screening incorporates 3D structural information of the known active ligands and it can be successfully applied in many chemoinformatics programs. Various pharmacophore models can be built based on the structure of the available seed (reference) compounds.

Combination with 3D docking methods.

When the 3D structure of the target protein is known or can be easily deduced, docking procedures can be involved in the focused library generation. 3D structure-based virtual screening offers a *de novo* approach to select compounds from focused libraries without using prior knowledge of the active reference compounds. This virtual screening involves docking of the candidate small molecules into a protein target followed by the application of a scoring function which estimates the probability if the small molecule is able to bind the protein with high affinity. In this case the prior information is the 3D structure of a target protein, which can be determined by NMR, X-ray crystallography or homology modeling.[82] 3D virtual screening by docking is substantially slower than the 2D similarity-based virtual screening, therefore, it limits its application when multimillion compound library is used for selecting the active compounds. Therefore, it is normally used in the 2nd round focused library selection and *in vitro* screening. Our group applied this method in fragment merging of glutaminyl cyclase inhibitor disovery.[81]

1.4. Summary of the literature of the protein targeted by the focused libraries; pharmacological importance

1.4.1. Phosphodiesterase-4/5

Phosphodiesterases (PDEs) belong to a family of cyclic nucleotide degrading enzymes, which control the intracellular levels of cAMP and cGMP.

The cyclic adenosine monophosphate (cAMP) and cyclic guanosine monophosphate (cGMP) are ubiquitous secondary messengers playing role in several processes in many different

tissues of the body. cAMP and cGMP are produced by the activation of adenylate cyclase and both soluble and particulate guanylate cyclase.

There are 11 known PDE families, produced in the central nerveous system (PDE1 - 11), with at least 21 subfamilies or subtypes which differs in structure, substrate specificity, body tissue dispersion, regulation by kinases, protein-protein interaction and inhibitor selectivity.[83,84] PDE-4 isozymes play role in pathologies with inflammatory symptoms such as asthma, chronic obstructive pulmonary disease (COPD), inflammatory bowel disease, atopic dermatitis (AD), psoriasis and rheumatoid arthritis (RA). PDE-4 inhibitors have importance also in CNS, including antidepressant, and memory-enhancing effect mostly in long-term treatment.

PDE-4 has four isotypes (PDE-4A-D), these catalyze the hydrolysis of cAMP. The catalytic domains of PDE-4A, PDE-4B, PDE-4C and PDE-4D were elucidated by X-ray crystallography.

PDE-4B and 4D have similar active sites, PDE-4C appears to be different from the other PDE-4 subfamilies. The development of PDE-4 subfamily-selective inhibitors is based on these structural differences.[85]

The first commercial PDE-4 inhibitor was roflumilast 1 (Daxas, Daliresp and Libertek), which was approved in Europe in 2010 and in the United States in 2011.[86]

1.4.2. 5-Hydroxytryptamine receptor 6

5-HT_6 is implicated in the pathogenesis of neurological and cognitive disorders including Alzheimer's disease, schizophrenia etc. as well as obesity. [87,88]

5-Hydroxytryptamine (5-HT, serotonin) is a major neurotransmitter in the central (CNS) and in the peripheral nervous system, which mediates many effects through its interaction with a family of receptors called 5-HT receptors. The mediated processes are mood, cognition, perception, pain, feeding behavior, smooth muscle contractility and platelet aggregation. Seven 5-HT families (5-HT_1–5-HT_7) are known, with 14 different subtypes. The 5-HT_3 receptor is an exception, it is a ligand-gated ion channel, while the other 5-HT receptors are G protein–coupled receptors.[89,90] 5-HT_6 receptor takes part in the glutamatergic system and also able to modulate the γ-aminobutyric acid (GABA) receptor, considering that 5-HT_6 receptor are expressed on GABAergic neurons.[91] The seven transmembrane-spanning human 5-HT_6 receptor was cloned and described previously.[92] 5-HT_6 antagonists are able to increase

acetylcholine levels, thereby these antagonists have therapeutic potential to improve normal and reduced memory.

It was shown with 5-HT_6 transfected cell lines that the receptor is able to positively couple to adenylyl cyclase.[93] The cyclase functional assay can be used to identify potential 5-HT_6 agonists based on stimulation of adenylyl cyclase activity and 5-HT_6 antagonists based on antagonism of 5-HT induced cAMP formation. Under agonist (first messenger) activation, adenylyl cyclase (AC) is induced to convert ATP to cyclic AMP (cAMP), which is a second messenger and activates protein kinase A (PKA). The coupling of 5-HT_6 receptors to other G proteins ($G_{i/o}$ or G_q) is another possibility to facilitate the characterization of agonist/antagonists of the receptor, and a Ca^{2+} signaling assay by using chimeric G protein has been described. Another signaling pathway is the mediating effect of 5-HT_6 receptors on K^+ channels.[94,95]

The modulation of 5-HT_6 receptor have an effect on the transmission of several neurotransmitters important in memory: acetylcholine and glutamate, as well as dopamine, γ-aminobutyric acid (GABA), epinephrine (E), and norepinephrine (NE). More 5-HT_6 antagonists have been developed to help to understand the relationship between 5-HT_6 system and memory consolidation in diverse learning paradigms.

The 5-HT_6 receptor is localized in the CNS, in an important area for learning and memory, and atypical antipsychotics and tricyclic antidepressants bind with high affinity to 5-HT_6 receptor. 5-HT_6 receptor antagonists increase neurotransmission at cholinergic and glutamatergic neurons. This fact seems to be useful for treatment of cognitive dysfunction.

Cognitive dysfunction is one of the specific symptom of dementia such as Alzheimer disease (AD), Parkinson-s disease (PD), Lewy-Body dementia (LBD) and cerebrovascular dementia (CVD) and schizophrenia.[96,97]

The first 5-HT_6 antagonists (SB-271046, phenyl–piperazine, Ro 63-0563, Ro-04-6790) were discovered in the late 1990's by high throughput screening (HTS) at Roche and GlaxoSmithKline. 5-HT_6 antagonists can be classified into four groups based on the typical pharmacophoric features: bisaryl sulfonamides, indoles, indole-like derivatives and non-sulfonyl compounds.[98] While AD is the major cause of dementia its mechanism is still poorly understood. By studying of the 5-HT_6 receptors a wide decline was found in the receptor expression in the cortex of AD patients.[99] About 470 papers have been published in the last ten years that attribute importance to receptors in learning and memory processes. From 1998 to 2011 there were 101 attempts to develop drugs targeting AD and only three new medicines were approved to treat the symptoms of the disease.[100]

These high numbers indicate that the effort towards studying potential 5-HT_6 antagonists is still a hot topic in development of CNS targeting agents, and this justifies our interest on this target.

2. Objectives, methods, experimental work

There general and specific objectives of the present work:

General objectives:

1. In general, testing and evaluating various integrated chemical biology model systems, which have an aim to indentify the interacting common portion of the chemical and biological space by combining *in silic*o selection of potential biological active compounds (generating target focused libraries) from huge databases with in-house developed *in vitro* biological assays for evaluating the biological activity.

2. Develop and improve 2D similarity search methods using available softwares and outside partners. Key drivers: effectiveness and the novelty. Integrating the simple and robust 2D similarity search with other approaches (3D alignment, pharmacophore modelling and search) in order to improve the hit rate and the novelty.

3. Develop rapid and reproducible biological assays for evaluating the focused libraries. Assay development and validation.

4. Analyzing the obtained biologically active compounds and the structural evolution of the *in silico* search pathway in order to draw conclusions how successful was the selection as well as what should be modified or changed in future *in silico* selection campaignes.

5. The obtained novel hit compounds can be starting points in future in-house drug discovery programs.

Specific objectives:

In order to investigate and achieve the general objectives two cases studies were selected and carried out: in the first case study we attempted to discover novel PDE-4 inhibitors combining 2D/3D similarity approaches, while in the 2nd case study novel 5-HT_6 antagonists were searched by integrating 2D similarity search and pharmacophore model building.

Objectives of the first case study:

1. Investigating and comparing the 2D and the 3D alignment similarity search for compounds obtained in a typical 2D similarity search

2. Identifying appropriate cut-off values for the 3D alignment similarity search.

3. Investigating the combination of the 2D and 3D similarity approaches by devising a combined 2D/3D fusion score

4. Evaluating the selection performance in *in vitro* measurements (hit rate, novelty, structure evolution)

Objectives of the 2nd case study:

1. Investigating and comparing the first round and second round 2D similarity search

2. Investigating and evaluating the performance of the second round 2D similarity search and the pharmacophore model based virtual screening using *in vitro* measurements (hit rate, novelty, structure evolution).

2.1. Methods using chemoinformatics programs

2D Similarity Search

We applied InstJChem software (ChemAxon Ltd., Budapest, Hungary) for 2D similarity search. InstJChem uses the Chemical Hashed Fingerprints for the 2D similarity search (as discussed earlier). Normally, in the initial similarity search phase a compound was defined as similar, if the Tanimoto coefficient was T2D $\geq$ 0.65 compared to any reference compound. In the following screening rounds (eg. in hit validation or hit refinement), we often applied higher Tanimoto similarity cut-off values.

The searchable drug-like chemical space, which is the major target of the similarity search, was composed by existing compounds: non-exclusive commercial libraries were available from the actual edition of the top vendor databases (~5 million compounds): ChemBridge, ChemDiv, Asinex, Enamine, IfLab, UkrOrg, AMRI, Specs, Maybridge, Interbioscreen etc.).

The biologically active chemical space is composed of known, biologically active reference (seed) compounds. Known PDE inhibitors and 5-HT_6 antagonists were collected from the available literature, PubChem and various commercial databases. Normally, we preferred if exact activity values were available for the compounds, particularly if the efficacy value was

below 1 μM, however, in certain cases compounds with unique structures without having specific activities were also included in the reference space in order to increase diversity.

Clustering

The compounds obtained in the 2D similarity search were in some cases clustered into groups (scaffolds/chemotypes) based on the chemical architecture (or structural similarity/dissimilarity) using Chemaxon's JKlustor program.

Calculation of the Physico-chemical Parameters and Property-based Filtering

The 2D similarity search can be refined using physico-chemical descriptors in property–based virtual screening. The physico-chemical parameters (Mwt, LogP, H-bond donors/acceptors, rotatable bonds and topological polar surface area) were determined by the Calculation Suit of InstJChem (ChemAxon Ltd.). The relevant physico-chemical parameters are described in the Literature section.

The property or parameter space of the known active compounds can be defined as the calculated min and max values, which were focused to the central 90 % of the range. removing 5-5 % from both the lowest (min) and the highest (max) values. Such focused property space was used for property-based filtering of the 2D similarity search results.

3D Ligand-Based Similarity Search for Ranking and Filtering of Focused Libraries

In order to increase the efficiency of the 2D ligand-based virtual screening we involved 3D similarity features in the first case study. We applied for 3D ligand-based similarity search flexible alignment through the Match algorithm of the Screen3D software (Screen3D 5.1, ChemAxon Ltd. Budapest). Match algorithm marks first specific atoms with predefined atom types and pharmacophore rules, which are followed by the calculations of the minimum and maximum possible intramolecular distances between every atom-atom pair in the molecule during a high-throughput dynamic, continual conformational scan. Intramolecular distances are then collected and distance range histograms were created. A single histogram represents the cumulative distance range distribution of the given type of atoms from the selected atoms. Then all the histograms (as "3D fingerprints") of two compounds are compared and a similarity score is calculated (3D Tanimoto), for further details see Kalászi et al.[78] In order to use the Screen3D software as a complement to the 2D virtual screening we decided to analyze first a dataset obtained during a previous PDE-5 *in*

silico selection campaign and and we generated the 3D similarity (flexible alignment) values (T3D) for a portion of a PDE-5 focused library for a comparison to the T2D similarity scores.

Diversity Selection

Besides the above methods, diversity selection was used reducing the virtual libraries to a reasonable and affordable size. For diversity selection Similarity Manager was used (CompuDrug Int., Sedona, AZ, USA).

Pharmacophore Model Building

The Phase module of the Schrodinger 2013 Suite was used for pharmacophore modelling and screening. The program provides six built-in types of pharmacophore features: H-bond acceptor (HBA), H-bond donor (HBD), hydrophobic (HYD), negative ionizable (NI), positive ionizable (PI), and aromatic ring (AR). and then the program automatically finds common pharmacophore features between the structures of the active molecules.

In pharmacophore model development 5-HT_6 receptors were previously extensively studied. Most of the major 5-HT_6 antagonists have distinct structural characteristics (Figure 6). Previously a 5-HT_6 pharmacophore model was reported by Lopez-Rodriguez et al.[101] arranging the these characteristic groups.

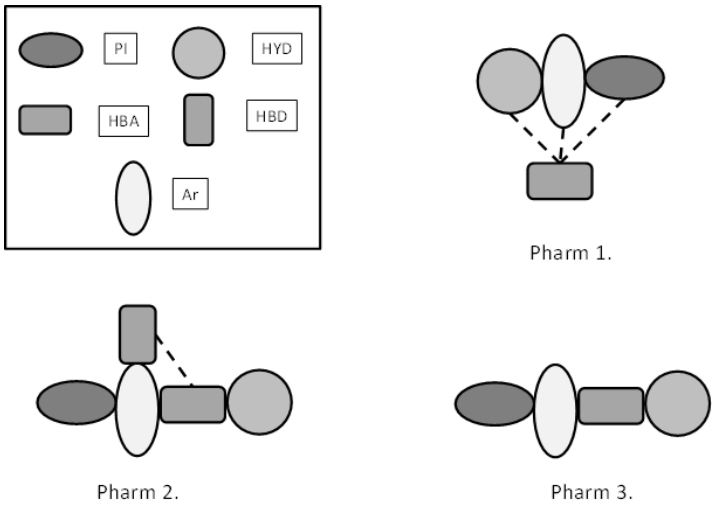


Figure. 6. Various group arrangements applied in 5-HT_6 pharmacophore models.

Based on the previous, reported study we attempted to build pharmacophore models applying a focused set of compounds as input structures. These structures derived from the seed compounds of the initial 2D similarity selection and first round biological screening. *(see*

Chapter 3.2.2. for details). The selected input structures were clustered using ChemAxon's JKlustor program into 3 groups based on similarities of the substructure (fragment) arrangement.

Cluster 1 contains aromatic/heteroaromic sulfonamides, where the N is embedded into a heteroaromatic ring; Cluster 2 contains aromatic/heteroaromic sulfonamides, where the N is coming from a primary amine; Cluster 3 contains aromatic/heteroaromic sulfones, where SO_2 makes a bridge between the rings.

The generated pharmacophore models are discussed under *Chapter 3.2.2.*

Finally, for pharmacophore-based virtual screening the ligand structures were prepared using Ligprep with default parameters. The conformers for each ligand were generated using Rapid Sampling with ConfGen using default parameters. The pharmacophore modeling was performed by Dr. Csaba Magyar (Inst. Enzymology).

3D Visualization, Ligand Docking and Pharmacophore Model Building

Schrodinger Suite 2013-3 package was used to demonstrate the primary interactions between the PDE-4B and the best hit compounds. The protein structures were prepared using the Protein Preparation Wizard of the Schrodinger Suite 2013-3 with the default settings. Default parameters were used in all Glide XP, Induced Fit Docking and Prime MM-GBSA ΔGbind calculations. The docking studies were perfomed by Dr. Csaba Magyar (Inst. Enzymology).

2.2. Assay development and biological activity screening

2.2.1. Phosphodiesterase-4/5

The PDE enzyme activity assay is described in detail for PDE-5A1, however, similar assays were applied for the other PDE isozymes.

PDE-5A1

The *in vitro* biological assay, we used, is based on PDE-5 cleavage of the phosphodiester bond of [^{3}H]cGMP resulting in [^{3}H]5'-GMP, which is further converted to [^{3}H]guanosine by snake venom nucleotidase. Unreacted cGMP is removed by solid phase extraction using an ion exchange resin. [^{3}H]guanosine content of the supernatant is determined by scintillation counting. The resulting signal is proportional to the amount of the cleaved cGMP. A recombinant full length PDE-5A1 enzyme was used for the assay

development. It was expressed in baculovirus-Sf9 system. Dypridamole, Zaprinast and Sildenafil were used as inhibitors to validate the assay.

