

# Chemically Advanced Template Search (CATS) for Scaffold-Hopping and Prospective Target Prediction for 'Orphan' Molecules

Michael Reutlinger,<sup>[a]</sup> Christian P. Koch,<sup>[a]</sup> Daniel Reker,<sup>[a]</sup> Nickolay Todoroff,<sup>[a]</sup> Petra Schneider,<sup>[a]</sup> Tiago Rodrigues,<sup>[a]</sup> and Gisbert Schneider<sup>\*[a]</sup>

**Keywords:** Drug design · Similarity searching · Combinatorial chemistry · Orphan drug · Polypharmacology · Self-organizing map · Virtual screening

Drug discovery is driven by the identification of new chemical entities (NCEs).<sup>[1,2]</sup> Virtual screening and de novo design techniques have been proven to serve this purpose, thereby complementing experimental biochemical and biological approaches.<sup>[3]</sup> Still, it remains a matter of debate, which particular molecular representation and similarity index are preferable for a given drug target in order to identify appropriate NCEs with minimal synthetic and testing effort involved.<sup>[4]</sup> Ligand-based chemical similarity approaches have been effectively applied to large-scale activity and target prediction for known drugs, some of the prominent methods being PASS developed by Poroikov et al.,<sup>[5]</sup> the techniques conceived by Mestres and co-workers,<sup>[6]</sup> and the similarity ensemble approach (SEA) implemented by the Shoichet group.<sup>[7]</sup> Here, we compared several popular two-dimensional molecular representations for their ability to retrieve actives (enrichment potential) and chemotypes (scaffold-hopping potential) from a collection of druglike bioactive compounds. Subsequently the applied chemical advanced template search (CATS)<sup>[8]</sup> was applied to predicting potential drug targets for a virtually assembled combinatorial compound library, from which we synthesized and successfully tested candidate compounds. The results demonstrate that CATS is not only suited for its intended purpose of NCE retrieval by scaffold-hopping,<sup>[9]</sup> but also for reliable target profiling of 'orphan' virtual molecules.<sup>[10]</sup> It thereby complements the suite of available validated tools for target prediction.

A framework for retrospective evaluation of similarity searching runs with different molecular representations ('descriptors') was established on basis of the COBRA collection of druglike bioactive compounds,<sup>[11]</sup> employing Euclidean distances for metric descriptors and the Tanimoto coefficient for fingerprint descriptors.<sup>[12]</sup> COBRA contains 12642 manually curated entries with 980 target protein subtype annotations. For 170 macromolecular drug targets with a minimum of 20 annotated active ligands per target, each compound annotated as 'active' was selected as a query in turn, and compared to all remaining compounds in the screening pool in terms of molecular descriptor simi-

larity, finally yielding sorted results lists with the most similar or least distant pool compounds sorted to the top. Although there are large collections of bioactive compounds available in the public domain,<sup>[13]</sup> we used the carefully compiled COBRA collection to i) reduce the risk of erroneous activity data and faulty compound structures,<sup>[14]</sup> and ii) avoid redundancy with existing tools that are based on such public structure-activity data. In addition, we intend to probe the value of a comparably small but well curated reference compound pool for target prediction.

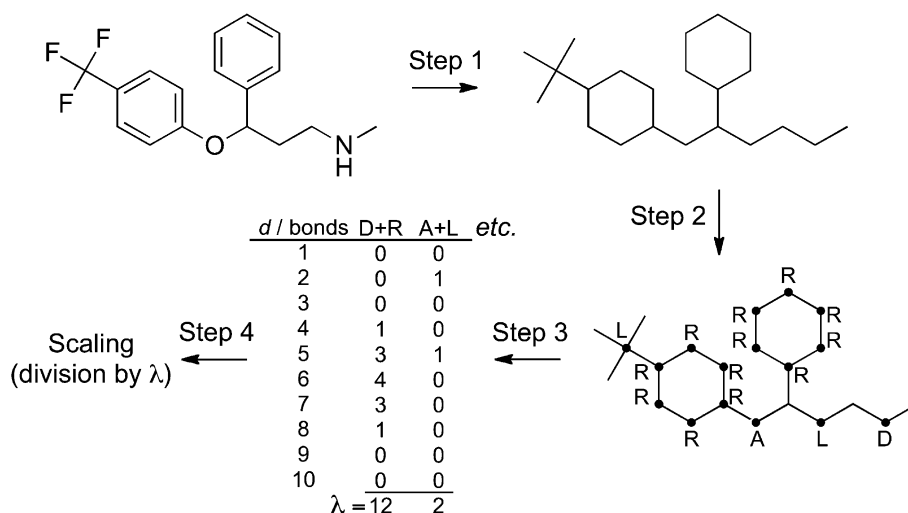
We used a representative set of descriptors and fingerprints for benchmarking. 'Morgan' fingerprints, closely related to extended-connectivity fingerprints (ECFP), are based on radial assessment of non-predefined potentially infinite molecular fragments.<sup>[17]</sup> The 'AtomPair' descriptor can be seen as a CATS predecessor merely denoting the occurrence of all pairs of atoms at a given topological distance.<sup>[18]</sup> The 'MACCS' keys represent substructure-based fingerprints,<sup>[19]</sup> and the 'RDkit' fingerprint implements a Daylight-like fingerprint based on hashed molecular subgraphs.<sup>[20]</sup> Latter fingerprints and descriptors were calculated using the open-source software package RDkit.<sup>[21]</sup> Finally, the 'MOE2D' descriptor consists of a standardized vector of physicochemical properties provided by the Molecular Operating Environment (v2011, Chemical Computing Group, Montreal).

