

Accepted Manuscript

Optimizing TRPM4 inhibitors in the MHFP6 chemical space

Clémence Delalande, Mahendra Awale, Matthias Rubin, Daniel Probst, Lijo C. Ozhathil, Jürg Gertsch, Hugues Abriel, Jean-Louis Reymond



PII: S0223-5234(19)30060-1

DOI: <https://doi.org/10.1016/j.ejmech.2019.01.048>

Reference: EJMECH 11055

To appear in: *European Journal of Medicinal Chemistry*

Received Date: 24 October 2018

Revised Date: 18 December 2018

Accepted Date: 19 January 2019

Please cite this article as: Clé. Delalande, M. Awale, M. Rubin, D. Probst, L.C. Ozhathil, Jü. Gertsch, H. Abriel, J.-L. Reymond, Optimizing TRPM4 inhibitors in the MHFP6 chemical space, *European Journal of Medicinal Chemistry* (2019), doi: <https://doi.org/10.1016/j.ejmech.2019.01.048>.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Optimizing TRPM4 Inhibitors in the MHFP6 Chemical Space

Clémence Delalande,^{a)} Mahendra Awale,^{a)} Matthias Rubin,^{b)} Daniel Probst,^{a)} Lijo C. Ozhathil,^{b)}

Jürg Gertsch,^{b)} Hugues Abriel^{b)} and Jean-Louis Reymond^{a)*}

^{a)} *Department of Chemistry and Biochemistry, National Center for Competence in Research NCCR TransCure, University of Berne, Freiestrasse 3, 3012 Bern, Switzerland*

^{b)} *Institute of Biochemistry and Molecular Medicine, National Center of Competence in Research NCCR TransCure, University of Bern, Bühlstrasse 28, 3012 Bern, Switzerland*

e-mail: jean-louis.reymond@dcb.unibe.ch

Abstract

We recently reported 4-chloro-2-(2-chlorophenoxy)acetamido)benzoic acid (CBA) as the first potent inhibitor of TRPM4, a cation channel implicated in cardiac diseases and prostate cancer. Herein we report a structure-activity relationship (SAR) study of CBA resulting in two new potent analogs. To design and interpret our SAR we used interactive color-coded 3D-maps representing similarities between compounds calculated with MHFP6 (MinHash fingerprint up to six bonds), a new molecular fingerprint outperforming other fingerprints in benchmarking virtual screening studies. We further illustrate the general applicability of our method by visualizing the structural diversity of active compounds from benchmarking sets in relation to decoy molecules and to drugs. MHFP6 chemical space 3D-maps might be generally helpful in designing, interpreting and communicating the results of SAR studies. The modified WebMolCS is accessible at <http://gdb.unibe.ch> and the code is available at <https://github.com/reymond-group/webMolCS> for off-line use.

Keywords: Ion channels, chemical space, structure-activity relationships, cheminformatics, virtual screening

Introduction

We recently reported the halogenated anthranilic acid derivative 4-chloro-2-(2-chlorophenoxy)acetamido)benzoic acid (CBA) as the first potent and selective inhibitor of TRPM4,¹ a calcium-activated nonselective cation channel expressed in many tissues and implicated in cardiac diseases² and prostate cancer.³ CBA shows reversible inhibition of TRPM4 overexpressed in HEK293 cells, and blocks TRPM4-mediated currents in human prostate cancer cells. CBA also acts as chemical chaperone to rescue the expression and activity of a loss-of-function TRPM4 genetic variant. We discovered CBA by selecting analogs of the known but weak and unselective inhibitors 9-phenanthrol,⁴ glibenclamide and flufenamic acid from commercial catalogs using a previously validated 3D-shape and pharmacophore similarity method,^{5,6} and screening them in a cell-based sodium uptake assay. The initial screening campaign also identified 4-chloro-2-(2-(4-chloro-2-methylphenoxy)propanamido) benzoic acid (LBA) with a lactic acid linker as potent TRPM4 inhibitor (Figure 1).

Analysis of the structural diversity of compounds used for the discovery of CBA using WebMolCS,⁷ a web-based application for creating interactive 3D-representations of chemical space (3D-maps) available at <http://gdb.unibe.ch>, had shown that CBA resided in a region of chemical space which only contained few commercial compounds and could only be explored by additional chemical synthesis. Initial structure-activity relationship (SAR) studies focused on variations of the 2-chlorophenoxy group, which uncovered 4-chloro-2-(1-naphthyloxyacetamido)benzoic acid (NBA) with an even stronger activity than CBA.¹ We also prepared several isosteres of the carboxyl group, however these efforts did not yield any active analogs.

Herein we report an additional SAR study leading to the discovery of two new TRPM4 inhibitors (**8** and **9**) with potency comparable to CBA. To design and interpret this new SAR study we used a modified version of WebMolCS displaying chemical space 3D-maps calculated using MHFP6 (MinHash fingerprint up to six bonds), a new molecular fingerprint using the principle of

circular substructures,^{8,9} and recently reported by our group.¹⁰ Molecular fingerprints are multidimensional vector representations of molecular structures used to quantify similarities between molecules.¹¹⁻¹⁴ Such comparisons can be calculated extremely rapidly, and can uncover similarities that are not easily perceived by visual inspection of molecules, although they are mostly valid in the perspective of drug design, as assessed by benchmarking studies.¹⁵ In these benchmarking studies MHFP6 outperforms ECFP4 (extended connectivity fingerprint up to four bonds),¹⁶ which was up to now the best fingerprint across a large number of different molecular fingerprints. We were therefore interested to test the suitability of MHFP6 for designing SAR studies.

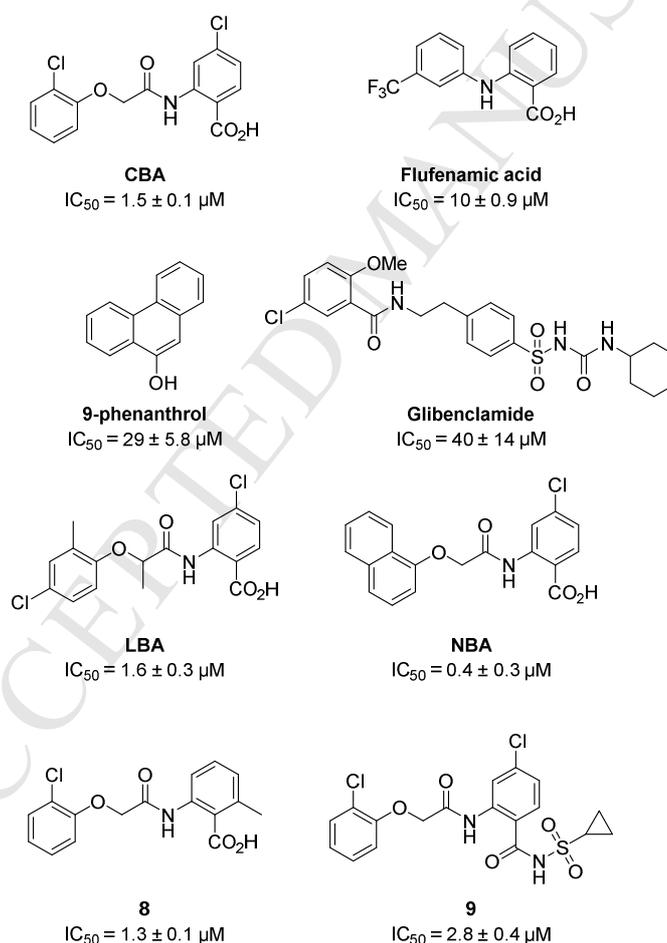


Figure 1. Structure and activity of TRPM4 inhibitors.