The PDE enzyme activity was determined using a modified method.[102] The enzyme was incubated for 20 min at 30 ºC in 20 mM Tris, pH 7.4, 5 mM $MgCl_2$, 0.15 μCi [^{3}H]cGMP (PerkinElmerNET337) and 0.1 μM cGMP(SigmaA6885) in the presence of Sildenafil (Sigma P3510) or the test compound in 100 μL volume. After incubation at 70 ºC for 2 min, samples were cooled on ice for 10 minutes. Then 25 μl of Crotalus atrox snake venom (Sigma V7000, 1 mg/mL) was added to each assay point and incubated at 30 ºC for 10 minutes. Then 200 μL of a 1:1:1 (v/v/v) Dowex 1X8 200–400 MESH CI resin (Sigma 44340), ethanol, and water were added, samples were shaken at RT for 20min and centrifuged at 1,000×*g* for 3 minutes. 50 μL of the supernatant was added to 200 μL Microscint 20 (PerkinElmer) and counted in a Wallac Microbeta1450 Trilux scintillation counter (PerkinElmer). The cyclic nucleotide concentration in the samples was below the K_m of the substrate for the enzyme. PDE-5A inhibitors were first tested at 10 μM concentration as a single point screening. Compounds that showed ≥ 55% inhibiton were selected for dose-response curves and IC_{50}s determination. The used concentration range of the compounds was from 5 to 100,000 nM using 10 different concentrations. Parallel measurement points were used for the assay. IC_{50}s were calculated using Origin7 (OriginLab) from the 10 point dose–response curve.

PDE-4B2

The *in vitro* biological assay is almost the same what we used for PDE-5A with the significant difference, that the substrate is [^{3}H]cAMP (substrate concentration was 1 μM) instead of [^{3}H]cGMP. A known PDE-4 inhibitor (3,5-dimethyl-1-(3-nitrophenyl)-1H-pyrazole-4-carboxylic acid ethyl ester) was used to validate the assay. In this case ≥ 47 % inhibition at 10 μM concentration was used as a selection criteria for the IC_{50} determination.

Determining PDE selectivity

For selectivity measurements PDE-2A3, PDE-3A, PDE-4D4, PDE-7A1, PDE-8B1, PDE-9A1, PDE-10A1 and PDE-11A4 were cloned and expressed in Sf9 cells similar to PDE-5 and PDE-4B2. In case of PDE-3, PDE-7, PDE-8 and PDE-4D4 the substrate was [^{3}H]cAMP and for PDE-2, PDE-9, PDE-10, and PDE-11 we used [^{3}H]cGMP as substrate, which were used universally at 1 μM concentration.

The following known inhibitors were used in the assays: BAY-60-7550 (PDE-2), trequinsin hydrochloride (PDE-3, PDE-8), BRL 50481 (PDE-7), tadalafil (PDE-5, PDE-11), BAY-73-6691 (PDE-9) and papaverine hydrochloride (PDE-10).

2.2.2. 5-Hydroxytryptamine receptor 6

CHO-K1 cells expressing stably the mitochondrially targeted aequorin (luminescent indicator) and Gα16 were used for the assay. The principle of the assay is that aequorin (derived from *Aequorea victoria*) complex emits blue light while binding Ca^{2+} ions.[103,104,105]
The aequorin and Gα16 expressing CHO-K1 cells were transiently transfected with plasmid harboring the gene expressing human 5-HT_6 using Roche X-treme GENE HP DNA Transfection Reagent. Cells were grown for 48 hours after transient transfection. One day before the assay the culture medium was changed to antibiotics free medium and the cells were grown for an additional 6 hours, then the cells were detached by gentle flushing with PBS/0.5 mM EDTA, recovered by centrifugation and resuspended at 1x106 cells/mL density in assay medium (DMEM/HAM's F12 D6434 with HEPES, without phenol red + 0.1% BSA + 2 mM glutamine) in a Falcon tube. Coelenterazine H (InVitrogen C6780) was added at a final concentration of 5 μM and the cells were incubated overnight at room temperature using constant shaking, protected from light. Before measurements the cells were diluted 2 fold in assay medium and incubated for 60 min.
The measurement was executed in Optiplate™-96 plates (PerkinElmer 6005290) and the luminescent emission was detected by an Appliskan™ (Thermo) plate-reader. Cell suspension (45μL)/well + 5 μl antagonist (diluted in assay medium) was preincubated in the plate, the reaction was initiated with the addition of agonist EMD 386088 (50 μL) at a concentration of 20 nM.[106]
The time lapse curves of receptor activation signals were recorded well by well. After detection of the baseline (8 sec) the agonist was injected in one well and the change of intracellular Ca^{2+} level released due to receptor activation was monitored for 45 seconds. Peak luminescence signal was used for evaluation of the measurements. For assay development and validation SB-271046 was used as antagonist control. All compounds were screened at a fixed concentration of 10 μM, followed by a 8-point dose response analysis.

3. Results and discussion

3.1. Discovery of potential PDE-4B inhibitors

One of the specific objectives of the present work was to improve the 2D ligand-based virtual screening by intergrating with additional methods. One approach is to combine 2D similarity search with 3D shape and conformational similarity of the compounds that might provide a more realistic picture of the similarity compared with the purely 2D structural connectivity. We analyzed the correlation between 2D similarity and 3D alignment to understand the relationship between the 2 approaches and draw conclusions for further use. Fortunately, a previous dataset was available from a PDE-5 inhibitor virtual and *in vitro* screening campaign[51], therefore, this dataset provided an excellent opportunity to analyze the 2D and 3D similarity of the focused library selected by 2D similarity search method as well as correlations of the identified hit compounds with the 3D similarities.

3.1.1. 2D/3D similarity correlation analysis of previously generated PDE-5 focused libraries

In the correlation study first we involved 5 PDE-5 inhibitor hit compounds (IC_{50} < 10 μM, Table 1), we identified in the previous study,**[51]** that were selected based on their 2D similarity to 3 seed compounds (#13-PDE-5[a], #18-PDE-5, #44-PDE-5, Table 1. column "Seeds"). (Note: the cut-off value was T2D=0.6 for the 2D similarity selection). Thus, we first generated their 3D similarity values (using Screen3D) towards the same seed compounds.

[a] Note: the numbering of the chemical structures correspond to the appropriate section and topic. In order to avoid any confusion for citing the compounds the numbers in the text has a post-fix indicating the particular topic (e.g. #X-PDE-5; #Y-PDE-4; #Z-5-HT_6).

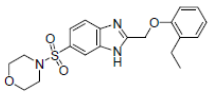
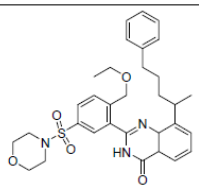
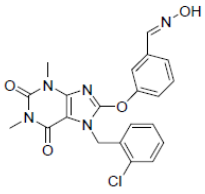
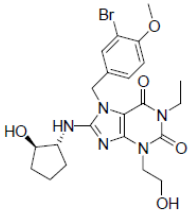
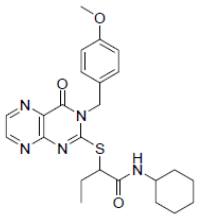
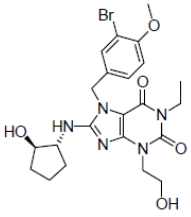
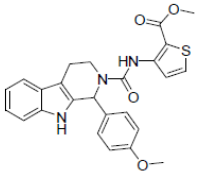
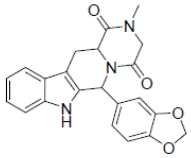
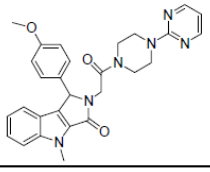
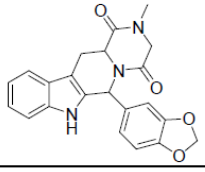
ID	Hits	IC_{50} (μM)	T2D	T3D	Seeds	Seed ID
1		3.65	0.63	0.47		13
2		1.99	0.61	0.38		18
3		7.24	0.61	0.31		18
4		0.19	0.69	0.35		44
5		6.74	0.67	0.38		44

Table 1. The generated 3D values from the 5 PDE-5 inhibitor hit compounds

It was not surprising that the T3D values are much lower than the T2D values.

In order to extend the study we calculated the 3D similarity values for all the compounds[71] we obtained from the 2D similarity selection to the 3 seed compounds with T2D ≥ 0.6 selectivities and the 2D/3D correlation is shown in Figure 7.

We found that while the 2D selection criteria was T2D ≥ 0.60, thus, the T2D values were above this value, the 3D Tanimoto scores (T3D) were above 0.2. Since the T3D values for the 5 hits were above 0.3, we indicated this value on the plot, even though many „2D similar" compounds fell into the 0.2-0.3 category.

We could conclude that even though, the compounds selected by 2D methods had similar molecular architectures (atomic connectivity), they are rather different in terms of shape and conformational flexibility, therefore, the lower T3D region suggests, that such 3D „dissimilarity" represents significantly different binding characteristics.

It is also interesting to correlate the 3D similarity values with the biological activity of the same cluster (41 compounds). While the 5 identified hits has T3D values higher than 0.3,

there are lots of 2D similar compounds that has 3D similarity values above 0.3 but they are not hits. We could call them as 'virtual false positives' (45% of all compounds measured). (Note: Hit compounds were defined if they exhibited 57% PDE-5 inhibitory activity at 10 μM concentration - this inhibition % value provided a reasonable number compounds for concentration dependency studies).

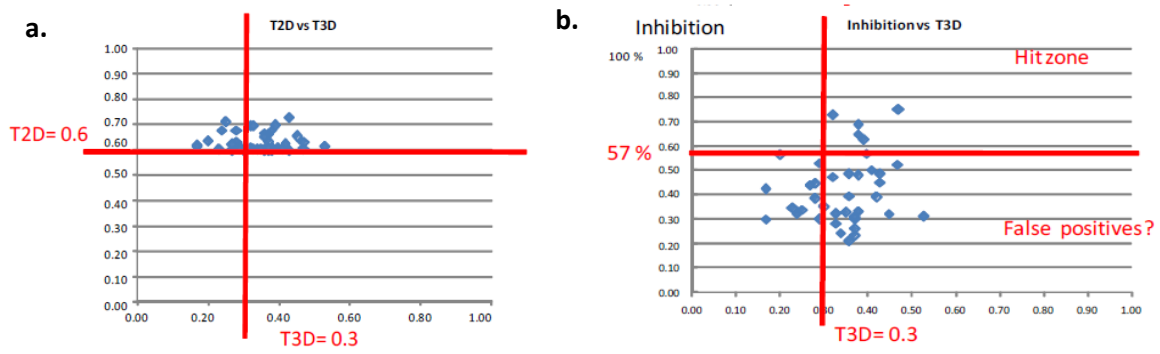


Figure 7a-b shows the scatter plot of correlating the 2D and 3D similarity scores of all the measured compounds (41, active and not active) derived from the three reference compounds (#13-PDE-5, #18-PDE-5, #44-PDE-5) in the first screening round. The five hits had T3D values > 0.3. In **Figure 7b** PDE-5 inhibition % values were correlated with the T3D values.

After the first round selection and screening we had the conclusion that T3D = 0.3 might have been a useful cut-off value even if it might result in numerous virtual false positives.

In a further study we analyzed the second round screening dataset (104 compounds), which were generated by 2D similarity search around the 5 first round screening hits using a higher selection criteria (T2D ≥ 0.8). We again calculated the 3D similarities for all the 2D selected compounds. While the T2D cut-off value was 0.8, the corresponding T3D values were between 0.35 and 0.95. The 2D/3D correlation is shown in **Figure 8**. There is a certain trend that with increasing 2D values the 3D values are also increasing but in a lower extent. Interestingly, the majority of the compounds have T3D values above 0.6.

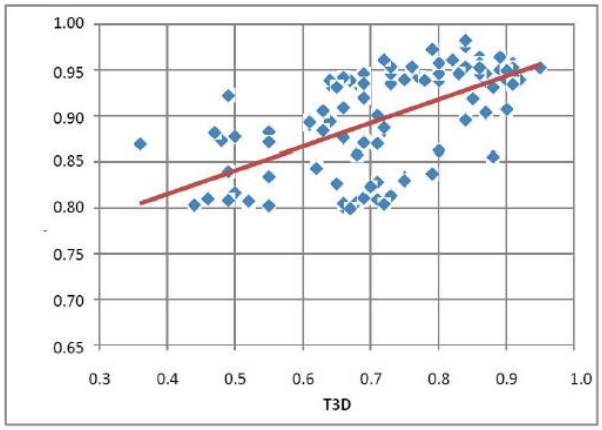


Figure 8. Correlation between T2D and T3D values of the compounds (104) selected for the second round screening

Most of the second round hit compounds (13 out of 20) derived from two PDE-5 reference compounds (#18-PDE-5, #44-PDE-5). These two series of compounds were analyzed separately.

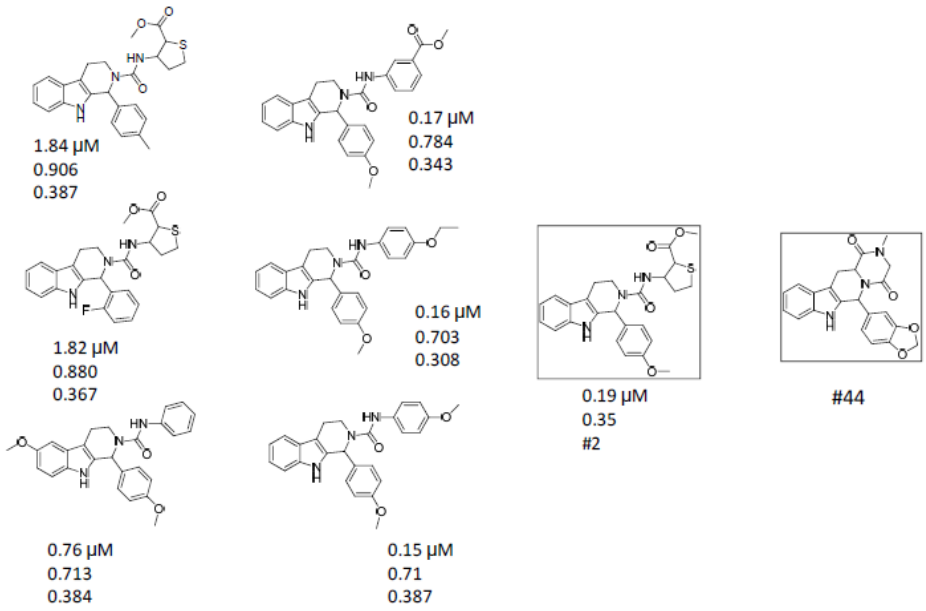


Figure 9. Second round hits derived from #44-PDE-5 seed and #2-PDE-5 first round hit with the T3D similarities to the first round hit and seed (reference compound)—Series #2-PDE-5.

Figures 9 and 10. represent the 3D similarities between the 2nd round hits (hit criteria: $IC_{50} \leq 2$ µM) and the first round hits as well as the reference compounds. PDE-5 inhibition (IC_{50} values) of compounds derived from #2-PDE-5 and #44-PDE-5 are shown in Figure 9. (upper value); the T3D values to #2-PDE-5 (in the middle) and the T3D values to the reference active compounds #44-PDE-5 (lower value).

Similarly, Figure 12 shows the IC_{50} values of compounds derived from #3-PDE-5 and #18-PDE-5 (upper value); the T3D values to #3-PDE-5 (middle) and the T3D values to the reference compounds #18-PDE-5 (lower value).

(Note: 2D similarity is over 0.8 for all the compounds according to the selection criteria).