At this point, we analyzed two versions of CATS vectors, namely the originally described CATS1<sup>[8]</sup> and CATS2, which distinguishes lipophilic from aromatic atoms during typing, thereby resulting in more pharmacophore type pairs and consequently a higher dimensionality of the descriptor

[a] M. Reutlinger,<sup>+</sup> C. P. Koch,<sup>+</sup> D. Reker,<sup>+</sup> N. Todoroff, P. Schneider, T. Rodrigues, G. Schneider  
ETH, Department of Chemistry and Applied Biosciences, Institute of Pharmaceutical Sciences  
Wolfgang-Pauli-Str. 10, CH-8093 Zurich, Switzerland  
fax: +41 44 633 13 79, tel: +41 44 633 74 38  
\*e-mail: gisbert.schneider@pharma.ethz.ch

[<sup>+</sup>] These authors contributed equally to this work.

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/minf.201200141>.

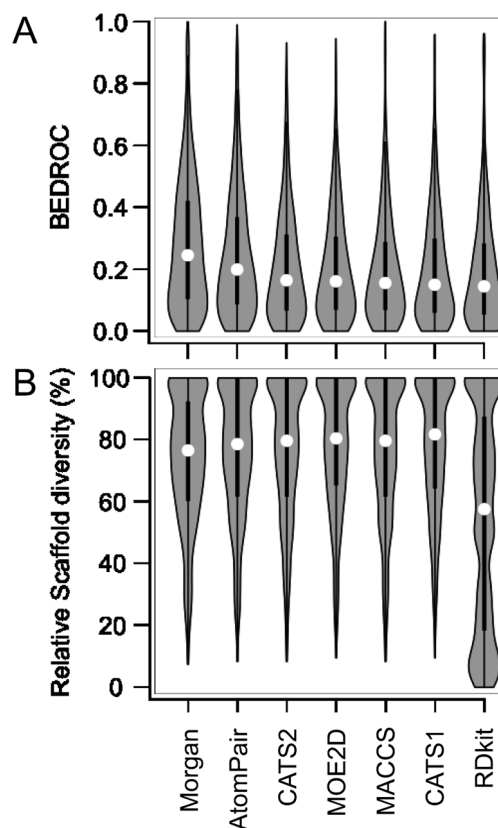


**Figure 1.** Principle of CATS descriptor calculation. The molecular structure (Step 1) is reduced to the molecular graph, and feature types are assigned (Step 2; L, lipophilic; R, aromatic; A, hydrogen-bond acceptor; D, hydrogen-bond donor). Then, atom pairs for all feature pairs are counted (Step 3), and the final descriptor values are scaled (Step 4). Here, the raw values were divided by the respective  $\lambda$  value (sum of atom type pair occurrences). Note that not all vertices in the molecular graph are considered ‘pharmacophoric’. These possess no feature types.

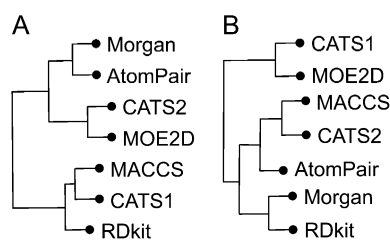
than CATS1, which lacks the aromatic atom type. For both descriptors we employed ‘types scaling’, which mitigates the potential dominance of prevalent pharmacophore feature types, and a maximal correlation distance of 10 bonds.<sup>[22]</sup> An example of CATS descriptor calculation is presented in Figure 1.

We employed the Receiver Operating Characteristic (ROC) related BEDROC score for actives-retrieval benchmarking.<sup>[15]</sup> For our study, the alpha level of the BEDROC method was set to 160.9, which corresponds to the top 1% of the screening list contributing 80% of the score. Murcko scaffold<sup>[16]</sup> diversity among the set of actives within the top 1% of respective screening lists served as measure for scaffold-hopping potential.