In the present report, we first modified our web-based application WebMolCS by enabling 3D-representations of the MHFP6 chemical space. We then computed a 3D-map of the MHFP6

chemical space representing the initial set of compounds used to discover our TRPM4 inhibitor CBA and used the map to select new analogs. We discuss the synthesis and testing of these analogs leading to the discovery of the new inhibitors **8** and **9**, and how MHFP6 chemical space 3D-maps provide an overview of the SAR study. Finally, we show that MHFP6 chemical space 3D-maps can be useful in a general context by illustrating structural differences among target specific compound series from the directory of useful decoys.¹⁷ MHFP6 chemical space 3D-maps might be generally helpful in designing, interpreting and communicating the results of SAR studies by providing overviews that are difficult to gain from R-group variation tables. The modified WebMolCS is accessible at <http://gdb.unibe.ch> and the code is available at <https://github.com/reymond-group/webMolCS> for off-line use.

Results and Discussion

Extending WebMolCS with MHFP6 chemicals space 3D-maps

In WebMolCS the user can load a series of up to 5,000 molecules written as a list of SMILES annotated with a name and optionally a property value such as an activity, and then obtain an interactive 3D-chemical space representation of the dataset (3D-map) within a few minutes. In these 3D-maps, molecules are represented as spheres colored-coded from red to blue either by similarity rank to the first molecule in the list (FP-Rank), the rank in the list of molecules (List-Rank), or the user-defined property value (ActualVals). The spheres, whose corresponding molecular structure is shown on mouse-over, are positioned at coordinates obtained by projecting a high-dimensional chemical space defined by a molecular fingerprint into 3D by dimensionality reduction. The spheres can be optionally connected with each other by a minimum spanning tree to facilitate the perception of nearest neighbor relationships, and the PC axes can be optionally displayed.

WebMolCS uses principal component analysis (PCA) as the dimensionality reduction method to project simple fingerprint spaces such as MQN (Molecular Quantum Numbers, 42D),^{18, 19} SMIfp (SMILES fingerprint, 34D)²⁰ or APfp (Atom-Pair fingerprint, 20D),²¹ because PCA can be

parallelized and is fast even for millions of molecules, as recently illustrated with our web-based visualization application Fearun.²² However, PCA is not possible with MHFP6 because this fingerprint is a set rather than a vector of bit indices, and can only serve to compute similarities between two molecules.¹⁰ We therefore implemented similarity mapping to project the MHFP6 chemical space into 3D. In similarity mapping, one first computes an N-dimensional similarity space whose N dimensions are the similarity values to N randomly selected reference molecules, and then performs a PCA of this similarity fingerprint to generate a 2D- or 3D-representation (see methods for details).²³⁻²⁵ The approach is interesting because the calculation of similarity maps is much faster than other dimensionality reduction methods for visualizing chemical space,²⁶⁻²⁸ and is therefore applicable to very large datasets.

Remarkably, similarity mapping allows to map chemical spaces for high-dimensional fingerprints that do not project well by PCA, including ECFP4 and MHFP6. In such cases, the choice of reference molecules for the similarity fingerprint is critical. In our preliminary studies with MHFP6, we found that similarity maps calculated using only actives as references, as implemented previously in 2D-maps of various sets of actives in relation to ChEMBL,²⁵ are straightforward to interpret because the first three principal components, which define the three dimensions of the map, usually cover over 80 % of data variance. In many cases PC1 itself covers 50-70% of data variance, reflecting the fact that most series of active compounds feature small and incremental structural variations. We therefore modified WebMolCS to read an optional fourth data column in the SMILES list of molecules called “satellite option”, which allows one to mark if a molecule should be used as reference (value = 1) or not (value = 0) for constructing a similarity map.

The creation of MHFP6 chemical space 3D-maps, as implemented in the modified version of WebMolCS including the satellite option to mark actives as references, is presented in detail in the methods section. Its use is illustrated in the SAR study and analysis of benchmarking sets presented below.

MHFP6 chemicals space maps of the initial TRPM4 screening set and SAR design

Using the modified WebMolCS described above, we computed a MHFP6 chemical space 3D-map featuring all compounds investigated during the discovery of our TRPM4 inhibitor CBA using only active compounds as references. The map color-coded by compound activity values and oriented with the main PC1 axis pointing to the right is shown in Figure 2. In this image, CBA appears at front right. The insert on top left shows the same map color coded by MHFP6 similarity rank to CBA, which illustrates that the PC1 axis reflects structural differences separating 9-phenanthrol and glibenclamide (far left) from CBA and analogs (far right). CBA and analogs are anthranilic acid amides related to flufenamic acid, which appears in the center of the map.

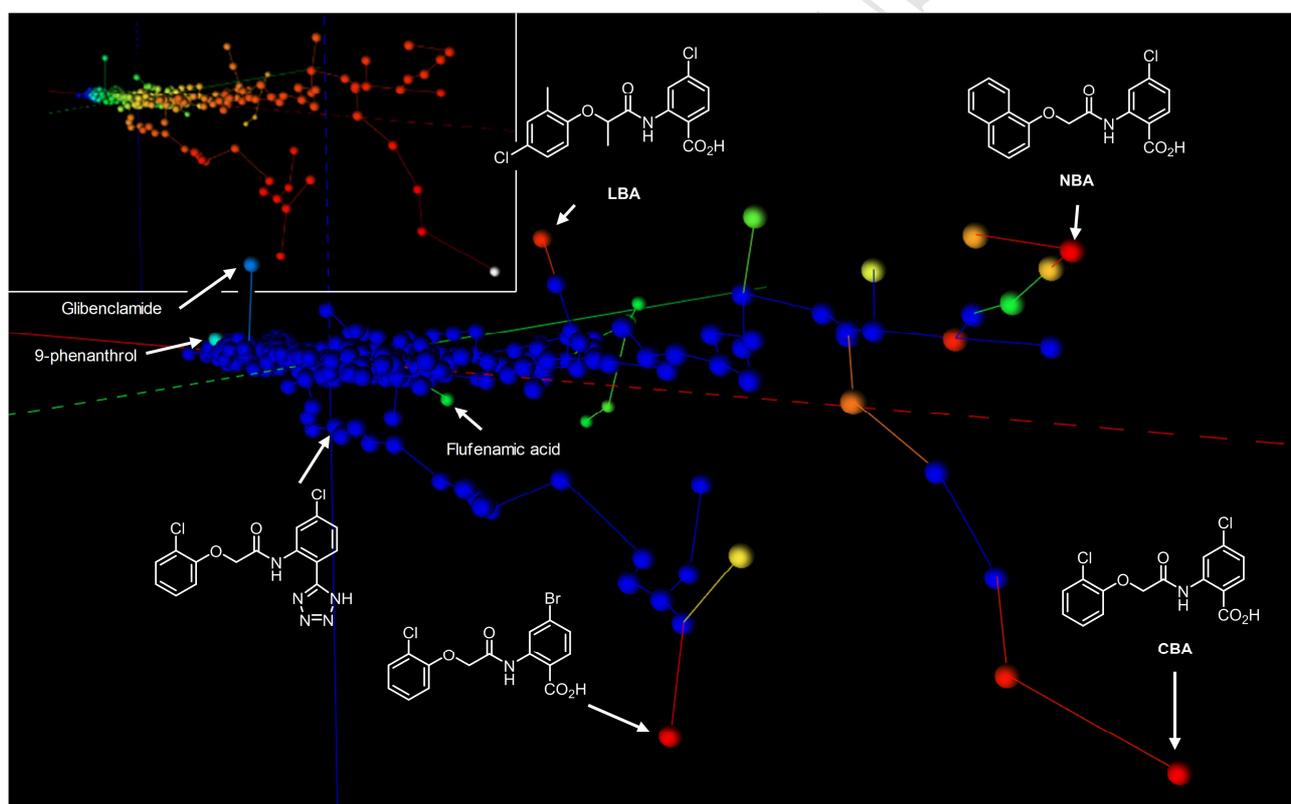


Figure 2. MHFP6 chemical space map of initial screening set leading to CBA. Main map: map color-coded by activity (blue = inactive, red = most active). Variance covered: red axis PC1 = 79 %, blue axis PC2 = 8 %, green axis PC3 = 3 %. Insert upper left: the same view color-coded by MHFP6 similarity rank (rank 1 = red, last rank = blue). The interactive map is available at <http://gdbtools.unibe.ch:8080/webMolCS/yourSIM.html?jobID=1544785130174&fp=MHFP>