The 3D similarities of the second rounds hits towards the corresponding first round hits were between 0.7 and 0.91 (#2-PDE-5/#44-PDE-5 series, Series #2-PDE-5, Figure 9); while between 0.47 and 0.69 (#3-PDE-5/#18-PDE-5 series, Series #3-PDE-5, Figure 10). In addition, the 3D similarity between the second round hits and the corresponding original seed compounds showed also differences in the two series; analogues derived originally from #44-PDE-5 seeds have higher T3D scores (between 0.31 and 0.39, Figure 9), while those compounds that derived from #18-PDE-5 seeds showed lower 3D similarity values (between 0.22 and 0.34, Figure 10). (Chapter 3.1.2.2., Table 2 and 3).

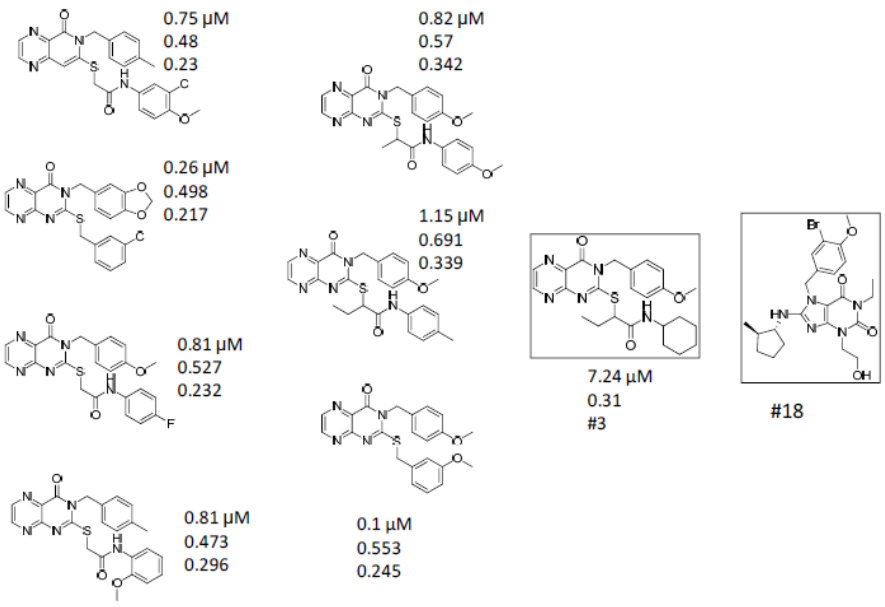


Figure 10. Second round hits derived from #18-PDE-5 seed and #3-PDE-5 first round hit with the T3D similarities to the first round hit and seed (reference compound)—Series #3-PDE-5.

While the T3D values were rather low in the case of Series #3-PDE-5, significant improvement can be observed in the inhibitory activities (from 7.24 µM to 0.1 µM, Figure 10). Consequently, such analogs could have significantly different binding characteristics than the first round hits they derived from. It is also a possible conclusion that the activity can significantly increased by replacing the saturated ring (cyclohexyl) in Series #3-PDE-5 with aromatic groups.

As a result of the above analysis we got to the conclusion that T3D values are much more sensitive measures than their T2D counterparts, and they reflect important structural features (e.g. conformational flexibility) that strongly depend on the structure. We could also make the statement that T3D = 0.3 would be a good cut-off value in the first round similarity selection and T3D = 0.5 or 0.6 in the second round (hit validation). We could get probably higher hit rate applying stricter filters, but then some compounds (chemotypes) would have been lost.
Furthermore, if we attempted to combine the T2D/T3D similarities into a single fusion score and the T2D + T3D values would range between 1.28 (0.8 + 0.47) and 1.71 (0.8 + 0.91) based on the similarity scores calculated for the 2nd round hits towards the first round hits. (Note: 0.47 was the lowest T3D value and 0.91 was the highest T3D value within the two analyzed sets: Series #2-PDE-5 and Series #3-PDE-5).
Summarising the above findings, we concluded that combination of the 2D similarity search with 3D similarity measures might be a useful and applicable approach in ligand-based virtual screening and we further refined this strategy in the selection of a PDE-4 focused library.

3.1.2. *In silico* PDE-4B focused library selection and its biological evaluation:

3.1.2.1. First round selection and screening

First, the typical 2D similarity search was carried out using 44 known PDE-4 inhibitors as seed compounds and the 5 M compound vendor repository as the drug-like chemical space. We obtained 4833 compounds, when the Tanimoto cut-off value was set as T2D $\geq$ 0.65. We calculated the property (parameter) space for the 44 known PDE-4 inhibitors and applied this property space as a filter, which allowed to reduce the number of the virtual hits to 1764.
Finally, we selected 200 compounds using simple diversity selection. After visual checking of the compound set we sent an request for 120 compounds, and we ordered the available compounds (105). *In vitro* measurements were done at 10 μM concentration, and we experienced inhibition (>47%) in case of nine compounds (Figure 11).
These compounds were selected for measuring the concentration dependency determining the IC_{50} values. Finally, seven compounds were considered as hits (IC_{50} values range from 0.05 to 16 μM). The hit rate of the first round screening was 6.6% (>47% inhibition) and the success rate was 2.8% if we counted only those hits that have IC_{50} values below 2 μM. (Note: 47%

inhibition was chosen instead of 50% to involve more compounds into the IC_{50} determination). Table 2-4 shows the first round screening results.

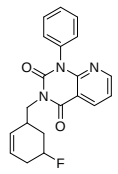
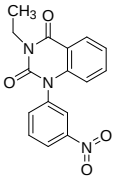
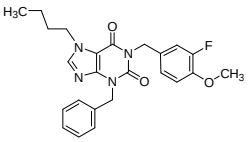
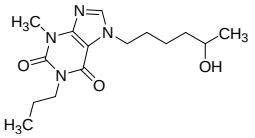
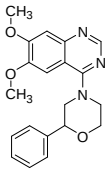
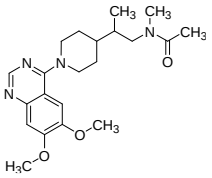
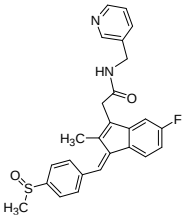
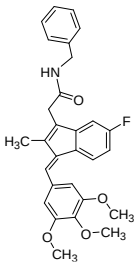
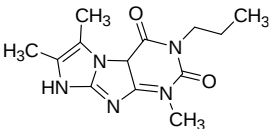
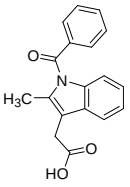
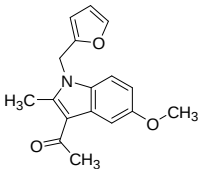
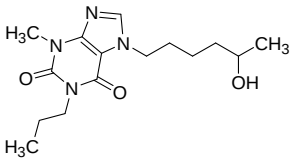
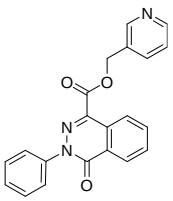
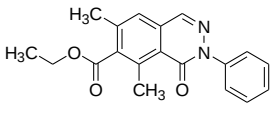
ID	Structure	inhibition	IC_{50} µM	T2D	Seed_ ID	Seed
2		0.92	0.053	0.66	38	
3		0.65	1.05	0.71	13	
5		0.71	1.97	0.7	42	
4		0.7	6.05	0.75	29	
1		0.47	11.05	0.72	13	
6		0.56	13.79	0.66	14	
7		0.6	16.5	0.77	24	

Table 2. 7 hits identified in the first screening round

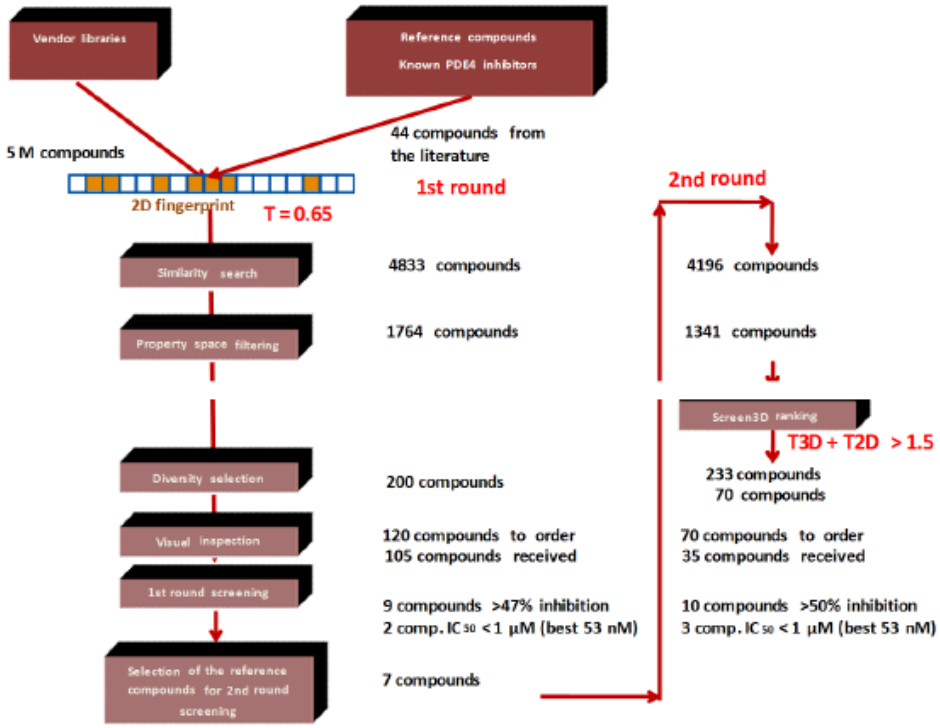


Figure 11. *In silico* selection scheme of the PDE-4B focused library

3.1.2.2. Second round library selection and screening. Generation of a fusion score combining 2D/3D similarities

For the second round library selection (hit validation) the seed input structures were the seven first round hits. First, a standard 2D similarity search was executed because Screen3D could only handle a couple of thousand compounds rather than a couple of million. After property space filtering 1341 compounds were obtained (T2D > 0.65). (Note: here we applied T2D ≥ 0.65 instead of 0.8 what we normally use in the second round selection in order to allow representing a bigger chemical space and giving a better chance for the 3D similarity (flexible alignment) to contribute.

In the next step we calculated the 3D similarity values for 1341 compounds towards their corresponding first round hits. Their correlation to the 2D scores is shown in Figure 12. We found a similar trend as with PDE-5 the 3D Tanimoto coefficient (T3D) ranged from 0.19 to 0.99 while the 2D cut-off value was 0.65. A weak linear correlation could be recognized again between the 2D and 3D similarity measures.

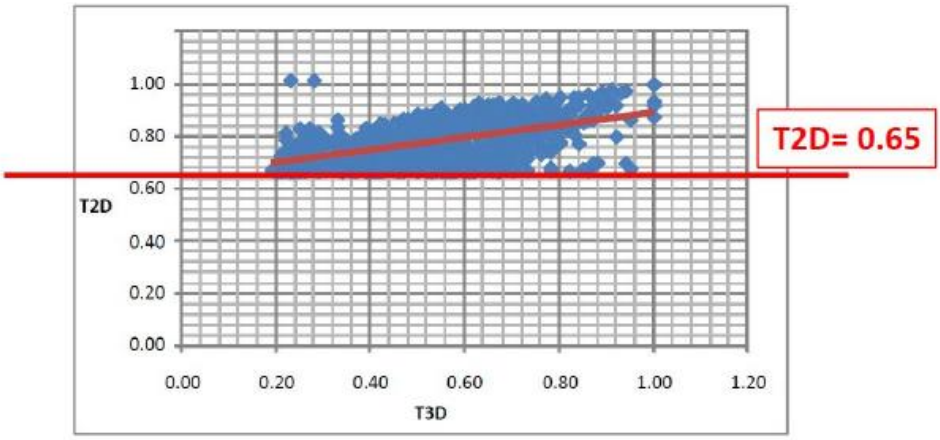
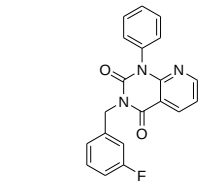
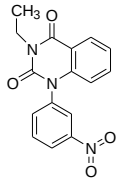
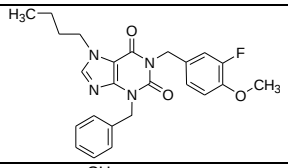
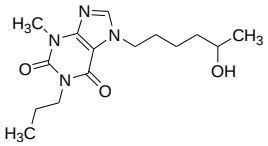
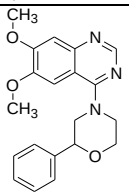
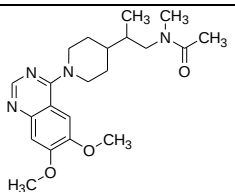
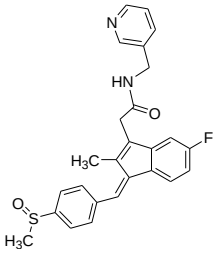
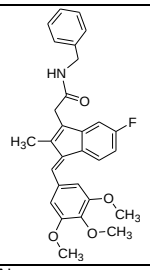
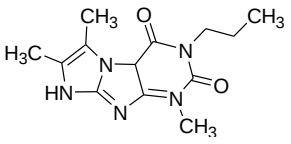
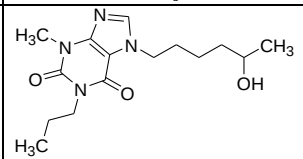
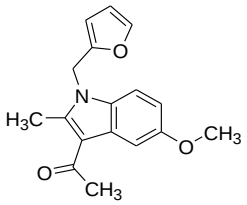
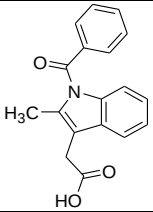


Figure 12. Correlation between the T2D and T3D values

ID	Structure	Inhibition	IC_{50} µM	T2D	Seed_ ID	Seed
2		0.92	0.053	0.66	38	
3		0.65	1.05	0.71	13	
5		0.71	1.97	0.7	42	
4		0.7	6.05	0.75	29	
1		0.47	11.05	0.72	13	
6		0.56	13.79	0.66	14	

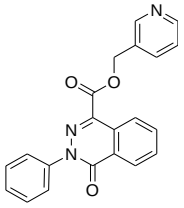
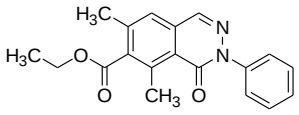
7		0.6	16.05	0.77	24	

Table 3. The result of the first screening round were 7 hits

After calculating the T3D values we decided not to use a single cut-off values (e.g. T3D > 0.3) as previously proposed, but we attempted to combine the two scores creating a more appropriate filtering tool and giving an opportunity to identify which similarity (2D or 3D) is more relevant for leading to active compounds.

Therefore, we decided to use a fusion score which handles the influence of T2D and T3D equally and aggregates the scores. A suitable cut-off value was applied as T2D + T3D ≥ 1.5 based on the previous analysis of the PDE-5 data as well as the following considerations. Actually, 1.5 fusion score is about the mean value of the calculated fusion scores (1.28 and 1.71) we deduced from the two series of PDE-5 we analyzed before. (Series #2-PDE-and Series #3-PDE-5, see Section 3.1.1.) The only difference is that we went down with the T2D cut-off value to T2D ≥ 0.65).

We had the following considerations as feasable scenarios:

(1). Since T2D was ≥ 0.65 (cut-off value) at such relatively low 2D similarity level only those compounds would remain in the selection, where the T3D similarity is high (> 0.85).

(2) Inversely, in the case of high T2D values (max. 0.96), compounds could be selected even in the case of lower T3D values (min. 0.54), which allows to select close 2D analogues. This supports one of the objectives of the second round screening since hit validation requires testing relatively close analogues.

(3) Our preference to equal weight of T2D and T3D scores can be similarly explained.

Applying this combined cut-off value (T2D + T3D ≥ 1.5) resulted in a relatively small number of compounds. In this way the library size was reduced from 1341 to 233. This level of size reduction is similar what we obtained previously in the first round screening by diversity selection. (Figure 13a)

The representation of the compounds derived from the seven first round hits varied significantly showing that the T3D value range and its distribution strongly depend on the structures (chemotypes) (Table 3). For example, more than 40% of the 233-membered set represented analogues that derived from #4-PDE-4 and #1-PDE-4, while other analogue

groups were underrepresented (#3-PDE-4 and #6-PDE-4). On the other hand, 90 compounds were excluded from the selection that exceeded T3D = 0.65, due to the lower than 1.5 cumulative score, which means T2D scores were less than 0.85).