Albeit state-of-the-art radial fingerprints and atom-pair fingerprints outperformed CATS descriptors in terms of the number of actives retrieved (Figure 2A), the latter ratify their intent of design by delivering the overall highest ratio of diverse scaffolds among retrieved actives. Scaffold-hopping potential was determined by examining the distribution of relative scaffold diversities  $r$ , which is the ratio of differing scaffolds  $s$  to the number of retrieved actives  $n$  among the top 1% of respective screening runs. While  $s$  correlates to the BEDROC scores when comparing different descriptors,  $r$  unveils the CATS1 descriptor as the most suitable descriptor for scaffold-hopping among the compared molecular representations (Figure 2B). In terms of BEDROC scores estimating the enrichment potential, radial fingerprints (Morgan) and Carhart-type atom pairs (AtomPair) performed similar, as did the CATS2 and MOE descriptors, while MACCS, CATS1, and RDkit fingerprints formed a third group (Figure 3A). With respect to scaffold-hopping



**Figure 2.** Comparison of molecular representations for their abilities to retrieve known actives (A) and scaffolds (B) from a collection of druglike bioactive compounds (COBRA). Violin plots show the shapes (gray), medians (white circle) and quartiles (thick lines) of the distributions.

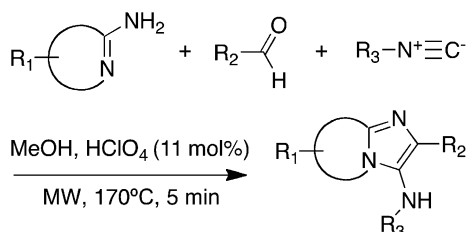


**Figure 3.** Similarity of molecular representations in terms of their enrichment (A) and scaffold-hopping potential (B). Pair-wise, one-sided Wilcoxon rank sum tests<sup>[24]</sup> were performed for the BEDROC score distributions of the descriptors. Clustering the obtained *p*-values with Ward's method<sup>[25]</sup> resulted in the depicted dendrograms.

potential, the groups vary, with CATS1 and MOE2D pairing up, as well as CATS2 and MACCS (Figure 3B). It might thus be advisable to select one method from each group for similarity searching and compare ranked results lists, e.g. by data fusion.<sup>[23]</sup> We wish to point out that the grouping of methods depicted in Figure 3 should be treated with caution, as the dendrograms are likely to vary for other reference data sets and chemotype/target coverage.

The outcome of this limited benchmark study is in agreement with a large-scale systematic analysis of 2D fingerprint methods by Sherman and co-workers, who conclude (...) *if the objective of a screen is to identify novel, diverse hits, then a less specific atom-typing scheme may be more appropriate.*<sup>[26]</sup> The CATS representation of molecular graphs and pharmacophoric features serves this purpose of finding new chemotypes. When using the descriptor, one should not expect highest possible enrichment of actives among the top-scoring virtual hits, but can anticipate surprising new ideas for synthesis and activity testing.

This intended permissiveness ('fuzziness')<sup>[27]</sup> of the CATS molecular representation, which is achieved by coarse-grained atom-typing and feature pair correlation, not only enables scaffold-hopping but may also be used for predicting mutual targets of structurally diverse bioactive ligands. Here, we started from an Ugi-type three-component combinatorial synthesis (Scheme 1)<sup>[28]</sup> and tested whether we could use CATS for 'de-orphanizing' some of the compounds by target identification. All prospective experiments were carried out with the CATS2 implementation.



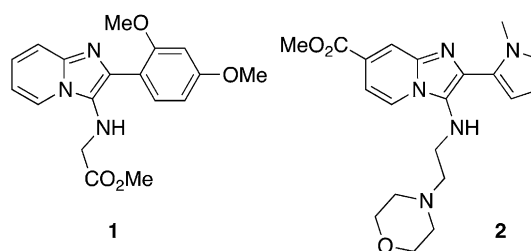
**Scheme 1.** Ugi-type three-component reaction used for constructing a virtual combinatorial library and synthesizing selected compounds.

We constructed a virtual combinatorial library from 12 aminopyridines, 40 aldehydes and 8 isocyanide building blocks, resulting in 3840 virtual products (Scheme 1 and Supporting Information). To predict potential bioactivities for these compounds we computed their CATS similarity values to known drugs and lead structures (COBRA v11.10). Briefly, we trained a self-organizing neural network (SOM, Kohonen network) on the pool of COBRA reference compounds and the virtual combinatorial products, followed by visualization of compound distributions as a two-dimensional toroidal map (Figure 4).<sup>[29,30]</sup> For the purpose of prediction, we only considered annotated targets of the reference compounds that were co-clustered with the combinatorial products. In this way, target predictions are limited to a conservative 'application domain' of a reference compound cluster, and the risk of false-positive prediction is reduced.<sup>[31]</sup> For further target prioritization, we computed *p*-values from the similarity score distribution between ligands binding to different targets (complete training data).<sup>[32]</sup> The *p*-values are an estimate of the probability of making a false-positive prediction (type-I error).