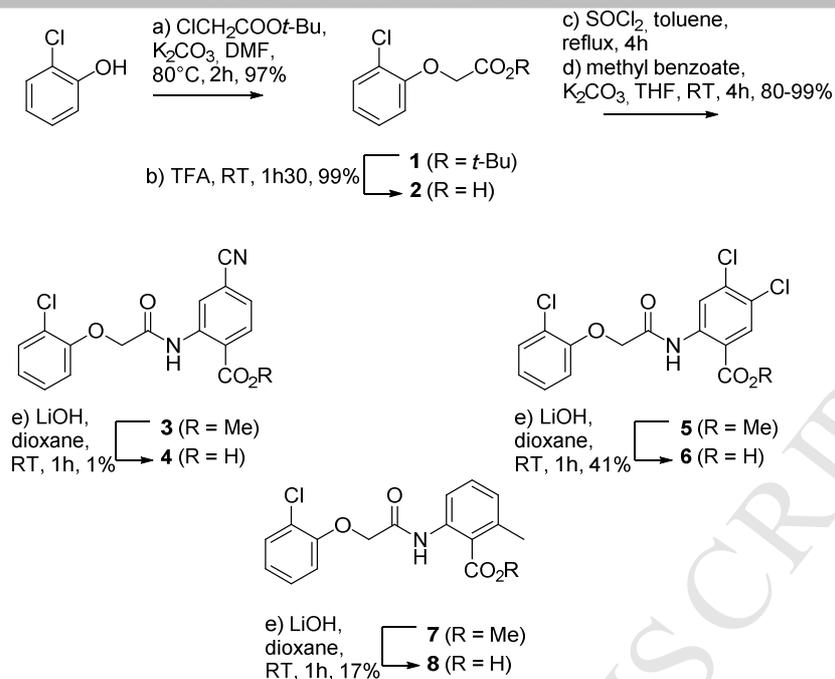
The 3D-map shows that our initial SAR covered a diversity of analogs of the 2-chlorophenoxy group that resulted in the identification of NBA (2-(1-naphthyloxyacetamido)-4-chloro-benzoic

acid) as a potent analog of CBA (Figure 2, upper right). The map also shows that only few active compounds are found among analogs featuring variations of the anthranilic acid moiety (Figure 2, lower center). Among variations of the 4-chloro substituent, only the 4-bromo analog of CBA showed good activity, while other substituents (F, Me, CF₃) at that position led to inactive analogs. Furthermore, any replacement of the carboxyl group abolished the activity including isosteres such as tetrazole (Figure 2, lower left).

Despite of this apparent lack of activity, the total number of anthranilic acid variations surveyed in our initial study was still low, suggesting to consider additional cases accessible either from commercial building blocks or by synthetic modifications. We set out to explore the chemical space around the bromo analog of CBA by substituting the bromo substituent with larger groups. Furthermore, despite of the inactivity of carboxylic isosteres, we aimed for an additional attempt with an *N*-sulfonyl carboxamide analog, a type of carboxyl group replacement that can be used to improve pharmacokinetic properties.^{29, 30} Note that the map also points to a relative lack of analogs around the lactic acid linker analogs LBA (top left of the image in Figure 4), however we decided not to pursue LBA further because both enantiomers had similar activities and the compound was prone to racemization.¹

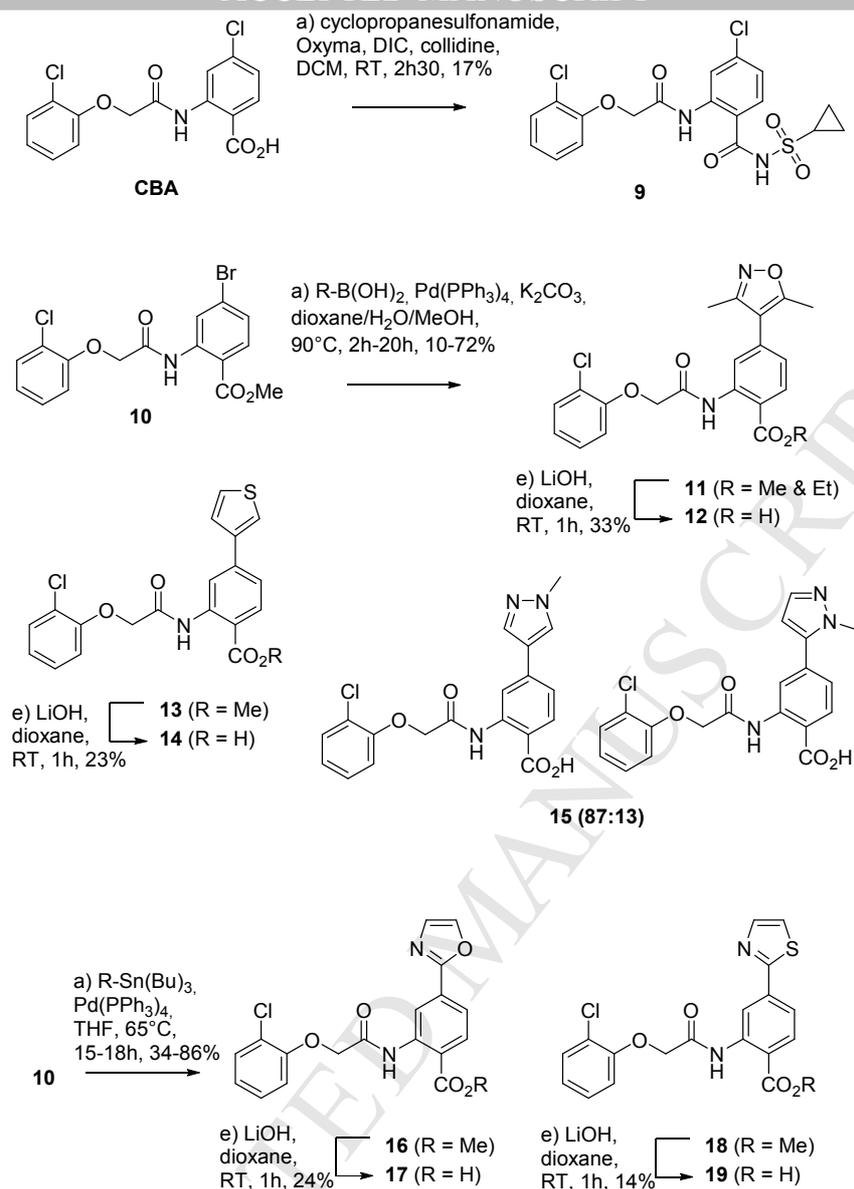
Synthesis

Alkylation of 2-chlorophenol with *tert*-butyl chloroacetate yielded ester **1**, which was deprotected by acidic treatment to **2**, and coupled with commercially available anthranilic acid esters to yield intermediates **3**, **5** and **7**. Basic hydrolysis of the methyl esters using lithium hydroxide yielded the 4-cyano (**4**), 4,5-dichloro (**6**) and 6-methyl (**8**) analogs of CBA as final products in moderate yields due to the competing hydrolysis of the anthranilic amide bond to both acidic and basic ester deprotection conditions (Scheme 1).



Scheme 1. Synthesis of CBA analogs from commercially available anthranilic esters.