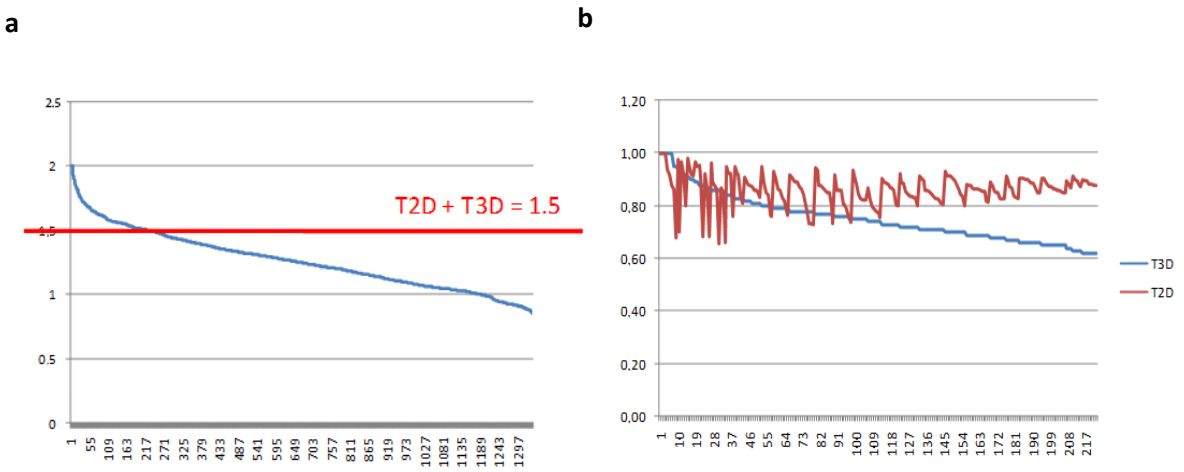


Figure 13a-b. Representation of the hit analogues.

After the 2D/3D combination filtering, which resulted in 233 compounds, the library was reduced to 70 by diversity selection and this compound set was ordered. Unfortunately, only 35 compounds were available; however, the obtained compound set reflected the distribution of the initial set of compounds (233). In general, the contribution of T2D is gradually increasing but fluctuating between 0.65 and 1, while T3D is decreasing from 1 to 0.6 among the compounds that have a fusion score of 1.5. (Figure 13b).

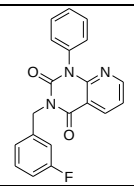
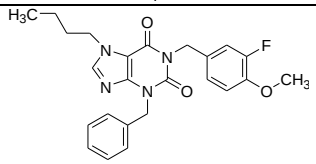
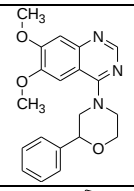
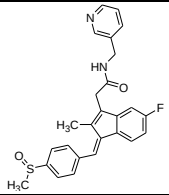
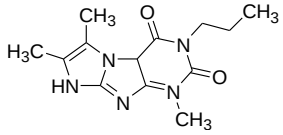
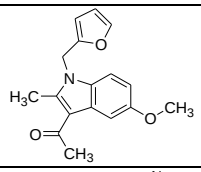
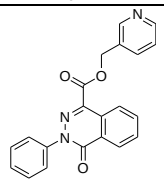
ID	Structure	Analogues in the 1341 Set	Lowest T3D	Analogues in the 233 Set	Lowest T3D	Percentage in the 233 Set
2		155	0.24	30	0.69	19.35
3		476	0.19	16	0.63	3.36
5		27	0.33	9	1.00	33.33
4		13	0.28	6	0.72	46.15
1		273	0.30	116	0.62	42.49
6		117	0.29	12	0.70	10.26
7		280	0.28	44	0.64	15.71

Table 4. Distribution of the T3D values in the entire second round selection library

In vitro screening of the small focused library resulted in 10 hits (IC_{50}= 0.053–3.2 μM, Table 4). The hit rate of the 2nd round screening was 28.5% (>50% inhibition, at 10 μM), while 10% hit rate was calculated if only those compounds were considered that had IC_{50} values ≤ 2 μM (seven compounds).

We also analyzed the T2D/T3D correlation of the hits. Such correlation and fusion score analysis applied on the 10 second round hits revealed that six out of the 10 hits have a fusion score around 1.5. which is close to the cut-off value (Figure 14).

It means that T3D values are apparently lower in all cases except one where T2/T3 ratio is < 1, and the T2D is around 0.8. (Table 5). Apparantly, the fusion score prefers this scenario and the selection leads to relatively close analogues (keeping the same chemotype) and helping to exclude false negatives, where the T2D is high but the T3D score is low.

This finding suggests that the fusion score approach is valuable in the hit validation stage particularly within the preferred chemotype.

Of course, this finding raises the question if the fusion score would have been decreased to 1.3, which would allow the involvement of more compounds with lower T2D or T3D values, would it lead to more hits or hits with higher frequency of novel chemotypes?

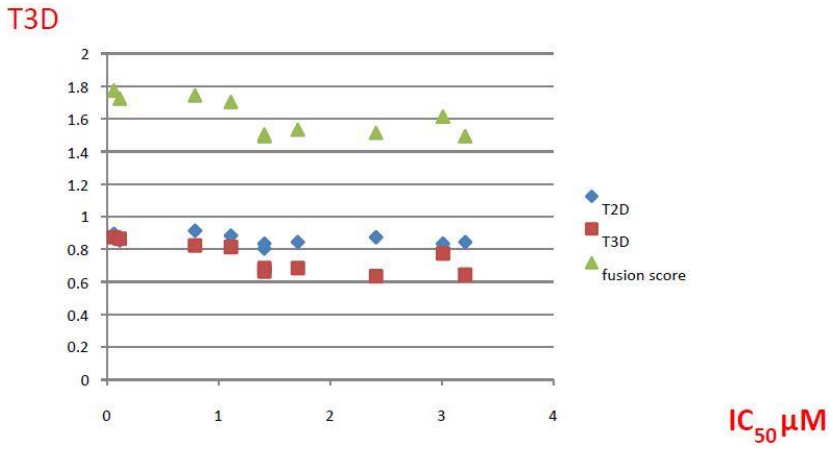
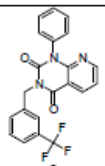
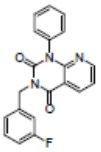
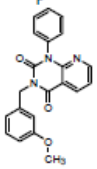
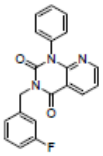
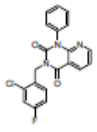
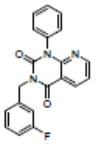
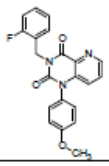
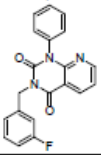


Figure 14. T2D/T3D and fusion score analysis of the 10 second round hits

ID	Structure	Inhibition at 10 µM (%)	IC50 (µM)	T2D	T3D	Fusion Score	T2D/T3D	First Round Hit_ID	First Round Hits	Chemo Type
3		83	0.053	0.9	0.88	1.78	1.02	2		A2
5		94	0.105	0.86	0.87	1.73	0.99	2		A2
4		90	0.779	0.92	0.83	1.75	1.1	2		A2
6		88	1.1	0.89	0.82	1.71	1.09	2		A2

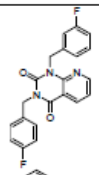
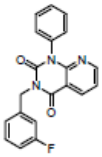
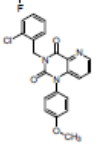
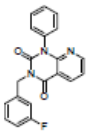
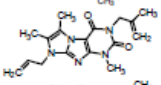
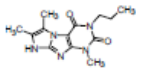
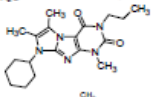
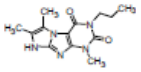
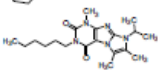
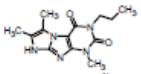
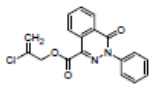
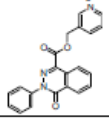
ID	Structure	Inhibition at 10 μM (%)	IC50 (μM)	T2D	T3D	Fusion Score	T2D/T3D	First Round Hit_ID	First Round Hits	Chemo Type
8		86	1.4	0.81	0.69	1.50	1.18	2		A2
7		89	3.0	0.84	0.78	1.62	1.08	2		A2
1		51	1.7	0.85	0.69	1.54	1.24	1		B1
2		61	1.4	0.84	0.67	1.51	1.25	1		B1
9		62	3.2	0.85	0.65	1.5	1.31	1		B1
10		62	2.4	0.88	0.64	1.52	1.38	7		A11

Table 5. 10 hits obtained in the second screening round

3.1.2.3. Similarity and novelty in the 2D similarity search towards identifying, novel PDE-4 inhibitors

Looking at the structural relationship of the first round hits and their seeds we could identify several pairs, where the core structure was retained (3 → 13, 5 → 42, 4 → 29, 6 → 14, 7 → 24, Table 4, Figure 15) together with few cases, where we could recognize some sort of scaffold hopping or at least modified scaffolds (1 → 13, 2 → 38).

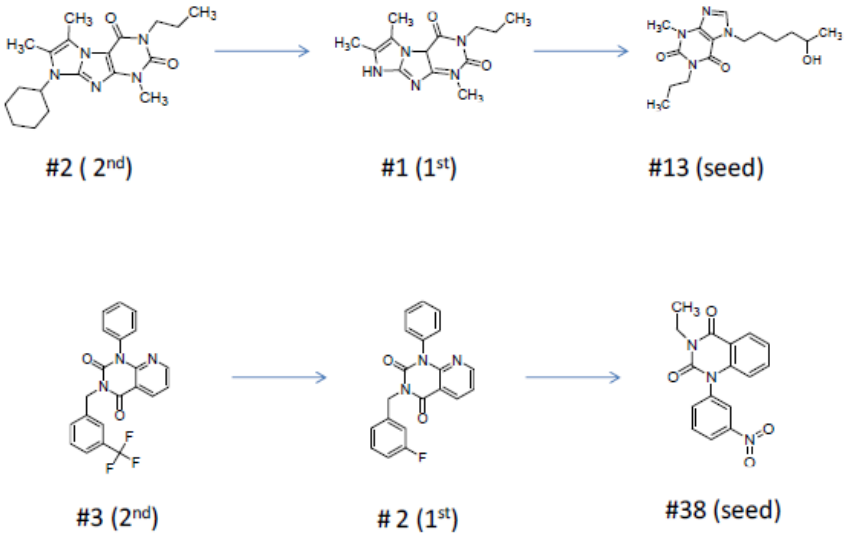


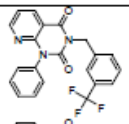
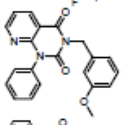
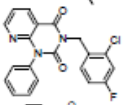
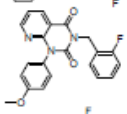
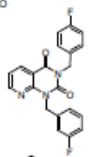
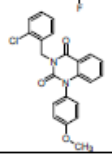
Figure 15. The hit evolution sequence of two second round hit compounds that involved scaffold hopping. #2-PDE-4 (2^{nd}) → #1-PDE-4 (1^{st}) → #1-PDE-4 3 (seed) (upper scheme) and #3-PDE-4 (2^{nd}) → #2-PDE-4 (1^{st}) → #38-PDE-4 (seed) (lower scheme)

We searched PubChem for such modified chemotypes, whether they have already been reported as potential PDE inhibitors. We found an analogue for hit #2-1st-PDE-4 (1st round, 53 nM) what we might have missed during the search for reference compounds (Figure 15, same chemotype with different substitution pattern), however, the reported activity of such analogue was only 950 nM).[107]

Furthermore, for chemotypes that represented in hits #1-1st-PDE-4 (1st round, 11 μM) and #2-2nd-PDE-4 (2nd round, 1.4 μM) we did not find any biological data in PubChem, the closest analogue exhibited biological activity as an adenosine A(1) and A(2A) antagonistic effect.[108]

3.1.2.4. Selectivity profiling of the second round hits

We found that most of the hit compounds showed some selectivity towards PDE-4B and in some cases the selectivity was surprisingly high (50–80 fold) over PDE-4D (Table 6). Some of the hits inhibited PDE-5 at low concentration, which could be useful in the therapy while PDE-10 inhibition could be avoided, since that target is more implicated in CNS disorders.

ID	Structure	PDE4B inh.% (10 μM)	PDE4B IC50 (μM)	PDE4D IC50 (μM)	PDE5 IC50 (μM)	PDE2 IC50 (μM)	PDE3 IC50 (μM)	PDE7 IC50 (μM)	PDE8 IC50 (μM)	PDE9 IC50 (μM)	PDE10A IC50 (μM)	PDE11 IC50 (μM)
3		83	0.053	35.1	n.i.	n.i.	n.i.	n.i.	n.i.	n.i.	n.i.	n.i.
5		94	0.105	1.12	n.i.	n.i.	n.i.	n.i.	n.i.	n.i.	n.i.	n.i.
4		90	0.779	7.8	n.i.	n.i.	n.i.	n.i.	n.i.	n.i.	n.i.	n.i.
6		88	1.1	10.7	n.i.	n.i.	n.i.	n.i.	n.i.	n.i.	n.i.	n.i.
8		86	1.4	7.5	5.8	n.i.	n.i.	n.i.	n.i.	n.i.	n.i.	n.i.
7		89	3.0	8	60	n.i.	n.i.	n.i.	n.i.	n.i.	n.i.	n.i.

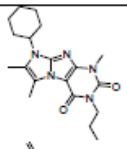
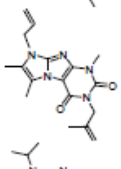
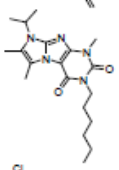
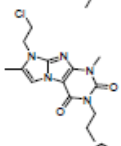
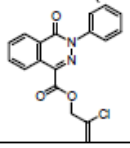
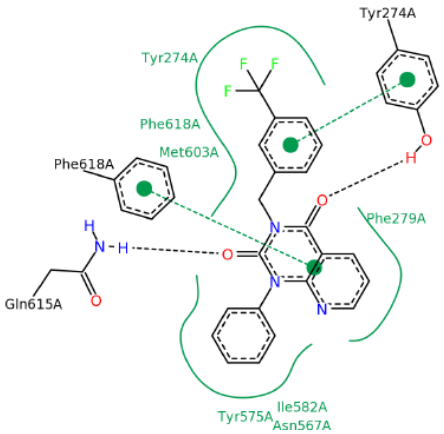
ID	Structure	PDE4B inh.% (10 µM)	PDE4B IC50 (µM)	PDE4D IC50 (µM)	PDE5 IC50 (µM)	PDE2 IC50 (µM)	PDE3 IC50 (µM)	PDE7 IC50 (µM)	PDE8 IC50 (µM)	PDE9 IC50 (µM)	PDE10A IC50 (µM)	PDE11 IC50 (µM)
2		61	1.4	24	n.i.	n.i.	n.i.	n.i.	n.i.	n.i.	40.8	n.i.
1		51	1.7	41	11.7	n.i.	n.i.	n.i.	n.i.	n.i.	2.4	40
9		62	3.2	n.i.	12.5	n.i.	n.i.	n.i.	n.i.	n.i.	0.6	n.i.
11		62	4.7	350	n.i.	n.i.	n.i.	n.i.	n.i.	n.i.	8.5	n.i.
10		62	2.4	13.5	32.8	n.i.	n.i.	n.i.	n.i.	n.i.	n.i.	n.i.