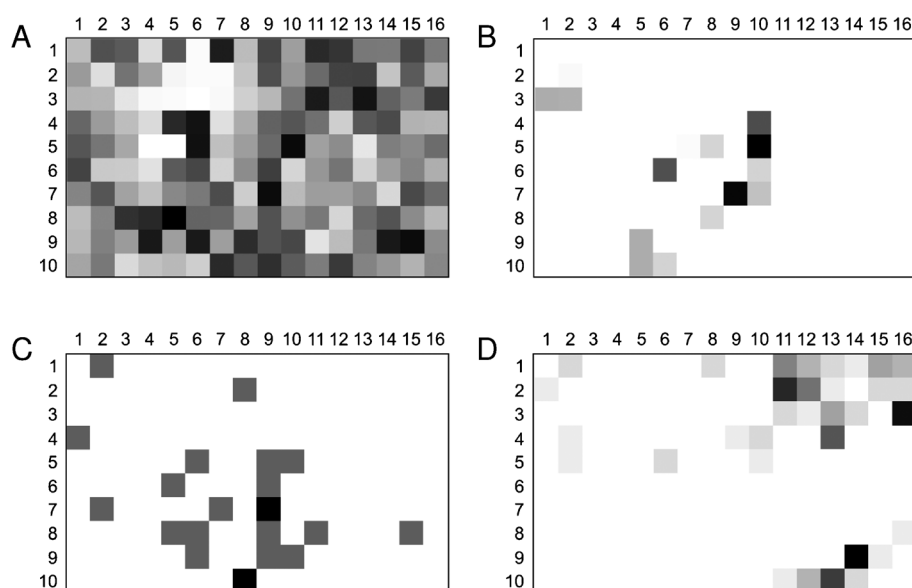
For the whole library, this method suggested six targets with average *p*-values < 0.01: phosphoinositide 3-kinase (PI3K), biphenyl-2,3-diol 1,2-dioxygenase, diacylglyceride *O*-acyltransferase, smoothed receptor, interleukin receptors, and cytochrome P450 reductase. We decided to investigate the PI3K prediction in more detail because this enzyme is a relevant drug target in antitumor research. Of note, the underlying scaffold was previously shown to afford PI3K $\alpha$  inhibitors.<sup>[33]</sup>

First, we synthesized and tested the nine top-predicted compounds for PI3K $\alpha$  inhibition. In total, four of them exhibited the desired activity. Compound 1 (Scheme 2) turned out to be the most active ( $IC_{50}$  = 131  $\mu$ M). Although the measured activities might be considered as weak, this result nevertheless proves the CATS + SOM-based approach valid for suggesting plausible macromolecular targets for small molecules.

We then synthesized and tested an additional set of 57 compounds from the virtual combinatorial library, for which the highest joint prediction scores for PI3K and DNA topoisomerases were computed. These were simply the top DNA topoisomerase hits that were also predicted to inhibit PI3K with relatively high confidence. Previous studies suggested that simultaneous inhibition of these two enzymes



**Scheme 2.** Structures of compounds 1 and 2.

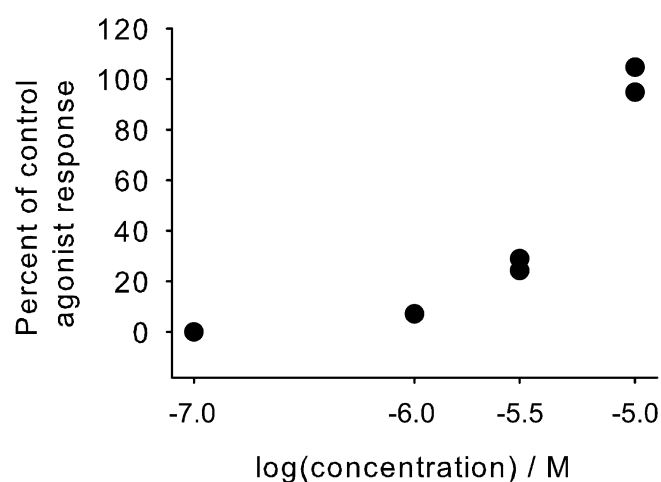


**Figure 4.** Toroidal self-organizing map (SOM) visualizing overall data density (A), distribution of the virtual combinatorial library (B), known PI3K inhibitors (C), and known muscarinic receptor ligands (D). 16×10 data clusters ('neurons', Voronoi fields) are shown as squares. Gray shading represents local compound density (note that the shading in each plot is scaled between minimal and maximal values). Compound 1 is located in cluster (9,7), compound 2 in cluster (10,5). For compound 1 an overlap with PI3K inhibitors is predicted. Compound 2 is found in a cluster that contains muscarinic receptor ligands and few PI3K inhibitors.