Further derivatives were prepared by modification of CBA and close analogs. First, CBA itself was converted by direct coupling to the corresponding *N*-cyclopropylsulfonyl carboxamide **9** (Scheme 2). Furthermore, palladium catalyzed (Suzuki or Stille) coupling of the aromatic bromide in methyl ester **10** with boronic acid or tributyl tin derivatives of five-membered aromatic heterocycles gave ester intermediates **11**, **13**, **16** and **18**, which were subsequently hydrolyzed as above to yield carboxylic acids **12**, **14**, **17** and **19**. Product **15** was obtained directly as the free carboxylic acid during the palladium-coupling step. All final products showed >95% purity (analytical HPLC).



Scheme 2. Synthesis of CBA analogs by derivatization.

TRPM4 inhibition

We measured the activity of the new analogs using a previously developed cell-based assay.¹ In this assay, HEK293 cells overexpressing TRPM4 are first depleted of Na⁺, and then loaded with the Na⁺ sensitive dye ANG-II using the acetoxymethyl (ACM) protected cell permeable precursor.

Intracellular esterases cleave the ACM groups to release the active sensor, which remains trapped inside cells. Exposure to 140 mM Na⁺ and 10 μM ionomycin to trigger Ca²⁺ influx and activation of TRPM4, results in Na⁺ influx into the cells that can be followed in real time by the increase in

ANG-II fluorescence (Figure 3). Screening of the nine newly synthesized analogs at 10 μM showed significant inhibition with all compounds except **7** and **14**. We then determined IC_{50} values for the seven active inhibitors and performed duplicate determinations for the two most active inhibitors **8** and **9** (Table 1).

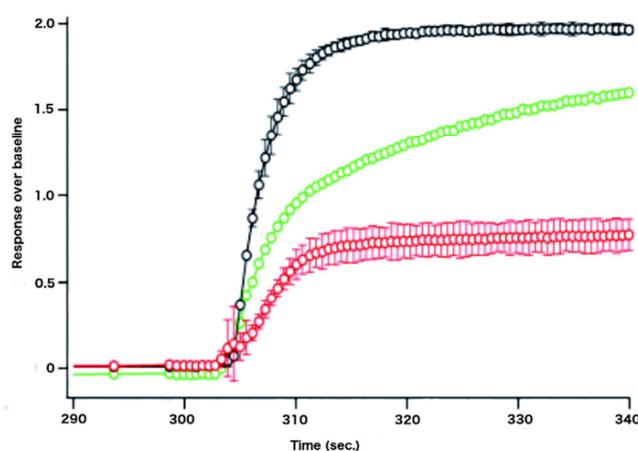
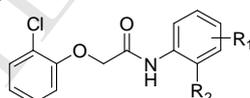


Figure 3. Fluorescent cell-based assay to determine TRPM4 modulation, measured in TRPM4 overexpressing cells. Red line: negative control (140 mM Na^+ , no ionomycin), black line: positive control (140 mM Na^+ , 10 μM ionomycin), green line: inhibitor (140 mM Na^+ , 10 μM ionomycin, 25 μM 9-phenanthrol).

Table 1. Inhibition of TRPM4 by CBA analogs.^{a)}



Cpd.	R ¹	R ²	IC ₅₀
6	4-cyano	COOH	3.6 μM
7	4,5-dichloro	COOH	30% inhibition at 10 μM
8	6-methyl	COOH	1.3 \pm 0.1 μM
9	4-chloro	CONHSO ₂ cProp	2.8 \pm 0.4 μM
14	3,5-dimethylisoxazol-4-yl	COOH	no inhibition up to 100 μM
15	thiophen-3-yl	COOH	5.9 μM
16	1-methyl-1H-pyrazol-4-yl & 1-methyl-1H-pyrazol-5-yl	COOH	9.1 μM
19	oxazol-2-yl	COOH	4.9 μM
20	thiazol-2-yl	COOH	4.7 μM

^{a)} Compound inhibition was measured in a cell-based assay. Concentration-response curve was constructed for all active inhibitors except **7** and the values were fitted with Hill equation to extrapolate IC_{50} values. Duplicates were performed for **8**, **9** and data are given as mean \pm sem.

Interpreting SAR results using a MHFP6 chemical space 3D-map

To gain an overview of the overall SAR study, we computed a new MHFP6 chemical space 3D-map using WebMolCS including all molecules used in the previous map combined with the newly designed analogs, again using only active compounds for the similarity map calculation. This new map is similar to the map in Figure 2 and is shown in a similar layout color-coded by compound activity (Figure 4). In this view, CBA analogs with variations of the 2-chlororophenoxy group appear at front right and CBA analogs with variation of the anthranilic acid appear in the lower central part of the image. The insert at upper right shows the same map color-coded by MHFP6 similarity rank to CBA, which illustrates the gradient of similarity along the PC1 axis from CBA itself at front right to 9-phenanthrol at far left.

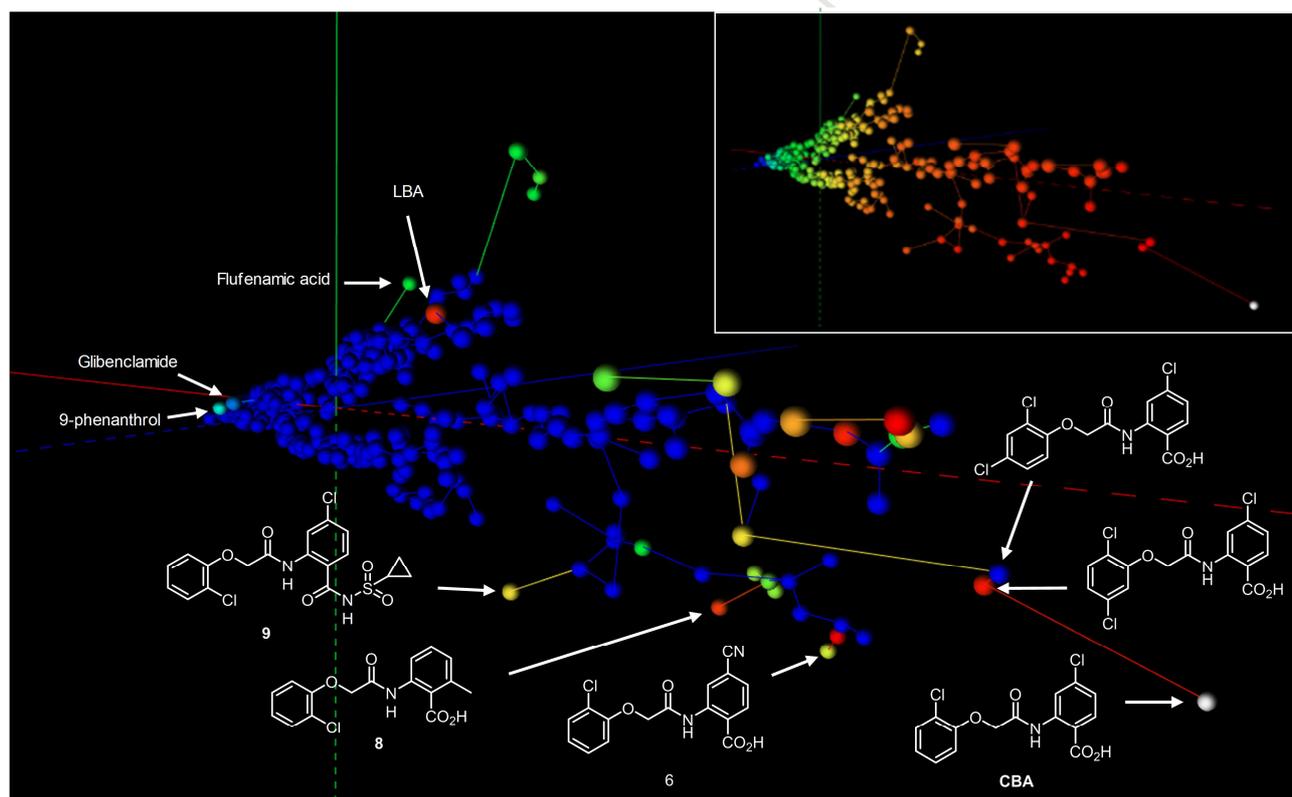


Figure 4. MHFP6 chemical space map of the overall SAR of TRPM4 analogs. Main map: map color-coded by activity (blue = inactive, red = most active). Variance covered: red axis PC1 = 78 %, blue axis PC2 = 8 %, green axis PC3 = 4 %. Insert upper right: the same view color-coded by MHFP6 similarity rank (rank 1 = red, last rank = blue). The interactive map is available at <http://gdbtools.unibe.ch:8080/webMolCS/yourSIM.html?jobID=1544605933001&fp=MHFP>