Table 6. The activity profile of 11 selected hit compounds (n.i. means no inhibition)

The most interesting finding was that #3-2nd-PDE-4 (2nd round) hit shows a high selectivity towards the closely related PDE-4D isozyme. Docking and interaction mapping of this compound revealed a possible explanation (performed by Csaba Magyar, Inst. Enzymology).

a



b

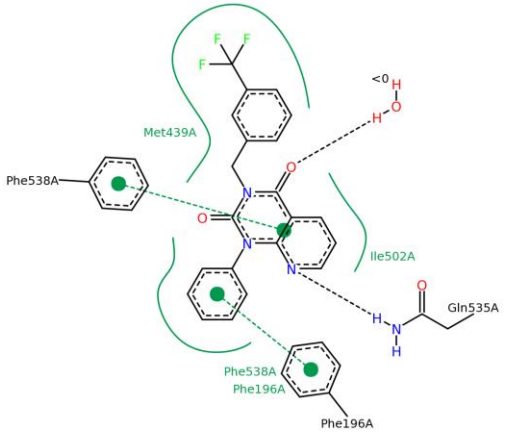


Figure 16a-b. Diagram of the most important interactions in the Glide XP calculated PDE-4B--#3-2nd-PDE-4 and PDE-4D--#3-2nd-PDE-4 complex identified by PoseView. Black dashed lines represent hydrogen bonds, green solid lines represent hydrophobic interactions and green dashed lines represent aromatic interactions.

In the case of ligand #3-2nd-PDE-4 in Figure 16a we can see that the Tyr^{274} residue(according to the PDB numbering), belonging to the UCR2 domain of PDE-4B, does not only contribute to the ligand binding through aromatic interaction but it also participates in the formation of a hydrogen bond. We docked ligand #3-2nd-PDE-4 into the 3G4G PDE-4D structure. The PoseView diagram of this complex can be seen in Figure 16b. The equivalent of Tyr^{274} is Phe^{196} residue in the PDE-4D structure, which cannot participate in the formation of a hydrogen bond. This additional H-bond might explain the high PDE-4B specificity of ligand #3-2nd-PDE-4(2nd round hit, IC_{50}, PDE-4B = 0.053 μM, IC_{50}, PDE-4D = 35.1 μM).

3.1.2.5. Observations regarding structure-activity relationships for the 1*H*-purine-2,6-diones chemotype

For the most common chemotype in our study (1*H*-purine-2,6-diones) we analyzed the various substitutions around the core in order to draw some conclusion regarding the structure—activity relationships. It is obvious that meta-substituted benzyls at R_1 position are favored together with unsubstituted phenyl groups at R_2 (Figure 17). 1*H*-purine-2,6-diones hits show good selectivity towards PDE-4D and other isozymes, except if R_2 is a benzyl group instead of phenyl, which causes reasonable PDE-5 inhibition (IC_{50} = 5.8 μM).

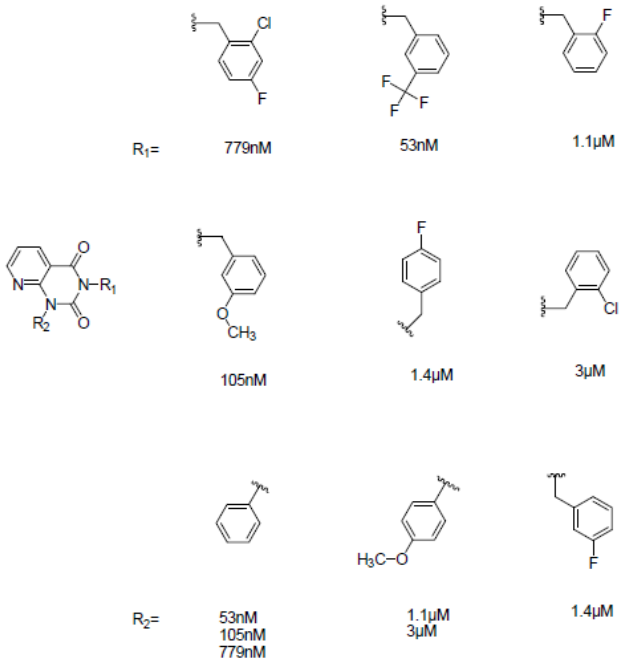


Figure 17. Preliminary structure-activity relationships of the most common chemotype of the 2nd round screening

3.1.3. Conclusion of the PDE-4 study

In conclusion, 3D shape-/comformational flexibility-based similarity search (Screen3D) is a sensitive and fast approach to select potentially active compounds. It can be used either independently or in combination with 2D similarity search (depending on the size of the library). Screen 3D is useful for selecting compounds up to 5000, but it is not practical if multi-million compounds could be virtually screened for 3D similarity. This method is particularly useful for filtering the 2D virtually prescreened libraries. While 2D similarity remains the method of choice due to its robustness it can be combined with 3D similarity measure. We proposed and applied a combination or fusion (2D/3D) score and a cut-off value. The novel fusion score approach led to improved hit rate in the second round (hit validation) screening phase. The improvement of the hit rate could be attributed to elimination of false positive 2D virtual hits (relatively high T3D values, e.g. > 0.6-0.65 if T2D is > 0.8). Since T3D is rather chemotype selective the fusion score or the T3D cut-off value could be applied differently according to individual chemical classes rather than using as an ultimate score. Further refinement and investigation is needed to invent a more general fusion score, which is less sensitive to structural class differences.

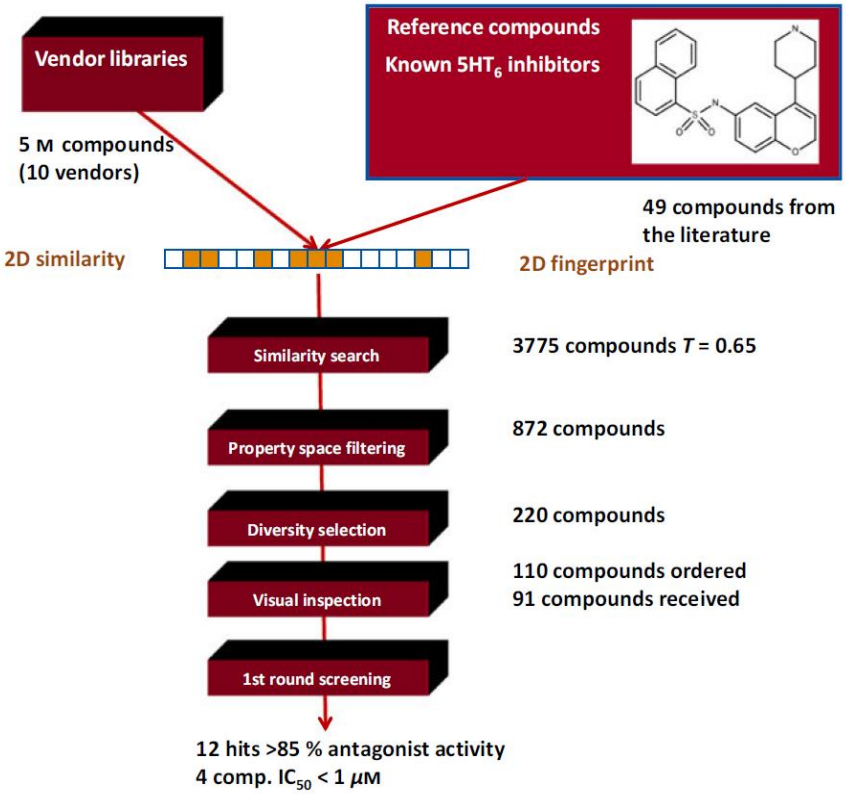
3.2. 5-HT_6 antagonist focused library selection and biological evaluation

3.2.1. Focused library selection and biological validation, first round

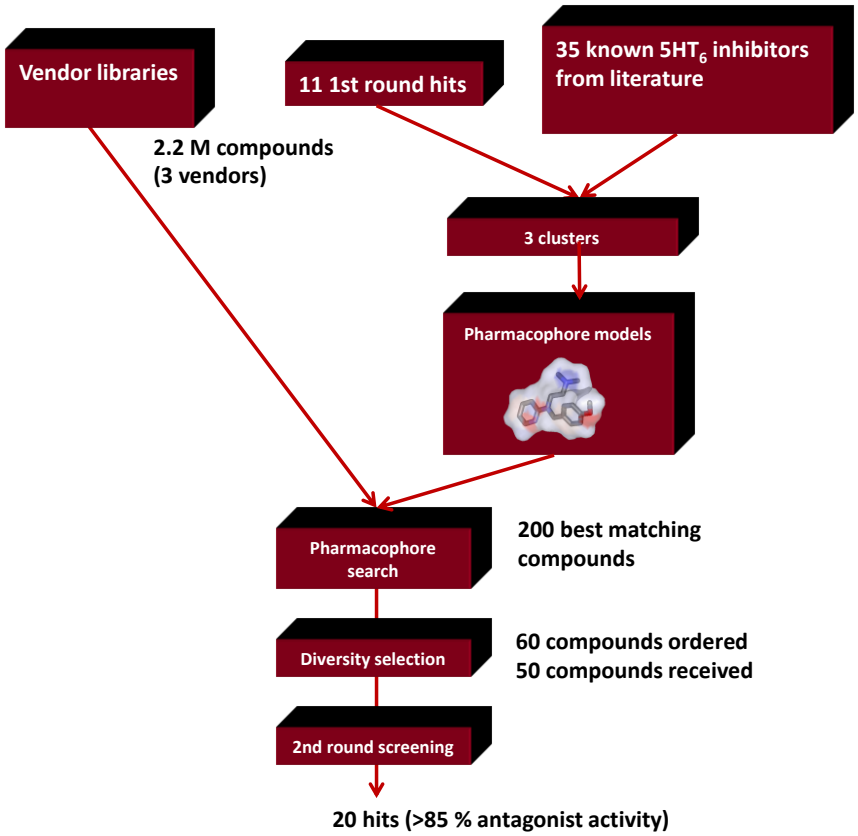
A multi-stage *in silico/in vitro* screening cascade was used for the selection of the 5-HT_6 target-focused libraries including 2D similarity selection, property-based filtering, diversity selection and biological screening. We collected 49 known 5-HT_6 antagonists („seeds") from the available literature to define the Reference Space, which represented various chemotypes (cca. 25).

At the compilation of the searchable drug-like Chemical Space, we used the 5M compounds' vendor databases (Figure 18). At the initial 2D similarity search (Figure 18a) we set the similarity threshold (T2D ≥ 0.65). Similarity search was followed by the „reference property space" filtering and diversity selection. The property (parameter) space was determined by calculating 6 typical physico-chemical parameters according for the reference compounds as described earlier. The central 90 % range (P5- P95) of the calculated parameters for the reference compounds was applied (Table 7). All the 6 parameters were applied by equal weight.

a.


Vendor libraries
5 M compounds
(10 vendors)
Reference compounds
Known 5HT6 inhibitors
49 compounds from
the literature
2D similarity
2D fingerprint
Similarity search
3775 compounds T = 0.65
Property space filtering
872 compounds
Diversity selection
220 compounds
Visual inspection
110 compounds ordered
91 compounds received
1st round screening
12 hits >85 % antagonist activity
4 comp. IC50 < 1 μM

b.


Vendor libraries
11 1st round hits
35 known 5HT6 inhibitors
from literature
2.2 M compounds
(3 vendors)
3 clusters
Pharmacophore models
Pharmacophore
search
200 best matching
compounds
Diversity selection
60 compounds ordered
50 compounds received
2nd round screening
20 hits (>85 % antagonist activity)

c.

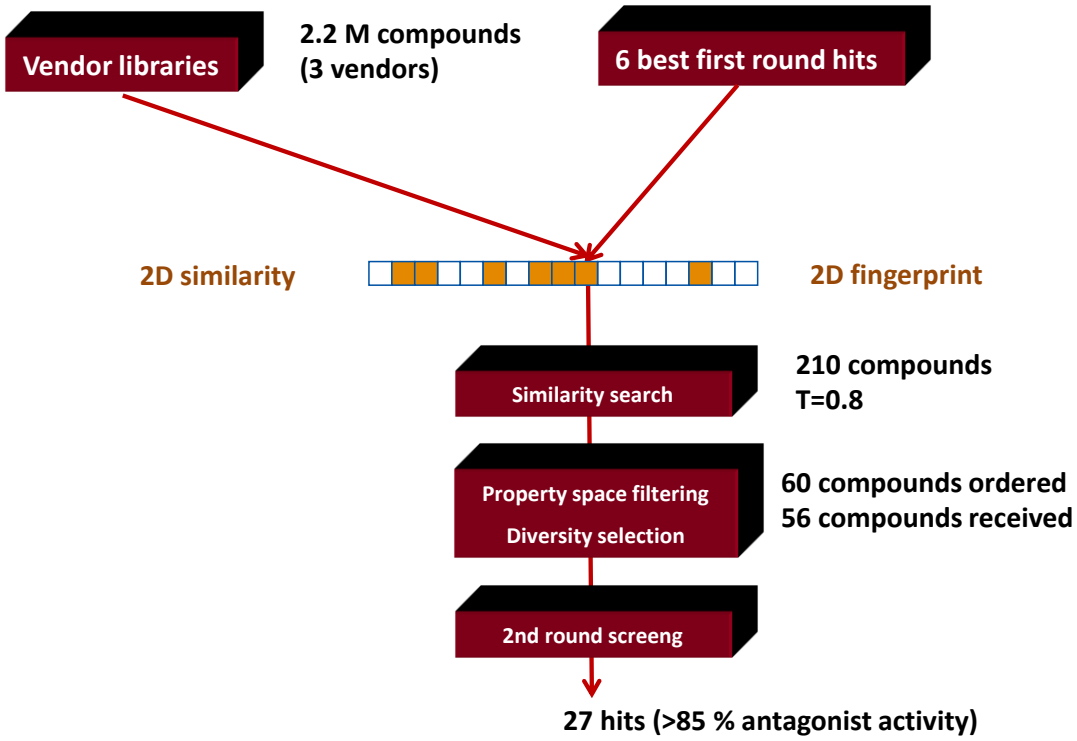


Figure 18a-c. Multi-stage *in silico/in vitro* screening cascade to select and validate 5-HT_6 focused libraries

After property space filtering the diversity of the selected molecule set was demonstrated by the LogP/Mwt graph of 872 compounds obtained by property space filtering of the similarity search results (Figure 18a vs. 19).

	MolWeight	LogP	TPSA	H bond donor	H bond acceptor	Rotatable bond
P95	476.0	4.72	96.60	3	7	5
P5	330.4	1.10	40.10	0	3	2

Table 7. The property space ranges for the known 5-HT_6 antagonists applied as seed compounds

In the first round multi-step focused library selection including property space filtering, diversity selection and visual inspection we finally acquired 91 compounds for *in vitro* screening.

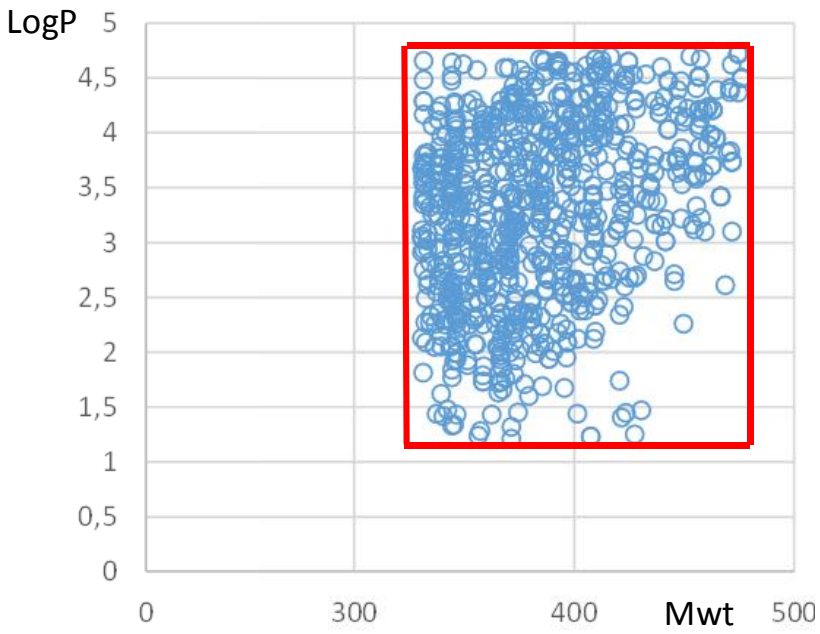


Figure 19. LogP/Mwt space of the compound set (872 compounds) after the 1st round similarity search and property space filtering

3.2.2.Second round screening and biological evaluation

We used a pharmacophore search algorithm for the second round focused library selection (Figure 18b). The 5-HT_6 antagonists have typical molecular architecture as shown before. For pharmacophore model building 49 structurally different seed compounds and 11 first round hit molecules were involved (Figure 18b).