might allow for more efficient chemotherapy with reduced chemoresistance of tumor cells.<sup>[34]</sup> Molecules with a target profile that includes both these targets will constitute an important step in anti-cancer research. Moreover, the scaffold of our library has already been proven to produce bioactive compounds against both those targets.<sup>[33,35]</sup> In fact, in the present study six of our compounds, at a concentration of 75  $\mu\text{M}$ , turned out to be moderately active against PI3K $\alpha$ , where compound 2 (Scheme 2) was the most potent ( $IC_{50} = 230 \pm 30 \mu\text{M}$ ). We wish to point out that we cannot completely rule out measurement artifacts caused by compound aggregation.<sup>[36]</sup> None of the 57 synthesized compounds inhibited human DNA topoisomerase II (EC 5.99.1.3), but in a preliminary test four of them inhibited bacterial DNA gyrase, a bacterial type II topoisomerase (EC 5.99.1.3) (data not shown). Apparently, the scaffold of the combinatorial library positions R-group vectors appropriately, but proper side-chain functionalities are required for potency and target selectivity. There is ample opportunity for optimizing compound 2 in this regard by including additional building blocks in the combinatorial synthesis.

For comparison, we also predicted targets for the obtained PI3K $\alpha$  inhibitors using SEA<sup>[39]</sup>. In SEA, compound 2 yielded no target predictions at all when using ChEMBL<sup>[40]</sup> as reference data. For the remaining compounds SEA reported maximal Tanimoto similarity below 0.35 and  $E$ -value > 1.2, rendering them low confidence predictions. Compound 1 was suggested as ligand of quinone reductase 2 (NQO2) and melatonin receptor 1B (MTNR1B). PI3K was not reported as a potential target for compound 1 by SEA.

Finally, it is of particular note that CATS suggested human muscarinic receptor 1 (M1) ranking among the top predictions on the target list computed just for compound 2. In a first cell-based functional assay<sup>[41]</sup> compound 2, in a concentration of 10  $\mu\text{M}$ , actually exhibited substantial M1 agonistic activity yielding  $34 \pm 5\%$  of the effect caused by 100 nM acetylcholine. Follow-up concentration-dependent activity determination yielded an approximate  $EC_{50}$  of 5  $\mu\text{M}$  for compound 2 (Figure 5). This result confirms the CATS +



**Figure 5.** Concentration-dependent agonistic activity of compound 2 on the human M1 receptor. Acetylcholine served as positive control agonist ( $EC_{50} = 1.9 \text{ nM}$ ). At ligand concentrations > 10  $\mu\text{M}$  compound 2 aggregated and interfered with the measurement (data not shown).

SOM-based target prediction as viable and de-orphanizes compound **2** as a novel (no entry in CAS<sup>[42]</sup>) functional M1 receptor agonist. We would like to mention that the SOM projection shown in Figure 4D may actually serve as a guide for structure optimization,<sup>[29,37]</sup> due to the fact that compound **2** is located in a sparsely populated region of the activity island formed by known muscarinic receptor ligands. Side-chain alteration could steer the design towards the center of the distribution thus potentially improving potency.<sup>[38]</sup>

In conclusion the results of this study corroborate CATS+SOM as a useful similarity-based approach for identifying pairs of molecules with similar bioactivity but different molecular scaffolds. Inclusion of the aromatic feature type in the CATS2 implementation increased enrichment in a retrospective analysis. Results of a preliminary prospective target-profiling study demonstrate that (i) the CATS2 descriptor may be employed to predict targets of virtually generated compounds with potential applications in de novo design and drug re-purposing, (ii) relying only on a single prediction algorithm bears the danger of missing relevant drug targets or focusing on false-positive predictions, and (iii) different molecular descriptors (here: CATS2; SEA with ECFP4 fingerprints) in combination with its associated knowledge base (here: COBRA or ChEMBL) complement each other in their domains of applicability. It will therefore be worthwhile to construct a prediction tool that is based on multiple reference databases, descriptors and models, e.g. as a jury decision approach. Whether activities in the micromolar range give rise to desired poly-pharmacology effects or turn out to be actually sufficient for drug re-purposing certainly depends on the particular pharmacological activity, therapeutic area, and intended application.<sup>[43]</sup> Many more practical examples will be required to allow for a statistically motivated assessment. Irrespective of the shortcomings of each method, our study validates ligand-based target prediction as viable for rapid compound profiling in medicinal chemistry and chemical biology.

## Experimental

**Synthesis and analytics.** Chemical synthesis was performed with a Biotage Initiator microwave synthesizer (Upsala, Sweden). Aminopyridine (1.0 mol. eq.), aldehyde (1.0 mol. eq.), isocyanide (1.0 mol. eq.) and perchloric acid (11 mol%) were dissolved in EtOH (1.1 mL × mmol<sup>-1</sup>). The solution was heated at 170 °C for 5 minutes under microwave irradiation. The resulting crude product was purified via preparative HPLC using CH<sub>3</sub>CN:H<sub>2</sub>O (+0.1% trifluoroacetic acid in each phase) as eluent, in a gradient of 5–50% CH<sub>3</sub>CN run over 16 minutes, to afford compounds **1** and **2** as yellow oils.