In the map in Figure 4, all newly synthesized analogs are connected by the spanning tree between the cyano analog **6**, the 6-methyl analog **8** and the N-sulfonylcarboxamide **9**. The map thus illustrates that the newly synthesized analogs improved coverage of chemical space in that region of the map, as suggested by their design. The map furthermore shows that exploring chemical space around the bromo analog (red dot connected to **6**) only uncovered less potent analogs such as the cyano derivative **6** and the heteroaromatic analogs **15**, **16**, **19** and **20** (shown as green dots around compound **8**). The map also points to the fact that TRPM4 inhibition is very sensitive to the precise substitution pattern. For example, the activity of CBA is preserved in its 2,5-dichlorophenoxy analog but disappears in the 2,4-dichlorophenoxy analog despite of their high similarity indicated by their nearest neighbor relationship in the 3D-map.

Visualizing actives, decoys and drugs in the MHFP6 chemical space

MHFP6 chemical space 3D-maps might be generally useful to inspect series of target specific actives. This is illustrated here with datasets from the directory of useful decoys, DUD-E,¹⁷ which groups series of molecules active on the same target mixed with decoy molecule with similar overall physico-chemical properties. MHFP6 chemical space 3D-maps show that series of actives from DUD-E typically consist in three to four different series of analogs, each featuring molecules with a specific substructure pattern and a range of different substituents, as illustrated here for inhibitors of thymidine kinase (kith), catechol-O-methyl transferase (comt), focal adhesion kinase (fak1), and farnesyl diphosphate synthase (fpps) (Figure 5).

In all cases, the decoys are well separated from the actives, reflecting the performance of MHFP6 in virtual screening benchmarks. The strong separation between actives and decoys in Figure 5 contrasts to poor separation between actives and inactives in our TRPM4 inhibition study (Figures 2 and 4). This difference reflects the fact that decoy molecules in benchmarking sets are molecules taken at random in a diverse pool of compounds only preserving the overall physico-

chemical properties of the actives but not their detailed substructures, while our TRPM4 series features many compounds sharing common scaffolds and small changes in substituents.

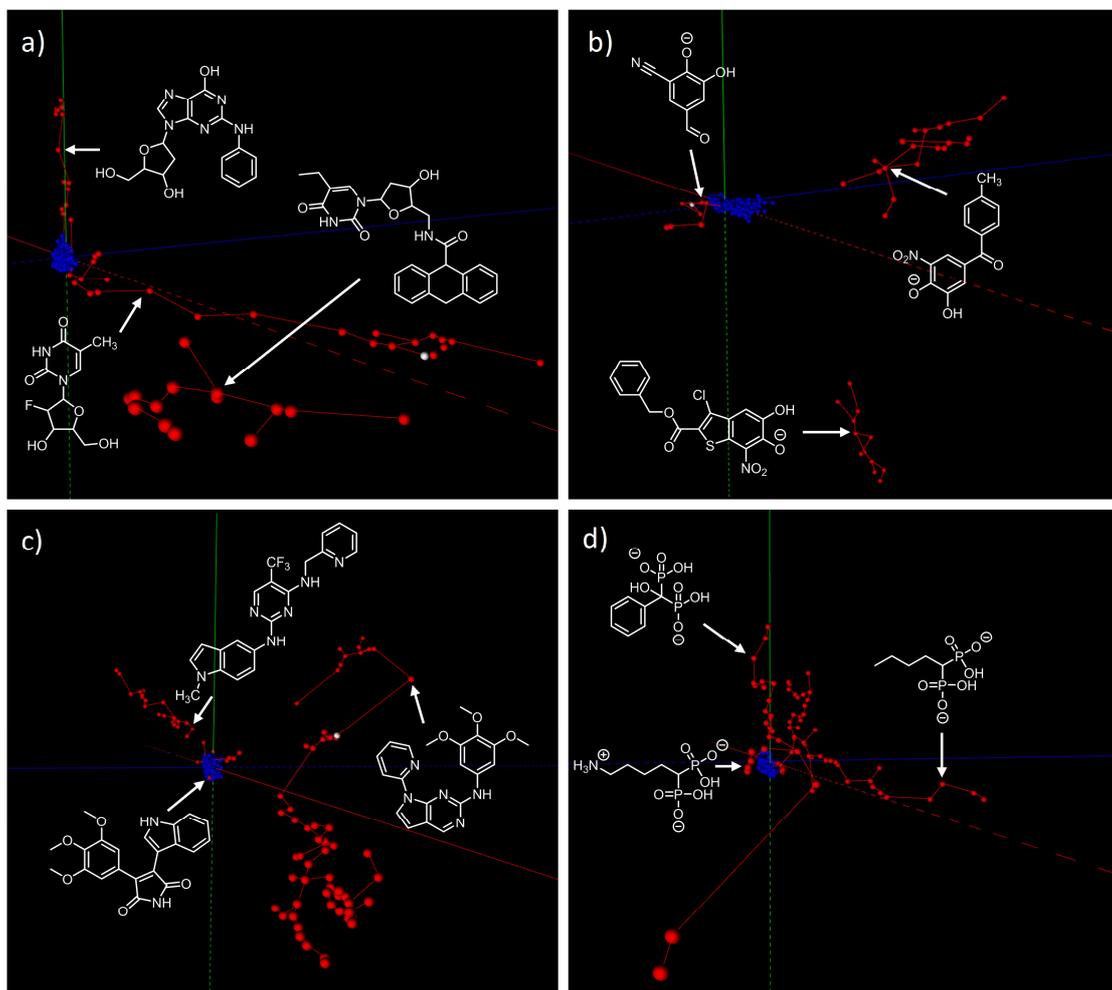


Figure 5. Visualization of benchmarking sets from DUD-E using MHFP6 similarity maps in WebMolCS, color coded as actives (red) or decoys (blue). Interactive 3D-maps can be accessed at the indicated weblinks. (a) Thymidine kinase (kith), PC1 = 70 % , PC2 = 9 % , PC3 = 5 % , <http://gdbtools.unibe.ch:8080/webMolCS/yourSIM.html?jobID=1544775928296&fp=MHFP>. (b) catechol O-methyl transferase (comt), PC1 = 72 % , PC2 = 11 % , PC3 = 3 % , <http://gdbtools.unibe.ch:8080/webMolCS/yourSIM.html?jobID=1544776368253&fp=MHFP>. (c) focal adhesion kinase 1 (fak1), PC1 = 70 % , PC2 = 8 % , PC3 = 5 % , <http://gdbtools.unibe.ch:8080/webMolCS/yourSIM.html?jobID=1544776863167&fp=MHFP>. (d) farnesyl diphosphate synthase (fpps), PC1 = 43 % , PC2 = 12 % , PC3 = 11 % , <http://gdbtools.unibe.ch:8080/webMolCS/yourSIM.html?jobID=1544777667633&fp=MHFP>.