Based on the structures of all ligands a common pharmacophore hypothesis was attempted consisting of four sites (AARR – acceptor/acceptor/ring/ring). These four-pharmacophore features covered 75% of all 60 active molecules. In order to increase the coverage by pharmacophore models, we divided the molecules into three distinct clusters, which allowed to create a pharmacophore hypothesis with five sites for these three datasets (cluster 1: AAHRR type; cluster 2: AADRR type; and cluster 3: AAHRR type).

The five-point pharmacophores contain two hydrogen bond acceptors (A), one hydrophobic group (H) or hydrogen bond donor group (D), and two aromatic rings (R).
Cluster 1 contained 12 seed molecules and 5 hits; Cluster 2 had 11 seeds and 6 hits, while cluster 3 (sulfone-bridge) contained only 12 seed molecules and no molecules from the first round hits.

The pharmacophore models with the highest survival score for all three clusters can be seen in Figure 20a, 20b and 20c. The first and the third model in Figure 20a and 20c look similar, but not identical. The most striking difference is the position of the hydrophobic site.

Pharmacophore model for Cluster #2 (Figure 20b) substantially differs from the others, since the 5th feature is a D (donor) instead of a H (hydrophobic) site.

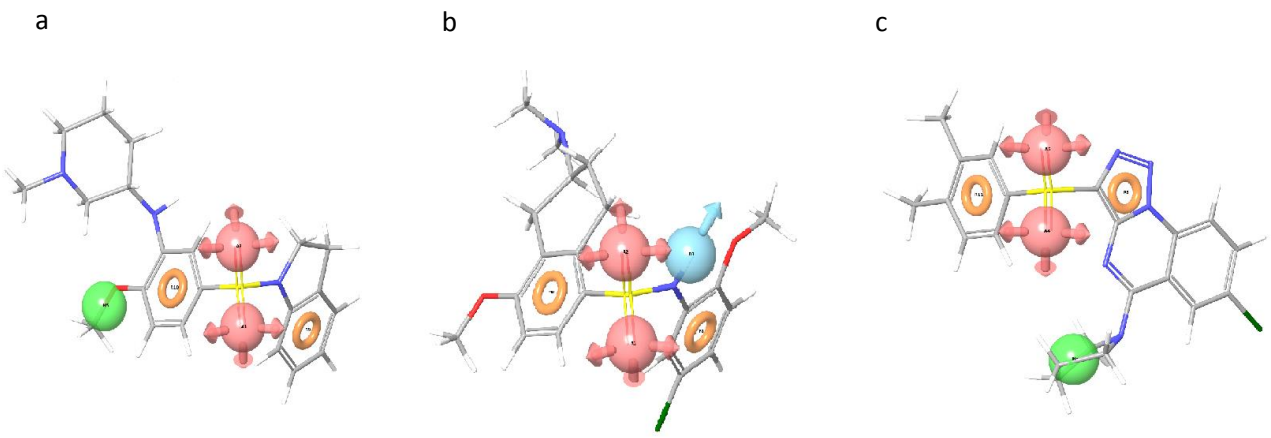


Figure 20a-c. Pharmacophore hypothesis of AAHRR,AADRR and AAHRR type created for Cluster #1, Cluster #2 and Cluster #3

In these figures orange rings represent aromatic rings and green spheres are for hydrophobic pharmacophore sites. Pink spheres are for hydrogen bond acceptor sites, while the cyan sphere is for a hydrogen bond donor site. The arrows define directionality.

All three pharmacophore models were used to screen a combined vendor database which contained over 2.2 million compounds (ChemDiv, Enamine, IBS).

We checked manually the 50 top scoring molecules coming from each pharmacophore model searching (altogether 150) and we bought 70 of them from the 3 compound suppliers. Sixty of them were received and tested for 5-HT$_6$ receptor antagonism.

Although the similarity of the pharmacophore model for Cluster #1 and #3, the result of the database screening was significantly different. In Cluster #2 the 5th feature is a D (donor) instead of a H (hydrophobic) site, so it can be an explanation for the difference in the results of this screening.

In comparison, we also carried out a typical, second round 2D similarity search using the 6 first round hits ($IC_{50} \leq 1$ μM) as seed compounds and the same 2.2 M vendor library was used as drug-like chemical space similarly to the pharmacophore-based *in silico* screening. (Figure 19c). Tanimoto ≥ 0.8 cut off value was used and after property space filtering as well as diversity based selection we had 60 compounds to order. Finally, we obtained 56 compounds for *in vitro* screening.

3.2.3. The results of the first and second round screening and 2D search

First round screening was purely based on 2D similarity search and property–based filtering. Compounds (12) having antagonist activity ≥ 85 % were considered as hits in a concentration of 10μM and selected for IC_{50} determination (Table 8). Altogether 9 compounds had IC_{50} values ≤ 2 μM; Between the 9 compounds there were 6 compounds having IC_{50} values ≤ 1 μM and 3 compounds IC_{50} values ≤ 500 nM. The best hit in the first round screening was #1-HT_6 (IC_{50}=20 nM).

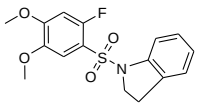
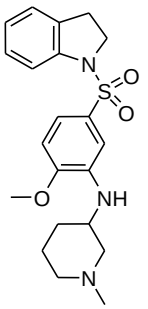
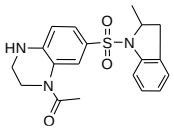
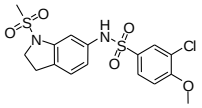
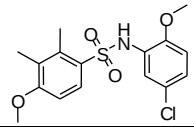
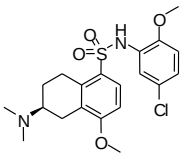
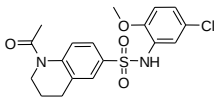
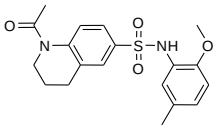
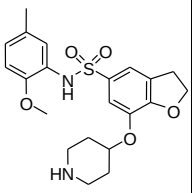
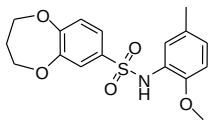
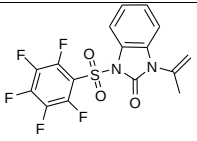
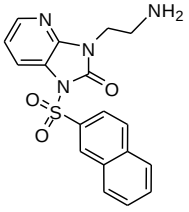
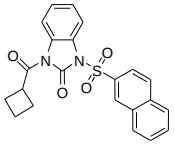
ID	Compound structure	Antagonist efficacy at 10μM(%)	IC50 (μM)	Similarity	Seed structure	Seed ID
1		96	0.020	0.73		2
4		95		0.67		
10		94	0.403	0.73		
12		94	0.753	0.73		12
7		88	2.46	0.66		
8		89	1.0	0.65		5
11		93	6.4	0.74		
9		96	2.0	0.71		31
2		92	2.8	0.71		

Table 8. First round hit compounds with their structural evolution from the 'seeds'

Low structural novelty was achieved, while the hit rate was 13% at T2D ≥ 0.65 similarity threshold and from the architecture of the parent (seed) compounds can we could recognize the selected compounds.

In the second round screening we involved pharmacophore models to select focused libraries that would give a better chance to identify active moieties with novel structures or chemotypes. We also applied routine 2D similarity search as a comparison.

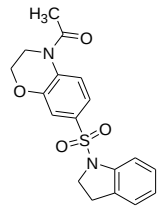
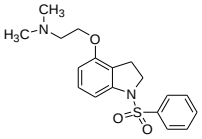
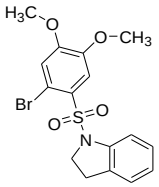
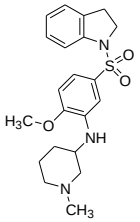
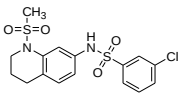
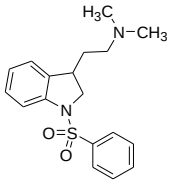
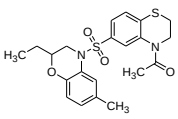
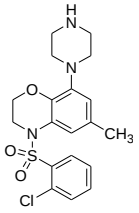
	Hit structure	**Mol Weight**	**IC_{50} pharmacophore selection (nM)**	**IC_{50} 2D similarity selection (nM)**	**Tanimoto similarity to the 1st round seeds**	**1st round seed structure**	**LE**	**LLE**	**$LLE_{(AT)}$**
1		358.41	45		25 (0.755)		0.51	11.3	0.73
2		398.27		1.9	2 (0.750)		0.52	8.77	0.63
3		400.9		3.6	20 (0.725)		0.46	9.42	0.62
4		432.55		4.9	40 (0.683)		0.39	8.28	0.50

Table 9. Best second round hit compounds with their structural evolution from the ‘seeds’

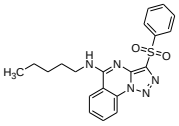
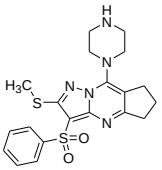
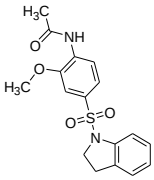
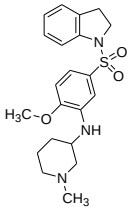
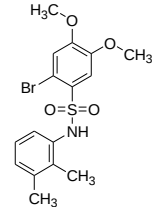
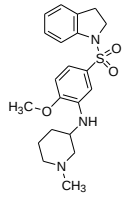
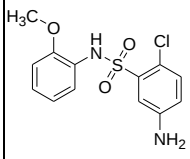
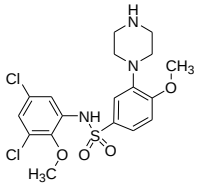
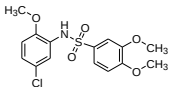
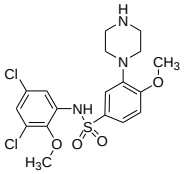
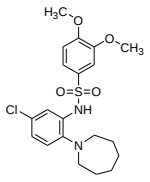
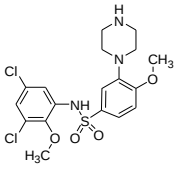
	Hit structure	Mol Weight	IC_{50} pharmacophore selection (nM)	IC_{50} 2D similarity selection (nM)	Tanimoto similarity to the 1st round seeds	1st round seed structure	LE	LLE	$LLE_{(AT)}$
5		395.47	15.3		38 (0.557)		0.38	6.03	0.41
6		346.4		24.8	2 (0.782)		0.43	8.61	0.60
7		400.28	27		2 (0.538)		0.45	6.43	0.49
8		312.77	30		35 (0.677)		0.52	8.23	0.68
9		357.8	31		35 (0.674)		0.48	7.70	0.57
10		424.94	33		35 (0.582)		0.37	6.10	0.41

Table 9 cont.

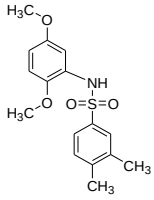
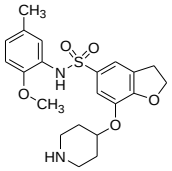
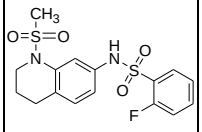
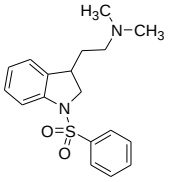
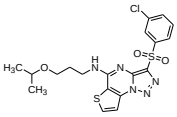
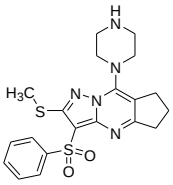
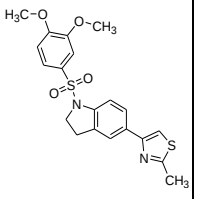
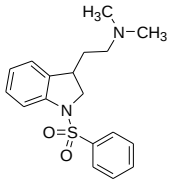
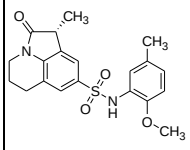
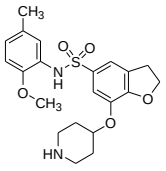
	Hit structure	Mol Weight	IC_{50} pharmacophore selection (nM)	IC_{50} 2D similarity selection (nM)	Tanimoto similarity to the 1st round seeds	1st round seed structure	LE	LLE	$LLE_{(AT)}$
11		321.39	33		5 (0.584)		0.47	7.08	0.55
12		384.44		41.2	20 (0.733)		0.40	8.43	0.57
13		465.97	45		38 (0.563)		0.33	5.88	0.38
14		416.51		59	20 (0.652)		0,35	6.68	0.44
15		386.46	88		5 (0.613)		0.36	6.86	0.46

Table 9. cont.

Table 9. shows the biological results for the best 15 compounds for the pharmacophore and 2D similarity search. To compare the chemotype switching (scaffold hopping) we analyzed the structural evolution of the hits coming from either the pharmacophore search or from the 2D similarity search. In the case of 2D similarity search the corresponding first rounds hits (as their seed compounds) are shown, while in the case of the pharmacophore model derived compounds the closest similar first round hit compounds were identified and displayed for easy comparison.

We found that pharmacophore search was a particularly powerful tool for selecting compounds from relatively large size compound libraries: 20 compounds out of 50 were found active (>85 % antagonist activity, hit rate = 40 %).

The average 2D similarity to the closest similar first round hit compounds was 0.586, which has already anticipated novelty, and really, 3 novel chemotypes were found compared with the first round hits, which served seeds for the 2D selection. In terms of the hit rate the purely 2D similarity search (T2D ≥ 0.8) around the 6 first round hits ($IC_{50} \leq 1$ μM) gave somewhat better results (27 compounds out of 56 were found active, hit rate= 51 %). The average 2D similarity values to the first round hits as seeds were 0.704 (at that level less chance for novelty was expected).

The calculated developability scores (LE, LLE, LLE_{AT}) in all cases of the molecules met the the required values (LE is preferred if >0.3; LLE if >5, and LLE_{AT} if >0.3).

3.2.4. Analysis of the hits

We searched the best hit compounds came from the 2nd round screening in the PubChem DB for novelty and biological activity. Those compounds were chosen mostly, where T2D < 0.6 to the first round hits (pharmacophore model-based selection). We identified compound #1-5-HT_6 by pharmacophore screening (T2D = 0.755, IC_{50}= 45 nM), although this compound is not completely novel, we can find it in the DB but no biological activity is given. Therefore, our most active compound represents a novel 5-HT_6 antagonist.

For compound #2-5-HT_6 (T=0.750, IC_{50}= 1.9 nM) only similar compounds (e.g. Figure 21) were listed in PubChem with various biological activities, however, 5-HT_6 antagonist activity was not reported.

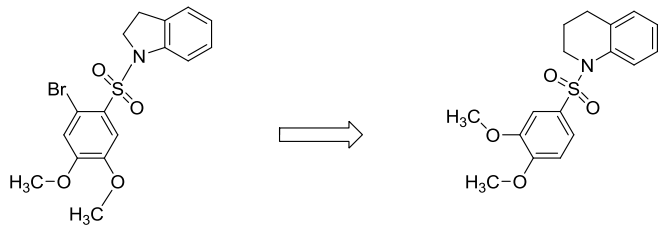


Figure 21. The structure of compound #2-5-HT_6 with the closest analogue from PubChem

For compound #3-5-HT_6 (T=0.725, IC_{50}= 3.6 nM) no close analogue was reported (although the chemotype has not changed much compared with the first round hits)

For compound #4-5-HT_6 (T=0.683, IC_{50}= 4.9 nM) a related analogue (Figure 22) is known as modulators of interaction between CendR peptide and neuropilin-1 (NRP-1) using Fluorescence Polarization assay.