Compound **1** (methyl 2-((2-(2,4-dimethoxyphenyl)imidazo[1,2-*a*]pyridin-3-yl)amino)acetate), 81%: <sup>1</sup>H-NMR (CD<sub>3</sub>OD, 400.13 MHz): δ 3.44 (3H, s, OCH<sub>3</sub>), 3.67 (2H, s, CH<sub>2</sub>),

3.75 (3H, s, OCH<sub>3</sub>), 3.80 (3H, s, OCH<sub>3</sub>), 6.57–6.61 (2H, m, Ar-H), 7.32–7.36 (1H, m, Ar-H), 7.53 (1H, d, *J*=8.0 Hz, Ar-H), 6.77–7.75 (2H, m, Ar-H), 8.68 (1H, d, *J*=2.4 Hz, Ar-H). <sup>13</sup>C NMR (CD<sub>3</sub>OD, 100.61 MHz): δ 48.54, 52.52, 56.18, 56.48, 99.77, 107.08, 108.44, 112.51, 117.43, 123.13, 126.42, 129.04, 132.62, 133.56, 137.46, 159.91, 164.60, 173.19. HRMS-ESI calc. (C<sub>18</sub>H<sub>19</sub>N<sub>3</sub>O<sub>4</sub>+H<sup>+</sup>): 342.1448, found: 342.1448.

Compound **2** (methyl 2-(1-methyl-1*H*-pyrrol-2-yl)-3-((2-morpholinoethyl)amino)imidazo[1,2-*a*]pyridine-7-carboxylate), 74%: <sup>1</sup>H-NMR (CD<sub>3</sub>OD, 400.13 MHz): δ 3.08 (2H, m, CH<sub>2</sub>), 3.21 (2H, t, *J*=6.4 Hz, CH<sub>2</sub>), 3.37 (2H, m, CH<sub>2</sub>), 3.45 (2H, t, *J*=6.4 Hz, CH<sub>2</sub>), 3.72 (3H, s, CH<sub>3</sub>), 3.81 (2H, m, CH<sub>2</sub>), 3.97 (2H, m, CH<sub>2</sub>), 4.05 (3H, s, CH<sub>3</sub>), 6.32 (1H, dd, *J*=3.8 Hz, Ar-H), 6.63 (1H, dd, *J*=3.8 Hz Ar-H), 7.06 (1H, m, Ar-H), 7.93 (1H, dd, *J*=1.6 and 7.2 Hz, Ar-H), 8.40 (1H, m, Ar-H), 8.80 (1H, dd, *J*=0.8 and 7.2 Hz, Ar-H). <sup>13</sup>C NMR (CD<sub>3</sub>OD, 100.61 MHz): 35.01, 41.04, 53.33, 53.84, 57.27, 64.81, 110.04, 114.52, 115.63, 116.03, 116.31, 118.78, 126.40, 127.62, 131.28, 134.24, 137.07, 165.19. HRMS-ESI calc. (C<sub>20</sub>H<sub>25</sub>N<sub>5</sub>O<sub>3</sub>+H<sup>+</sup>): 384.2030, found: 384.2031.

We used dynamic light scattering (Brookhaven 90Plus) to determine potential aggregation of compound **2** in aqueous solution with 1% DMSO. Aggregate particles were observable at concentrations ranging from 15.5–250 μM.

**Self-organizing map.** We use our software tool *molmap* for generating a toroidal SOM containing 160 clusters arranged in a 16 × 10 rectangular grid, as described previously,<sup>[31]</sup> with number of training cycles = 10<sup>6</sup> and Gaussian neighborhood radius = 8.

**CATS molecular descriptor.** Descriptor calculation was performed with a proprietary Java-based software tool (for licensing options, contact G. S.). Free online access to demonstration software is provided at URL: <http://modlab-cadd.ethz.ch/>

**Biochemical activity determination.** Activity against PI3Kα was measured by Reaction Biology Corp. (Malvern, PA, USA) in a 10-dose IC<sub>50</sub> determination (*n*=3), in the presence of 10 μM ATP. Preliminary DNA topoisomerase and gyrase inhibition tests were performed with a compound concentration of 5 mM by Inspiralis Ltd (Norwich, UK).

## Conflict of Interest Statement

G. S. is a scientific consultant to pharmaceutical industry and a co-founder of AlloCyte Pharmaceuticals Ltd, Basel.

## Acknowledgements

This research was supported by the *Swiss National Science Foundation* (Grant 205321-134783), *Deutsche Forschungsgemeinschaft* (FOR1406TP4), and *OPO-Foundation Zürich*. The *Chemical Computing Group Inc.* (Montreal, Canada) provided a research license of MOE.