MHFP6 chemical space 3D-maps can also be constructed using actives from one, two, three of four different targets as references. In this case we added approximately the 2,200 approved drugs listed in DrugBank for comparison, as illustrated here for the four target specific series discussed above (Figure 6).³¹ Each series of actives appears as an elongated group separated from the approved

drugs, which cluster in the center of the map because none of them is considered as reference in this display. The 3D-maps highlight the main structural patterns in each target specific series. For focal adhesion kinase (fak1), this motif is a pyrrolopyridine, for thymidine kinase (kith), it is a thymidine-like substructure, for farnesyl diphosphate synthase (fpps) a diphosphonate group, and for catechol-O-methyl transferase (comt) a catechol. These maps show that our MHFP6 chemical space 3D-maps also provide useful differentiation between compound series with different bioactivities.

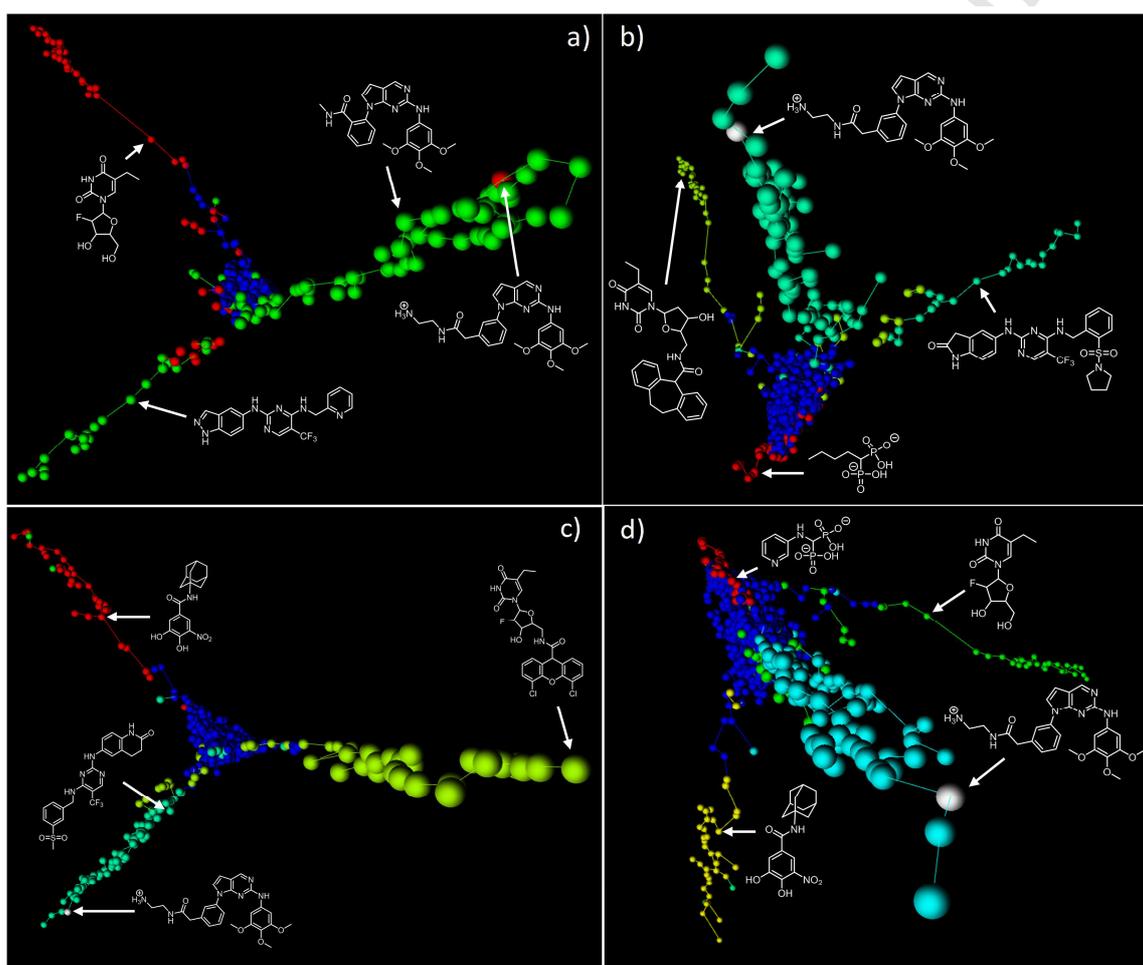


Figure 6. Visualization of active sets from DUD-E in combination with approved drugs from DrugBank using MHFP6 similarity maps in WebMolCS, color coded by set (actual value option in controls). (a) green: Fak1 (focal adhesion kinase 1), red: kith (thymidine kinase), blue: DrugBank, <http://gdbtools.unibe.ch:8080/webMolCS/yourSIM.html?jobID=1539954046166&fp=MHFP>. (b) green: kith (thymidine kinase), red: fpps (farnesyl diphosphate synthase), cyan: Fak1 (focal adhesion kinase 1), blue: DrugBank, <http://gdbtools.unibe.ch:8080/webMolCS/yourSIM.html?jobID=1539951847766&fp=MHFP>. (c) green: kith (thymidine kinase), red: comt (catechol O-methyl transferase), cyan: Fak1 (focal adhesion kinase 1), blue: DrugBank, <http://gdbtools.unibe.ch:8080/webMolCS/yourSIM.html?jobID=1539951512218&fp=MHFP>. (d) green: kith (thymidine kinase), red: fpps (farnesyl diphosphate synthase), cyan: Fak1 (focal adhesion kinase 1), yellow: comt (catechol O-methyl transferase), blue: DrugBank, <http://gdbtools.unibe.ch:8080/webMolCS/yourSIM.html?jobID=1539952304206&fp=MHFP>.

Conclusion

In summary, we modified our web-based application WebMolCS to enable 3D-maps displaying similarity relationships between compounds calculated with MHFP6, a recently reported molecular fingerprint, which outperforms ECFP4 in virtual screening benchmarking studies. We then used our modified WebMolCS to compute a MHFP6 chemical space 3D-map displaying a set of compounds used to discover CBA, a potent inhibitor of the cation channel TRPM4. The map helped us to design a new series of analogs, which we synthesized and tested for TRPM4 inhibition, leading to the discovery of compounds **8** and **9** as two new CBA analogs with comparable potencies. We also found that MHFP6 chemical space 3D-maps provide insightful representations of various series of target specific actives and their relationship to decoy molecules and to drugs. We propose MHFP6 chemical space 3D-maps as an alternative to R-group variation tables for designing, interpreting and communicating the results of SAR studies.

Methods

Cheminformatics

Loading datasets. The WebMolCS webpage at <http://gdb.unibe.ch> takes as input dataset an annotated list of molecules in which each line of the file represents a different molecule and comprises the following entries: 1) the SMILES of the molecule; 2) the name or code for the molecule; 3) a value for color coding, which may be a number representing activity as pIC_{50} or a number corresponding to a selective subset; 4) a satellite selection value, set to 1 if the compound should be considered as reference for the similarity map calculation, and 0 if it should not be considered as reference.

MHFP6 fingerprint calculation. Besides the series of fingerprints already present, the modified WebMolCS web application now includes an option for computing similarity maps in the MHFP6

chemical space, which uses the 128D version of MHFP6. Calculating the MHFP6 fingerprint of a molecule comprises the following operations: 1) write the SMILES of all circular substructures up to a diameter of six bonds around each atom; 2) convert each SMILES to a value using a Hash function; 3) Apply the MinHash procedure to this set of values. For the 128D MHFP6 considered here, MinHash applies 128 different Hash functions to each value in the primary set, yielding 128 new values for each substructure. For each of the 128 different hash functions, only the smallest of the hashed values is kept. These minimum hashed values are collected to form a secondary set of values, which is the MHFP6 fingerprint. The method is probabilistic in that each substructure only has a certain probability of being considered in the final set of secondary values generated by the MinHash procedure. The resulting 128D MHFP6 is a set of 128 values. The similarity between two molecules is calculated from their MHFP6 fingerprints as the ratio of the number of common values to the total number of different values in the two sets (Jaccard index).