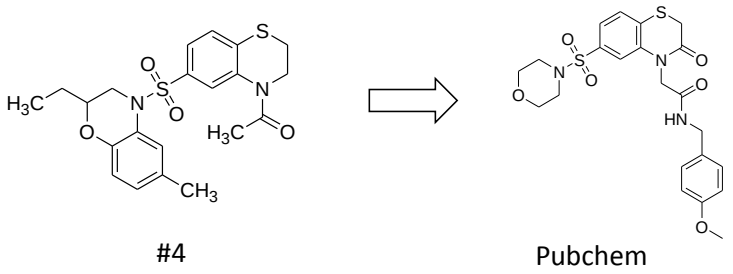


Figure 22. The structure of compound #4-5-HT_6 with the closest analogue from PubChem

Compound #5-5-HT_6 was identified by pharmacophore screening (T= 0.557, IC_{50}= 15.2 nM) and represent a novel chemotype in our studies. The primary biological activity of a closest chemotype (Figure 23) is a corticotropin releasing factor-binding protein antagonism.

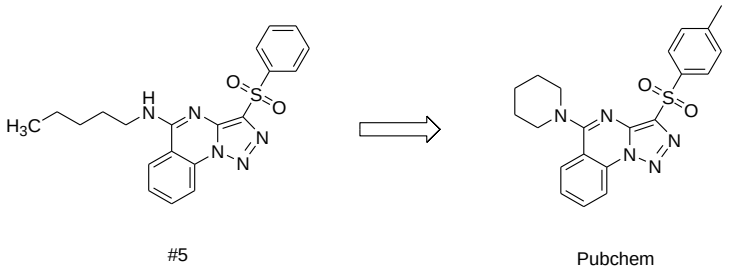


Figure 23. The structure of compound #5 with the closest analogue from PubChem

Compound #13-5-HT_6 was identified by pharmacophore screening (T= 0.563, IC_{50}= 45 nM) and represent another chemotype in our virtual screening, however, a structurally related ring-system is reported in 5-HT_6 receptor antagonists.[109] (Figure 24).

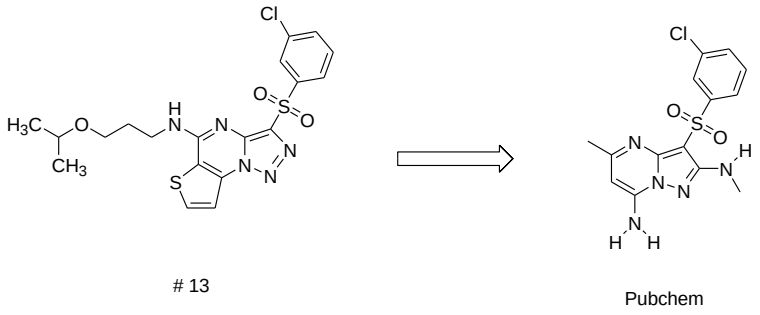


Figure 24. The structure of compound #13-5-HT_6 with the closest analogue from PubChem

3.2.5. Conclusion of the 5-HT_6 study

The integration of the 2D similarity search with pharmacophore matching screening looks a powerful combination for selecting focused libraries with enriched novelty. Our approach with clustering active molecules and developing distinct pharmacophore hypotheses for all clusters resulted in improved models. In the first screening round 12 compounds showed > 85

% antagonist efficacy out of the 91 screened, the hit rate was 13 %. In the second round (hit validation) screening phase pharmacophore models were built and applied as a complementary approach to the 2D similarity search method. The pharmacophore search resulted in a high hit rate (40 %) and led to novel chemotypes compared to the seed structures, while 2D similarity search had a slightly better hit rate (51 %), but lacking novelty.

Ligand efficiency indices demonstrated the utility of the 2D/3D combined virtual screening cascade presented here in the early-phase drug discovery.

4. Summary and conclusion of the two screening projects

Drug discovery has a high demand to identify novel, active structures in a fast and efficient way, therefore, *in silico* methods replace the conventional, random high-throughput screening. Two-dimensional similarity search is a widely used *in silico* technique in the pharma research selecting compounds from huge compound repositories based on theb structure of known, biologically active compounds. This method is robust and fast but has several weak points. The traditional fingerprint-based 2D similarity approaches are able to find similar structures that in fact resembles to the parent, seed compounds, therefore, to achieve novelty and higher hit rates are critical question. Increasing the hit rate is also a financial issue if the focused library provides more hits and less „junk" molecules it reduces the acquisition and screening costs. Achieving novelty together with increased hit rate requires combination with other in silico techniques.

In the first study, generating and evaluating a PDE-4 focused library in combination with 3D similarity assessment is a viable option to enrich the focused library with potential hits and removing the formally 2D similar compounds, thus increasing the hit rate. Conformational flexibility, shape alignment let us enter the '3D world' and it was confirmed that compounds with high 3D similarities could have similar binding features. In the PDE-4 study we applied Screen3D after characterizing the significant differences between 2D and 3D similarities using a previously generated PDE-5 focused set. Using this experience we devised a combination of 2D and 3D smilarity score ($T2D + T3D \geq 1.5$) and applying that fusion score we reached higher hit rate. Further investigation would lead to further improvement.

Another key problem that the multi-million compound repositories contain lots of redundant structures and difficult to achieve structural novelty. Three dimensional ligand and structure-based approaches could help to increase novelty. Pharmacophore models combine the 3D characteristics of the hit compounds and could be applied in combination with the robust 2D methods. In the second study, we compared the 2D ligand based similarity

selection with the 3D pharmacophore search in the second round screening phase using the same compound database and first round hits. In this study pharmacophore search was found superior in finding novel structures. Since pharmacophore model building requires much expertise (expensive software) and time, furthermore screening multi-million compound DBs are computational intensive a suitable combination with the robust 2D similarity search methods is more favorable in virtual screening.

The integrated virtual screening approaches resulted in several, novel biologically active compounds and the hit rate was much higher than the typical random screening.

PDE-4B: In the first screening round seven compounds were considered as hits (IC_{50} ranges from 0.05 to 16 μM) and the hit rate was 6.6% (>47% inhibition at 10 μM). The highest activity hit was 53 nM. The hit rate of the 2nd round screening was 28.5% (>50% inhibition at 10 μM), while 10% hit rate was calculated if only those compounds were considered that had an IC_{50} values < 2 μM. The highest activity hit was 53 nM (different from the first round hit). Seven compounds showed at least 10-fold selectivity towards PDE-4B over PDE-4D.

5-HT_6: In the first screening round, 12 compounds showed >85% antagonist efficacy of the 91 screened, and the hit rate was 13%. In the second-round (hit validation) screening phase, pharmacophore models were built and applied as a complementary approach to the 2D similarity search method. The pharmacophore search resulted in a high hit rate (40%, 20 compounds of 50 were found active, >85% antagonist activity) and three novel chemotypes were found compared with the first round hit compounds. The comparative 2D similarity search had slightly better hit rate (51%, 27 compounds out of 56 were found active), but lacking novel chemotypes. The highest activity hit was 1.9 nM.

Acknowledgements

This work was carried out in the TargetEx Ltd. in Budapest in a cooperation with the Institute of Pharmaceutical Chemistry, University of Szeged. I would like to express my deep gratitude to my supervisor, Prof. Dr. Ferenc Fülöp, head of the Institute of Pharmaceutical Chemistry, for his guidance and for providing the possibility to prepare my thesis under his management.
Especially thanks to my supreme supervisor, to Dr. György Dormán for his inspiring ideas, his helpful discussions and his continuous support through of my work. I am very grateful to Dr. Sándor Cseh, the head of the TargetEx Ltd. for his productive cooperation and support.
I would like to thank to Dr. Csaba Magyar for his computational contributions (pharmacophore model and 3D docking studies).
I would like to thank to all of my colleagues for their help and encouragement and specially to Dr. István Hajdú, for his scientific support and advice.
And finally, I wish to express my thanks to my husband and my family for their patience and encouragement.

5. References

[1] Paul, S. M., Mytelka, D. S., Dunwiddie, C. T., Persinger, C. C., Munos, B. H., Lindborg, S. R., & Schacht, A. L. (2010). How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat. Rev. Drug Discov. 9*(3), 203-214.

[2] Overington, J. P., Al-Lazikani, B., & Hopkins, A. L. (2006). How many drug targets are there?. *Nat. Rev. Drug Discov. 5*(12), 993-996.

[3] Rask-Andersen, M., Almén, M. S., & Schiöth, H. B. (2011). Trends in the exploitation of novel drug targets. *Nat. Rev. Drug Discov. 10*(8), 579-590.

[4] Ward, D., Martino, O. I., Simpson, S., & Stevens, A. J. (2013). Decline in new drug launches: myth or reality. BMJ Open 2013;3:e002088.

[5] Kola, I., & Landis, J. (2004). Can the pharmaceutical industry reduce attrition rates?. *Nat. Rev. Drug Discov. 3*(8), 711-716.

[6] Kodadek, T. (2011). The rise, fall and reinvention of combinatorial chemistry. *Chem. Commun. 47*(35), 9757-9763.

[7] Macarron, R. (2006). Critical review of the role of HTS in drug discovery. *Drug Discov. Today, 11*(7), 277-279.

[8] Schuffenhauer, A., Ruedisser, S., Marzinzik, A., Jahnke, W., Selzer, P., & Jacoby, E. (2005). Library design for fragment based screening. *Curr. Top. Med. Chem. 5*(8), 751-762.

[9] Blum, L. C., & Reymond, J. L. (2009). 970 million druglike small molecules for virtual screening in the chemical universe database GDB-13. *J. Am.Chem. Soc. 131*(25), 8732-8733.

[10] Schmuck, C., & Wich, P. (2006). Combinatorial receptor finding—large and random vs. small and focused libraries. *New J. Chem. 30*(10), 1377-1385.

[11] Flagstad, T., Min, G., Bonnet, K., Morgentin, R., Roche, D., Clausen, M. H., & Nielsen, T. E. (2016). Synthesis of sp3-rich scaffolds for molecular libraries through complexity-generating cascade reactions. *Org. Biomol. Chem. 14*(21), 4943-4946.

[12] Congreve, M., Carr, R., Murray, C., & Jhoti, H. (2003). A 'rule of three'for fragment-based lead discovery?. *Drug Discov. Today*, *8*(19), 876-877.

[13] Lipinski, C. A. (2004). Lead-and drug-like compounds: the rule-of-five revolution. *Drug Discov. Today: Technologies*, *1*(4), 337-341.

[14] Carr, R. A., Congreve, M., Murray, C. W., & Rees, D. C. (2005). Fragment-based lead discovery: leads by design. *Drug Discov. Today*, *10*(14), 987-992.

[15] Khanna, I. (2012). Drug discovery in pharmaceutical industry: productivity challenges and trends. *Drug Discov. Today*, *17*(19), 1088-1102.

[16] Scannell, J. W., Blanckley, A., Boldon, H., & Warrington, B. (2012). Diagnosing the decline in pharmaceutical R&D efficiency. *Nat. Rev. Drug Discov*, *11*(3), 191-200.

[17] Khanna, R., Guler, I., & Nerkar, A. (2016). Fail often, fail big, and fail fast? Learning from small failures and R&D performance in the pharmaceutical industry. *Acad. Manage. J.l*, *59*(2), 436-459.

[18] Bajorath, J. (2002). Integration of virtual and high-throughput screening. *Nat. Rev. Drug Discov. 1*(11), 882-894.

[19] H Ma, X., Zhu, F., Liu, X., Shi, Z., X Zhang, J., Y Yang, S., ... & Z Chen, Y. (2012). Virtual screening methods as tools for drug lead discovery from large chemical libraries. *Curr.Med. Chem. 19*(32), 5562-5571

[20] Lionta, E., Spyrou, G., K Vassilatis, D., & Cournia, Z. (2014). Structure-based virtual screening for drug discovery: principles, applications and recent advances. *Curr. Top.Med. Chem. 14*(16), 1923-1938.

[21] McInnes, C. (2007). Virtual screening strategies in drug discovery. *Curr. Opin. Biol. 11*(5), 494-502.

[22] Drwal, M. N., & Griffith, R. (2013). Combination of ligand-and structure-based methods in virtual screening. *Drug Discov. Today: Technologies*, *10*(3), e395-e401.

[23] Gorse, A. D. (2006). Diversity in medicinal chemistry space. *Curr. Top.Med. Chem. 6*(1), 3-18.

[24] Böhm M. Virtual Screening of Chemical Space: From Generic Compound Collections to Tailored Screening Libraries. pp. 3-26. in: *Virtual screening: principles, challenges, and practical guidelines* (Vol. 48). Eds. Mannhold, R., Kubinyi, H., & Folkers, G. (2011).. C. Sotriffer (Ed.). John Wiley & Sons.

[25] Irwin, J. J., & Shoichet, B. K. (2005). ZINC–a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model. 45*(1), 177.

[26] Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S., & Coleman, R. G. (2012). ZINC: a free tool to discover chemistry for biology. *J. Chem. Inf. Model. 52*(7), 1757.

[27] http://www.chemspider.com

[28] https://www.emolecules.com/

[29] Russ, A. P., & Lampel, S. (2005). The druggable genome: an update. *Drug Discov. Today*, *10*(23), 1607-1610.

[30] The PubChem Project. Available online: http://pubchem.ncbi.nlm.nih.gov/ (accessed on 03 May 2017).

[31] CheEMBL-European Bioinformatics Institute. Available online: https://www.ebi.ac.uk/chembl/ (accessed on 03 May 2017)

[32] Binding DB—The Binding Database. Available online: http://www.bindingdb.org/ (accessed on 03 May 2017)

[33] Olah, M.; Rad, R.; Ostopovici, L.; Bora, A.; Hadaruga, N.; Hadaruga, D.; Moldovan, R.; Fulias, A.; Mracec, M.; Oprea, T. I. WOMBAT and WOMBAT-PK: Bioactivity Databases for Lead and Drug Discovery. In Chemical Biology: From Small Molecules to Systems Biology and Drug Design; Schreiber, S. L., Kapoor, T. M., Wess, G., Eds.; Wiley-VCH: New York, 2007; pp 760−786.

[34] https://pharmaintelligence.informa.com/products-and-services/data-and-analysis/pharmaprojects

[35] https://integrity.thomson-pharma.com/integrity/xmlxsl/

[36] Cortés-Cabrera, Á., Murcia, P. A. S., Morreale, A., & Gago, F. (2015). Ligand-Based Drug Discovery and Design. *In Silico Drug Discovery and Design: Theory, Methods, Challenges, and Applications* (pp. 99-121). CRC Press.

[37] Koeppen, H., Kriegl, J., Lessel, U., Tautermann, C. S., & Wellenzohn, B. (2011). Ligand-Based Virtual Screening. *Virtual Screening: Principles, Challenges, and Practical Guidelines*, 61-85.

[38] Maggiora, G., Vogt, M., Stumpfe, D., & Bajorath, J. (2013). Molecular similarity in medicinal chemistry: miniperspective. *J. Med. Chem. 57*(8), 3186-3204.

[39]Maggiora, G.M. and V. Shanmugasundaram. 2011. Molecular similarity measures. *Method. Mol. Biol.* 672: 39–100.

[40] Martin,Y.C., Kofron, J.L.,and Traphagen, L.M. (2002) Do structurally similar molecules have similar biological activity? *J. Med. Chem.* 45, 4350–4358.

[41] Butina, D. 1999. Unsupervised database clustering based on daylight's fingerprint and Tanimoto similarity: A fast and automated way to cluster small and large data sets. *J. Chem. Inf. Comput. Sci*. 39(4): 747–750.

[42] Willett, P. (2011). Similarity searching using 2D structural fingerprints. *Chemoinformatics and computational chemical biology*, 133-158.

[43] Willett P (2006) Similarity-based virtual screening using 2D fingerprints. *Drug DiscovToday* 11:1046–1053.

[44] Venkatraman, V., Pérez-Nueno, V. I., Mavridis, L., & Ritchie, D. W. (2010). Comprehensive comparison of ligand-based virtual screening tools against the DUD data set reveals limitations of current 3D methods. *J. Chem. Inf. Model. 50*(12), 2079-2093.