## References

- [1] F. Pammolli, L. Magazzini, M. Riccaboni, *Nat. Rev. Drug Discov.* **2011**, *10*, 428–438.
- [2] S. M. Paul, D. S. Mytelka, C. T. Dunwiddie, C. C. Persinger, B. H. Munos, S. R. Lindborg, A. L. Schacht, *Nat. Rev. Drug Discov.* **2010**, *9*, 203–214.
- [3] G. Schneider, *Nat. Rev. Drug Discov.* **2010**, *9*, 273–276.
- [4] a) P. Willett, *Methods Mol. Biol.* **2011**, *672*, 133–158; b) M. Vogt, J. Bajorath, *Methods Mol. Biol.* **2011**, *672*, 159–173; c) E. J. Gardiner, J. D. Holliday, C. O'Dowd, P. Willett, *Future Med. Chem.* **2011**, *3*, 405–414; d) S. Renner, G. Schneider, *ChemMedChem* **2006**, *1*, 181–185; e) R. Garcia-Serna, J. Mestres, *Expert Opin. Drug Metab. Toxicol.* **2010**, *6*, 1253–1263.
- [5] V. Poroikov, D. Filimonov, A. Lagunin, T. Glorizova, A. B. Zakharov, *SAR QSAR Environ. Res.* **2007**, *18*, 101–110.
- [6] a) F. M. Areias, J. Brea, E. Gregori-Puigjané, M. E. Zaki, M. A. Carvalho, E. Domínguez, H. Gutiérrez-de-Terán, M. F. Proença, M. I. Loza, J. Mestres, *J. Bioorg. Med. Chem.* **2010**, *18*, 3043–3052; b) E. Gregori-Puigjané, J. Mestres, *Comb. Chem. High Throughput Screen.* **2008**, *11*, 669–676; c) E. Gregori-Puigjané, J. Mestres, *Curr. Opin. Chem. Biol.* **2008**, *12*, 359–365.
- [7] E. Lounkine, M. J. Keiser, S. Whitebread, D. Mikhailov, J. Hamon, J. L. Jenkins, P. Lavan, E. Weber, A. K. Doak, S. Côté, B. K. Shoichet, L. Urban, *Nature* **2012**, *486*, 361–367.
- [8] a) G. Schneider, W. Neidhart, T. Giller, G. Schmid, *Angew. Chem. Int. Ed.* **1999**, *38*, 2894–2896; b) U. Fechner, L. Franke, S. Renner, P. Schneider, G. Schneider, *J. Comput. Aided Mol. Des.* **2003**, *17*, 687–98.
- [9] G. Schneider, P. Schneider, S. Renner, *QSAR Comb. Sci.* **2006**, *25*, 1162–1171.
- [10] a) A. Koutsoukas, B. Simms, J. Kirchmair, P. J. Bond, A. V. Whitmore, S. Zimmer, M. P. Young, J. L. Jenkins, M. Glick, R. C. Glen, A. Bender, *J. Proteomics* **2011**, *74*, 2554–2574; b) J. Achenbach, P. Tiikkainen, L. Franke, E. Proschak, *Future Med. Chem.* **2011**, *3*, 961–968.
- [11] P. Schneider, G. Schneider, *QSAR Comb. Sci.* **2003**, *22*, 713–718.
- [12] P. Willett, *Drug Discov. Today* **2006**, *11*, 1046–1053.
- [13] A. M. Wassermann, J. Bajorath, *Expert Opin. Drug Discov.* **2011**, *6*, 683–687.
- [14] C. Kramer, T. Kalliokoski, P. Gedeck, A. Vulpetti, *J. Med. Chem.* **2012**, *55*, 5165–5173.
- [15] J. F. Truchon, C. I. Bayly, *J. Chem. Inf. Model.* **2007**, *47*, 488–508.
- [16] G. W. Bemis, M. A. Murcko, *J. Med. Chem.* **1996**, *39*, 2887–2893.
- [17] D. Rogers, M. Hahn, *J. Chem. Inf. Model.* **2010**, *50*, 742–754.
- [18] R. E. Carhart, D. H. Smith, R. Venkataraghavan, *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 64–73.
- [19] MACCS-II, MDL Information Systems/Symyx, Santa Clara, CA, **1984**.
- [20] *Fingerprint Toolkit*, Daylight Chemical Information Systems Inc., ch. 6: *Fingerprints - Screening and Similarity*; [http://www.daylight.com/products/finger\\_kit.html](http://www.daylight.com/products/finger_kit.html).
- [21] G. Landrum, *RDKit: Open-Source Cheminformatics*; <http://www.rdkit.org>.
- [22] U. Fechner, G. Schneider, *QSAR Comb. Sci.* **2004**, *23*, 19–22.
- [23] J. D. Holliday, E. Kanoulas, N. Malim, P. Willett, *J. Cheminform.* **2011**, *3*, 29.
- [24] F. Wilcoxon, *Biometrics Bull.* **1945**, *1*, 80–83.
- [25] J. H. Ward Jr, *J. Am. Stat. Assoc.* **1963**, *58*, 236–244.