Similarity map calculation. To produce the MHFP6 similarity map, WebMolCS selects as references either 50 molecules taken at regular interval across the entire dataset, or the subset of molecules with the satellite option value set to 1. The server then performs the following operations: 1) compute the MHFP6 fingerprint similarity value of each compound in the list to each of the reference compounds; 2) calculate the PCA on the resulting similarity fingerprint; 3) construct a 3D-representation of the dataset using the first three principal components (PCs) as coordinates, in which each molecule appears as a sphere positioned at its (PC1, PC2, PC3) coordinates, and the molecular structure is shown on mouse-over. The color-code for the spheres can be selected in the menu, and the spheres can be connected with each other by lines representing the minimum spanning tree, which connects nearest neighbors in the 3D-space. The entire computation from the input dataset to the MHFP6 similarity map requires approximately 2.2 minutes of server time for a set of 1,000 molecules.

Availability. The WebMolCS webpage at <http://gdb.unibe.ch> is freely accessible for use. The source code for the application and instructions to setup WebMolCS in a dockerized container are available on Github at <https://github.com/reymond-group/webMolCS>.

Synthesis

Reagents and solvents were purchased from commercial sources and were used without further purification. MilliQ deionised water and HPLC-grade acetonitrile were employed for RP-HPLC solvents. Synthesis were followed both by Alugram Xtra Sil G/UV₂₅₄ with detection of the compounds by UV at 214 nm or by UPLC-MS. RP-UPLC spectra were obtained using a Dionex Ultimate 3000 RSLC System and a Dionex Acclaim RSLC 120 column. Three methods were performed as follows: A gradient from 0.05 % TFA to 9:1 acetonitrile/water 0.05 % TFA during 2.2 min then 2.8 min of 9:1 acetonitrile/water 0.05 % TFA isocratic (method A). A gradient from 0.05 % TFA to 9:1 acetonitrile/water 0.05 % TFA during 4.5 min then 2.5 min of 9:1 acetonitrile/water 0.05 % TFA isocratic (method B). A gradient from 0.05 % TFA to 9:1 acetonitrile/water 0.05 % TFA during 7.5 min then 2.5 min of 9:1 acetonitrile/water 0.05 % TFA isocratic (method C). Low-resolution mass spectra were conducted on a Thermo Scientific LCQ Fleet by electron spray ionization (ESI) in the positive mode. Automated column chromatography purification was executed on an Interchim Puriflash 430 system on RediSepRf prepacked columns. Preparative RP-HPLC was performed on a Waters Prep LC4000 Chromatography system using a Reprospher 100 column from Dr. Maisch GmbH with a flow of 40 mL/min and two lines of solvents: a 0.1 % TFA and a 9:1 acetonitrile/water 0.1 % TFA. Compounds were analyzed by UPLC-MS and their purity assigned by analytical RP-UPLC. NMR spectra were recorded on: BRUKER AC 300 for ¹H-300 MHz; ¹³C-75 MHz; ¹⁹F-282 MHz measurements or BRUKER AC 400 for ¹H-400 MHz; ¹³C-101 MHz measurements. Chemical shifts are given in parts per million (δ) referenced to TMS ($\delta = 0.00$ ppm ¹H-, ¹³C-NMR). Coupling constants are given in Hertz. High resolution mass spectra measurements were performed by the “Service of Mass Spectrometry” of the Department of

Chemistry and Biochemistry at the University of Bern on a Thermo Scientific LTQ OrbitrapXL

(group of PD Dr. Stefan Schürch).

***tert*-Butyl 2-(2-chlorophenoxy)acetate (1).** Potassium carbonate (46.7 mmol; 6.45 g) was added to a solution of 2-chlorophenol (23.3 mmol; 2.38 mL) in DMF (10 mL) and stirred for 10 min at RT. The *tert*-butyl chloroacetate (25.7 mmol; 3.67 mL) was added then the mixture was heated to 80°C for 2 h until the reaction was complete. After full conversion, the hot mixture was poured into a mixture of crushed ice and water (~50 mL). The mixture was stirred for 30 min then the precipitate was filtered off and dried to afford a white solid (5.50 g; 97%), which was used without further treatment. ¹H NMR (300 MHz, CDCl₃) δ 7.38 (dd, *J* = 7.9, 1.6 Hz, 1H), 7.19 (ddd, *J* = 8.2, 7.6, 1.6 Hz, 1H), 6.93 (td, *J* = 7.7, 1.4 Hz, 1H), 6.82 (dd, *J* = 8.2, 1.3 Hz, 1H), 4.60 (s, 2H), 1.48 (s, 9H).

2-(2-Chlorophenoxy)acetic acid (2). The ester **1** (22.7 mmol; 5.50 g) was dissolved in trifluoroacetic acid (22.7 mL) and stirred at RT for 1 hour and a half until the reaction was complete. The crude was concentrated under reduced pressure. The obtained solid was taken back with DCM several times then evaporated to dryness to afford a white solid (4.20 g; 99%), which was used without further treatment. ¹H NMR (300 MHz, DMSO) δ 13.10 (s, 1H), 7.43 (dd, *J* = 7.9, 1.6 Hz, 1H), 7.27 (ddd, *J* = 8.5, 7.5, 1.6 Hz, 1H), 7.06 - 6.92 (m, 2H), 4.80 (s, 2H).

Methyl 2-(2-(2-chlorophenoxy)acetamido)-4-cyanobenzoate (3). Acid **2** (2.20 mmol; 0.42 g) was dissolved in 12 mL toluene and SOCl₂ (82.3 mmol; 6.00 mL) was added. The mixture was refluxed overnight, and the solvent was removed on high vacuum to give the crude acyl chloride. This product was dissolved in a small amount of THF and added to a solution of methyl 2-amino-4-cyanobenzoate (2.00 mmol; 0.35 g) and K₂CO₃ (4.00 mmol; 0.55 g) in THF (29 ml) stirred for 15 min at RT. The reaction mixture was stirred at RT for 4 h until the reaction was complete. The reaction was diluted with water (30 mL) and extracted with EtOAc (3 x 30 mL). The combined organics were washed with brine, dried with Na₂SO₄ and evaporated to dryness. The crude product was recrystallized in acetonitrile to give white crystals (0.55 g; 80%). ¹H NMR (400 MHz, CDCl₃)

δ 11.90 (s, 1H), 9.16 (d, $J = 1.4$ Hz, 1H), 8.13 (d, $J = 8.2$ Hz, 1H), 7.42 (ddd, $J = 12.3, 8.0, 1.5$ Hz, 2H), 7.25 (td, $J = 8.0, 1.6$ Hz, 1H), 7.05 – 6.96 (m, 2H), 4.72 (s, 2H), 3.95 (s, 3H). ^{13}C NMR (101 MHz, CDCl_3) δ 167.45, 166.71, 153.10, 140.55, 131.72, 130.96, 128.10, 126.38, 124.51, 123.86, 123.53, 119.61, 117.77 (d, $J = 5.3$ Hz), 114.82, 69.11, 53.13. HRMS m/z $[\text{M} + \text{H}]^+$ calcd for $\text{C}_{17}\text{H}_{14}\text{ClN}_2\text{O}_4$ 345.0637, found: 345.0627. mp: 159.6-160.1°C.