[45] Schuffenhauer, A.; Gillet, V. J.; Willett, P. Similarity Searching in Files of Three-Dimensional Chemical Structures: Analysis of the BIOSTER Database Using Two-Dimensional Finger prints and Molecular Field Descriptors. *J. Chem. Inf. Model.* 2000, 40, 295–307.

[46] Bajorath, J. (2002). Integration of virtual and high-throughput screening. *Nat. Rev. Drug Discov. 1*(11), 882-894.

[47] Edwards, B. S., Bologa, C., Young, S. M., Balakin, K. V., Prossnitz, E. R., Savchuck, N. P., ... & Oprea, T. I. (2005). Integration of virtual screening with high-throughput flow cytometry to identify novel small molecule formylpeptide receptor antagonists. *Mol. Pharmacol. 68*(5), 1301-1310.

[48] Polgár, T.; Keseru, G.M. Integration of virtual and high throughput screening in lead discovery settings. *Comb. Chem. High Throughput Screen*. 2011, 14, 889–897.

[49] Polgár, T., Baki, A., Szendrei, G. I., & Keserű, G. M. (2005). Comparative virtual and experimental high-throughput screening for glycogen synthase kinase-3β inhibitors. *J. Med. Chem., 48*(25), 7946-7959.

[50] Stahura F.L., Bajorath J. (2004) Virtual screening methods that complement HTS. Comb Chem High Throughput Screen;7:259–269.

[51] Tömöri T, Hajdú I, Barna L, Lőrincz Z, Cseh S and Dormán G, Combining 2D and 3D *in silico* methods for rapid selection of potential PDE-5 inhibitors from multi-million compounds' repositories. Biological evaluation, *Mol Divers.* 16(1), 59-72 (2012).

[52] Di, L., Kerns, E. H., & Carter, G. T. (2009). Drug-like property concepts in pharmaceutical design. *Curr. Pharm. Design*, *15*(19), 2184-2194.

[53] Lipinski C A, Lombardo F, Dominy B W ,Feeney P J. (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv Rev* 23:3–25.

[54] Veber DF, Johnson SR, Cheng H-Y, Smith BR, Ward KW, Kenneth DK (2002) Molecular properties that influence the oral bioavailability of drug candidates .*J .Med. Chem* .45:2615–2623.

[55] Schultes, S., de Graaf, C., Haaksma, E. E., de Esch, I. J., Leurs, R., & Krämer, O. (2010). Ligand efficiency as a guide in fragment hit selection and optimization. *Drug Discov. Today: Technologies*, *7*(3), e157-e162.

[56] High-throughput ADMETox Estimation: *In vitro* & *in silico* Approaches, Eds. Darvas F and Dormán G, BioTechniques Press, Eaton Publishing (2002)

[57] Rishton, G. M. (1997). Reactive compounds and in vitro false positives in HTS. *Drug Discov. Today*, *2*(9), 382-384.

[58] Baell, J. B., & Holloway, G. A. (2010). New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J. Med. Chem. 53*(7), 2719-2740.

[59] Boda, K., Seidel, T., & Gasteiger, J. (2007). Structure and reaction based evaluation of synthetic accessibility. *J. Comput. Aid. Mol. Des.*, *21*(6), 311-325.

[60] Morphy R (2006) The influence of target family and functional activity on the physicochemical properties. *J. Med. Chem*. 49:2969– 2978.

[61] Wager TT, Hou X, Verhoest PR, Villalobos A (2010) Moving beyond rules: the development of a central nervous system multiparameter optimization (CNS MPO) approach to enable alignment of druglike properties. *ACS Chem. Neurosci.* 1:435–449

[62] Tyrchan C, Blomberg N, Engkvist O, Kogej T, Muresan S. (2009) Physicochemical property profiles of marketed drugs, clinical candidates and bioactive compounds. *Bioorg. Med. Chem. Lett.* 19:6943–6947.

[63] Martin, Y. C. Molecular Diversity: How We Measure It? Has It Lived Up to Its Promise? *Farmaco* 2001, 56, 137-139.

[64] Cao, Y., Jiang, T., & Girke, T. (2008). A maximum common substructure-based algorithm for searching and predicting drug-like compounds. *Bioinformatics*, *24*(13), i366-i374.

[65] Papp, Á., Gulyás-Forró, A., Gulyás, Z., Dormán, G., Ürge, L., & Darvas, F. (2006). Explicit Diversity Index (EDI): a novel measure for assessing the diversity of compound databases. *J . Chem. Inf. Model.*, *46*(5), 1898-1904.

[66] Maggiora, G. M., & Shanmugasundaram, V. (2011). Molecular similarity measures. *Chemoinformatics and computational chemical biology*, 39-100.

[67] Willett P,Winterman V (1986) A comparison of some measures of intermolecular structural similarity. *Quant. Struct-Act Rel.* 5:18–25.

[68] Glen, R. C., & Adams, S. E. (2006). Similarity metrics and descriptor spaces–which combinations to choose?. *Mol. Inform. 25*(12), 1133-1142.

[69] Dixon SL, Koehler RT (1999) The hidden component of size in two dimensional fragment descriptors: side effects on sampling in bioactive libraries. *J. Med. Chem.* 42:2887–2900.

[70] Whittle, M., Gillet, V. J., Willett, P., & Loesel, J. (2006). Analysis of data fusion methods in virtual screening: similarity and group fusion. *J. Chem. Inf. Model*, *46*(6), 2206-2219

[71] Xue L, Stahura FL, Godden JW, Bajorath J (2001) Fingerprint scaling increases the probability of identifying molecules with similar activity in virtual screening calculations. *J. Chem. Inf. Comp. Sci*. 41:746–753.

[72] Hert J, Willett P, Wilton DJ, Acklin P, Azzaoui K, Jacoby E, Schuffenhauer A (2005) Enhancing the effectiveness of similarity-based virtual screening using nearest-neighbour information. *J. Med. Chem.* 48:7049–7054.

[73] Medina-Franco, J. L., Martinez-Mayorga, K., & Meurice, N. (2014). Balancing novelty with confined chemical space in modern drug discovery. *Expert Opin. Drug Dis.*, *9*(2), 151-165.

[74] Flachner B, Tömöri T, Hajdú I, Dobi K, Lőrincz Z, Cseh S and Dormán G, Rapid *in silico* selection of an MCH1-R antagonists focused library from multi-million compounds' repositories. Biological evaluation *Med. Chem. Res.* 23(3) 1234-1247 (2014).

[75] Heikamp, K. and J. Bajorath. 2013. The future of virtual compound screening. *Chem. Biol. Drug Des.* 81(1): 33–40

[76] Cortes-Cabrera, A., F. Gago, and A. Morreale. 2012. A reverse combination of structure-based and ligand-based strategies for virtual screening. *J. Comput. Aid. Mol. Des.* 26(3): 319–327.

[77] Decornez H, Gulyás-Forró A, Papp A, Szabó M, Sármay G, Hajdú I, Cseh S, Dormán G, Kitchen DB Design, selection, and evaluation of a general kinase-focused library. *Chem. Med. Chem*. 4(8), 1273-8 (2009).

[78] Kalaszi, A.; Szisz, D.; Imre, G.; Polgar, T. Screen3D: A novel fully flexible high-throughput shape-similarity search method. *J. Chem. Inf. Model.* 2014, 54, 1036–1049.

[79] Hann MM, Keserü GM (2012) Finding the sweet spot: the role of nature and nurture in medicinal chemistry. *Nat. Rev. Drug. Discov.* 11:355–365

[80] Murray CW, Rees DC (2009) The rise of fragment-based drug discovery. *Nat. Chem.* 1:187–192.

[81] Szaszkó, Mária, et al. "Identification of potential glutaminyl cyclase inhibitors from lead-like libraries by *in silico* and in vitro fragment-based screening." *Mol. Divers.* (2017): 1-12.

[82] Reddy AS, Pati SP, Kumar PP, Pradeep HN, Sastry GN (2007) Virtual screening in drug discovery—a computational perspective. *Curr. Protein Pept. Sc.* 8:329–351.

[83] Kate Loughney, Teresa R. Hill, Vincent A. Florio, Lothar Uher, Guy J. Rosman, Sharon L. Wolda, Bryan A. Jones, Monique L. Howard, Linda M. McAllister-Lucas, William K. Sonnenburg, Sharron H. Francis d, Jackie D. Corbin d, Joseph A. Beavo, Ken Ferguson, Isolation and characterization of cDNAs encoding PDE-5A, a human cGMP-binding, cGMP-specific 3',5'-cyclic nucleotide phosphodiesterase1, *Gene* 216 (1998) 139–147

[84] Conti M, Beavo J. Biochemistry and physiology of cyclic nucleotide phosphodiesterases: essential components in cyclic nucleotide signalling. *Annu. Rev. Biochem.* 2007;76:481-511

[85] Press NJ, Banner KH. PDE-4 inhibitors - a review of the current field. *Progr. Med. Chem.* 2009;47:37-74

[86] Victoria Boswell-Smith,[1] Domenico Spina,[1] and Clive P Page,Phosphodiesterase inhibitors, *Brit. J. Pharmacol.* 2006 Jan; 147(Suppl 1): S252–S257.

[87] Holenz, J.; Pauwels, P.J.; Diaz, J.L.; Merce, R.; Codony, X.; Buschmann, H. (2006) Medicinal chemistry strategies to 5-HT(6) receptor ligands as potential cognitive enhancers and antiobesity agents. *Drug. Discov. Today* 11, 283-299.

[88] Heal, D.J.; Smith, S.L.; Fisas, A.; Codony, X.; Buschmann, H. (2008) Selective 5-HT6 receptor ligands: progress in the development of a novel pharmacological approach to the treatment of obesity and related metabolic disorders. *Pharmacol. Ther.* 117, 207-231.

[89] Hoyer, D.; Clarke, D.E.; Fozard, J.R.; Hartig, P.R.; Martin, G.R.; Mylecharane, E.J.; Saxena, P.R.; Humphrey, P.P. (1994) International Union of Pharmacology classification of receptors for 5-hydroxytryptamine (Serotonin). *Pharmacol. Rev.* 46, 157-203.

[90] Hoyer, D.; Martin, G.R. (1996) Classification and nomenclature of 5-HT receptors: a comment on current issues. *Behav. Brain Res.* 73, 263-268.

[91] Woolley, M.L.; Marsden, C.A.; Fone, K.C.F. (2004) 5-ht6 receptors. Curr Drug Targets *CNS Neurol. Disord.* 3, 59-79.

[92] Kohen, R.; Metcalf, M.A.; Khan, N.; Druck, T.; Huebner, K.; Lachowicz, J.E.; Meltzer, H.Y.; Sibley, D.R.; Roth, B.L.; Hamblin, M.W. (1996) Cloning, characterization, and chromosomal localization of a human 5-HT6 serotonin receptor. *J. Neurochem.* 66, 47-56.

[93] Messina, D.; Annesi, G.; Serra, P.; Nicoletti, G.; Pasqua, A.; Annesi, F.; Tomaino, C.; Ciro- Candiano, I.C.; Carrideo, S.; Caracciolo, M. et al. (2002) Association of the 5-HT6 receptor gene polymorphism C267T with Parkinson's disease. *Neurology* 58, 828-829.

[94] Sebben, M.; Ansanay, H.; Bockaert, J.; Dumuis, A. (1994) 5-HT6 receptors positively coupled to adenylyl cyclase in striatal neurones in culture. *Neuroreport* 5, 2553-2557.

[95] Bonsi, P.; Cuomo, D.; Ding, J.; Sciamanna, G.; Ulrich, S.; Tscherter, A.; Bernardi, G.;Surmeier, D.J.; Pisani, A. (2007) Endogenous serotonin excites striatal cholinergic interneurons via the activation of 5-HT 2C, 5-HT6, and 5-HT7 serotonin receptors: implications for extrapyramidal side effects of serotonin reuptake inhibitors. *Neuropsychopharmacol*. 32, 1840-1854

[96] Hebert, L.E.; Scherr, P.A.; Bienias, J.L.; Bennett, D.A.; Evans, D.A. (2003) Alzheimer disease in the US population: prevalence estimates using the 2000 census. *Arch. Neurol.* 60, 1119-1122.

[97] Mount, C.; Downton, C. (2006) Alzheimer disease: progress or profit?. *Nat. Med.* 12, 780-784.

[98] Benhamu, B.; Martin-Fontecha, M.; Vazquez-Villa, H.; Pardo, L.; Lopez-Rodriguez, M.L. (2014) Serotonin 5-HT6 receptor antagonists for the treatment of cognitive deficiency in Alzheimer's disease. *J. Med. Chem.* 57, 7160-7181.

[99] Garcia-Alloza, M.; Hirst, W.D.; Chen, C.P.L.; Lasheras, B.; Francis, P.T.; Ramirez, M.J. (2004) Differential involvement of 5-HT(1B/1D) and 5-HT6 receptors in cognitive and noncognitive symptoms in Alzheimer's disease. *Neuropsychopharmacol.* 29, 410-416.

[100] Liu, K.G.; Robichaud, A.J. (2009) 5-HT6 Antagonists as Potential Treatment for Cognitive Dysfunction. *Drug Develop. Res.* 70, 145-168.

[101] Lopez-Rodrıguez M.L., Benhamu B., de la Fuente T., Sanz A., Pardo L., Campillo M. (2005) A three-dimensional pharmacophore model for 5-hydroxytryptamine6 (5-HT6) receptor antagonists. *J. Med. Chem.;*48:4216– 4219.

[102] Thompson WJ, Appleman MM (1971) Multiple cyclic nucleotide phosphodiesterase activities from rat brain. *Biochemistry-US* 10:311–316.

[103] Brini, M.; Marsault, R.; Bastianutto, C.; Alvarez, J.; Pozzan, T.; Rizzuto, R. (1995) Transfected aequorin in the measurement of cytosolic Ca2+ concentration ([Ca2+]c). A critical evaluation. *J. Biol. Chem.* 270, 9896-9903.

[104] Button, D.; Brownstein, M. (1993) Aequorin-expressing mammalian cell lines used to report Ca2+ mobilization. *Cell Calcium* 14, 663-671.

[105] Stables, J.; Green, A.; Marshall, F.; Fraser, N.; Knight, E.; Sautel, M.; Milligan, G.; Lee, M.; Rees, S.(1997) A bioluminescent assay for agonist activity at potentially any G-protein-coupled receptor. *Anal. Biochem.* 252, 115-126.

[106] Mattsson, C.; Sonesson, C.; Sandahl, A.; Greiner, H.E.; Gassen, M.; Plaschke, J.; Leibrock, J.; Böttcher, H. (2005) 2-Alkyl-3-(1,2,3,6-tetrahydropyridin-4-yl)-1H-indoles as novel 5-HT6 receptor agonists. *Bioorg. Med. Chem. Lett.* 15, 4230-4234.

[107] Nam, G.; Yoon, C.M.; Kim, E.; Rhee, C.K.; Kim, J.H.; Shin, J.H.; Kim, S.H. Syntheses and evaluation of pyrido[2,3-dlpyrimidine-2,4-diones as PDE 4 inhibitors. *Bioorg. Med. Chem. Lett.* 2001, *11*, 611–614.

[108] Drabczyńska, A.; Yuzlenko, O.; Köse, M.; Paskaleva, M.; Schiedel, A.C.; Karolak-Wojciechowska, J.; Handzlik, J.; Karcz, T.; Kuder, K.; Müller, C.E.; *et al.* Synthesis and biological activity of ricyclic cycloalkylimidazo-, pyrimido- and diazepinopurinediones. *Eur. J. Med. Chem.* 2011, *46*, 3590–3607.

[109] Ivachtchenko, A.V.; Golovina, E.S.; Kadieva, M.G.; Kysil, V.M.; Mitkin, O.D.; Tkachenko, S.E.; Okun, I.M. (2011) Synthesis and structure-activity relationship (SAR) of (5,7-disubstituted 3-phenylsulfonyl-pyrazolo[1,5-a]pyrimidin-2-yl)-methylamines as potent serotonin 5-HT(6) receptor (5-HT(6)R) antagonists. *J. Med. Chem.* 54, 8161-8173.