- [26] M. Sastry, J. F. Lowrie, S. L. Dixon, W. Sherman, *J. Chem. Inf. Model.* **2010**, *50*, 771–784.
- [27] a) A. Klenner, M. Hartenfeller, P. Schneider, G. Schneider, *Drug Discov. Today Technol.* **2010**, *7*, e237–e244; b) G. Schneider, *Drug Discov. Today Technol.* **2012**, *9*, <http://dx.doi.org/10.1016/j.ddtec.2012.06.001>.
- [28] H. Bienaymé, K. Bouzid, *Angew. Chem. Int. Ed.* **1998**, *37*, 2234–2237.
- [29] a) G. Schneider, M. Nettekoven, *J. Comb. Chem.* **2003**, *5*, 233–237; b) P. Schneider, Y. Tanrikulu, G. Schneider, *Curr. Med. Chem.* **2009**, *16*, 258–266.
- [30] P. Schneider, K. Stutz, L. Kasper, S. Haller, M. Reutlinger, F. Reisen, T. Geppert, G. Schneider, *Pharmaceuticals* **2011**, *4*, 1236–1247.
- [31] a) G. Schneider, Y. Tanrikulu, P. Schneider, *Future Med. Chem.* **2009**, *1*, 213–218; b) M. Schmuker, M. de Bruyne, M. Hähnel, G. Schneider, *Chem. Cent. J.* **2007**, *1*, 11; c) M. Schmuker, G. Schneider, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 20285–20289.
- [32] a) V. Hähnel, N. Todoroff, T. Rodrigues, G. Schneider, *Future Med. Chem.* **2012**, *4*, 1897–1906; b) G. Schneider, *ACS National Meeting, San Diego, CA, USA, March 25–29, 2012*.
- [33] O. Kim, Y. Jeong, H. Lee, S. S. Hong, S. Hong, *J. Med. Chem.* **2011**, *54*, 2455–2466.
- [34] H. G. Yu, Y. W. Ai, L. L. Yu, X. D. Zhou, J. Liu, J. H. Li, X. M. Xu, S. Liu, J. Chen, J. F. Liu, *Int. J. Cancer* **2008**, *122*, 433–443.
- [35] A. T. Baviskar, C. Madaan, R. Preet, P. Mohapatra, V. Jain, A. Agarwal, S. K. Guchhait, C. N. Kundu, U. C. Banerjee, P. V. Bharatam, *J. Med. Chem.* **2011**, *54*, 5013–5030.
- [36] K. E. Coan, B. K. Shoichet, *J. Am. Chem. Soc.* **2008**, *130*, 9606–9612.
- [37] a) T. Noeske, A. Jirgensons, I. Starchenkova, S. Renner, I. Jaunzeme, D. Trifanova, M. Hechenberger, T. Bauer, V. Kaus, C. G. Parsons, G. Schneider, T. Weil, *ChemMedChem* **2007**, *2*, 1763–1773; b) S. Renner, M. Hechenberger, T. Noeske, A. Böcker, C. Jatzke, M. Schmuker, C. G. Parsons, T. Weil, G. Schneider, *Angew. Chem. Int. Ed. Engl.* **2007**, *46*, 5336–5339; c) R. Steri, J. Achenbach, D. Steinhilber, M. Schubert-Zsilavec, E. Proschak, *Biochem. Pharmacol.* **2012**, *83*, 1674–1681.
- [38] a) M. Reutlinger, W. Guba, R. E. Martin, A. I. Alanine, T. Hoffmann, A. Klenner, J. A. Hiss, P. Schneider, G. Schneider, *Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 11633–11636; b) G. Schneider, M. Hartenfeller, M. Reutlinger, Y. Tanrikulu, E. Proschak, P. Schneider, *Trends Biotechnol.* **2009**, *27*, 18–26.
- [39] M. J. Keiser, B. L. Roth, B. N. Armbruster, P. Ernsberger, J. J. Irwin, B. K. Shoichet, *Nat. Biotech.* **2007**, *25*, 197–206.
- [40] A. Gaulton, L. Bellis, J. Chambers, M. Davies, A. Hersey, Y. Light, S. McGlinchey, R. Akhtar, F. Atkinson, A. P. Bento, B. Al-Lazikani, D. Michalovich, J. P. Overington, *Nucl. Acids Res.* **2012**, *40*, D1100–1107.
- [41] C. Sur, P. J. Mallorga, M. Wittmann, M. A. Jacobson, D. Pascarella, J. B. Williams, P. E. Brandish, D. J. Pettibone, E. M. Scolnick, P. J. Conn, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 13674–13679.
- [42] *SciFinder*, Chemical Abstracts Service: Columbus, OH, **2012**, RN 58-08-02 (accessed Nov. 9, 2012).
- [43] a) T. I. Oprea, J. Mestres, *AAPS J.* **2012**, *14*, 759–763; b) X. Jalcencas, J. Mestres, *Med. Chem. Commun.* **2013**, *4*, 80–87.

Received: November 13, 2012

Accepted: January 18, 2013

Published online: February 7, 2013