2-(2-(2-chlorophenoxy)acetamido)-4-cyanobenzoic acid (4). A 1.4 M aqueous solution of lithium hydroxide (9.24 mmol; 6.60 mL) was added to a solution of **3** (1.54 mmol; 0.53 g) in 13 mL dioxane and stirred at RT for 1 h until the reaction was complete. The reaction mixture was quenched with 3.1 mL of a 1M HCl solution and concentrated under reduced pressure. The crude was first filtrated on a silica pad with ethyl acetate acidified with acetic acid then purified by automated flash chromatography on 25 g of silica gel (gradient from 98 % 2 % DCM/MeOH to 95 % 10% DCM/MeOH) two times. The impure fractions obtained from the column were taken back in water and the formed precipitate was filtered and then triturated in ethanol three times to give a pure white solid (0.004 g; 1 %). ^1H NMR (400 MHz, DMSO) δ 14.32 (s, 1H), 11.88 (s, 1H), 8.98 (d, $J = 1.2$ Hz, 1H), 8.14 (d, $J = 8.2$ Hz, 1H), 7.66 (dd, $J = 8.2, 1.4$ Hz, 1H), 7.48 (dd, $J = 7.9, 1.4$ Hz, 1H), 7.35 – 7.28 (m, 1H), 7.23 – 7.17 (m, 1H), 7.07 – 7.00 (m, 1H), 4.91 (s, 2H). ^{13}C NMR (101 MHz, DMSO) δ 167.94, 167.49, 152.65, 139.96, 132.25, 130.19, 128.35, 126.52, 123.20, 122.76, 121.74, 120.78, 117.89, 115.77, 114.47, 68.12. HRMS m/z $[\text{M} + \text{H}]^+$ calcd for $\text{C}_{16}\text{H}_{12}\text{ClN}_2\text{O}_4$ 331.0480, found: 331.0473. mp = 30°C. UPLC (method A): $t_R = 2.18$ min., purity: 99 %.

Methyl 4,5-dichloro-2-(2-(2-chlorophenoxy)acetamido)benzoate (5). Acid **2** (1.10 mmol; 0.20 g) was dissolved in 12 mL toluene and SOCl_2 (82.3 mmol; 6.00 mL) was added. The mixture was refluxed overnight, and the solvent was removed on high vacuum to give the crude acyl chloride. This product was dissolved in a small amount of THF and added to a solution of methyl 2-amino-4,5-dichloro-benzoate (1.00 mmol; 0.22 g) and K_2CO_3 (2.01 mmol; 0.28 g) in THF (13 mL) stirred for 15 min at RT. The reaction mixture was stirred at RT for 4 h until the reaction was complete. The reaction was diluted with water (30 mL) and extracted with EtOAc (3 x 30 mL). The combined

organics were washed with brine, dried with Na₂SO₄ and evaporated to dryness to give a yellow solid which was used without further purification (0.38 g; 97%). ¹H NMR (300 MHz, CDCl₃) δ 11.83 (s, 1H), 9.01 (s, 1H), 8.11 (s, 1H), 7.44 (dd, *J* = 7.8, 1.5 Hz, 1H), 7.23 (dd, *J* = 7.9, 1.5 Hz, 1H), 7.05 – 6.95 (m, 2H), 4.70 (s, 2H), 3.92 (s, 3H). ¹³C NMR (75 MHz, CDCl₃) δ 167.06, 153.04, 139.04, 132.08, 130.83, 127.95, 126.92, 123.35, 122.57, 117.54, 115.84, 114.70, 69.04, 52.80. One quaternary carbon was not observed and is expected to be close to the baseline. HRMS *m/z* [M + Na]⁺ calcd for C₁₆H₁₂Cl₃NO₄Na 409.9724, found: 409.9716. mp: 158.2-160.4°C.

4,5-Dichloro-2-(2-(2-chlorophenoxy)acetamido)benzoic acid (6). A 1.4 M aqueous solution of lithium hydroxide (5.82 mmol; 4.16 mL) was added to a solution of **5** (0.97 mmol; 0.38 g) in 12 mL dioxane and stirred at RT for 1 h until the reaction was complete. The reaction mixture was quenched with 1.9 mL of a 1M HCl solution and concentrated under reduced pressure. The crude was taken back in 0.1 % TFA and the formed precipitate was filtered and washed with water twice. The obtained precipitate was first taken back in HCl 1M two times and filtrated and then triturated by MeOH to give a white solid (0,15 mg; 41%). ¹H NMR (400 MHz, DMSO) δ 14.25 (s, 1H), 11.85 (s, 1H), 8.91 (s, 1H), 8.07 (s, 1H), 7.55 – 7.40 (m, 1H), 7.30 (s, 1H), 7.23 – 7.12 (m, 1H), 7.02 (s, 1H), 4.87 (s, 2H). ¹³C NMR (101 MHz, DMSO) δ 167.32, 152.62, 139.26, 136.30, 132.22, 130.17, 128.29, 124.95, 122.74, 121.79, 121.31, 117.12, 114.46, 68.14. HRMS *m/z* [M + Na]⁺ calcd for C₁₅H₁₀Cl₃NO₄Na 395.9568, found: 395.9568. HRMS *m/z* [M - H]⁺ calcd for C₁₅H₉Cl₃NO₄ 371.9603, found: 371.9613. mp: 267.6-268.6°C. UPLC (method A): *t_R* = 2.75 min., purity: 98 %.

Methyl 2-(2-(2-chlorophenoxy)acetamido)-6-methylbenzoate (7). Acid **2** (1.10 mmol; 0.20 g) was dissolved in 12 mL toluene and SOCl₂ (82.3 mmol; 6.00 mL) was added. The mixture was refluxed overnight, and the solvent was removed on high vacuum to give the crude acyl chloride. This product was dissolved in a small amount of THF and added to a solution of methyl 2-amino-6-methyl-benzoate (1.00 mmol; 0.17 g) and K₂CO₃ (2.00 mmol; 0.28 g) in THF (17 ml) stirred for 15 min at RT. The reaction mixture was stirred at RT for 3 h until the reaction was complete. The reaction was diluted with water (30 mL) and extracted with EtOAc (3 x 30 mL). The combined

organics were washed with brine, dried with Na₂SO₄ and evaporated to dryness to give a brown oil which was used without further purification (0.39 g; 99%). ¹H NMR (300 MHz, CDCl₃) δ 10.07 (s, 1H), 8.17 (d, *J* = 8.3 Hz, 1H), 7.44 (dd, *J* = 7.9, 1.6 Hz, 1H), 7.37 (t, *J* = 8.0 Hz, 1H), 7.24 (dd, *J* = 8.1, 1.6 Hz, 1H), 7.06 – 7.00 (m, 2H), 7.00 – 6.92 (m, 1H), 4.68 (s, 2H), 3.85 (s, 3H), 2.43 (s, 3H). ¹³C NMR (75 MHz, CDCl₃) δ 168.47, 166.39, 153.12, 138.68, 136.37, 131.50, 130.82, 128.15, 127.49, 123.59, 123.27, 122.25, 120.25, 114.45, 68.93, 52.32, 21.88. HRMS *m/z* [M + H]⁺ calcd for C₁₇H₁₇ClNO₄ 334.0841, found: 334.0845.

2-(2-(2-chlorophenoxy)acetamido)-6-methylbenzoic acid (8). A 1.4 M aqueous solution of lithium hydroxide (6.93 mmol; 4.95 mL) was added to a solution of **7** (1.16 mmol; 0.39 g) in 14 mL dioxane and stirred at RT for 1 h until the reaction was complete. The reaction mixture was quenched with 2.3 mL of a 1M HCl solution and concentrated under reduced pressure. The crude was taken back in 0.1 % TFA and the formed precipitate was filtered and washed with MeOH twice. The obtained precipitate was first triturated by *tert*-butylmethylether several times then DCM several times to give a white solid (0.06 mg; 17%). ¹H NMR (400 MHz, DMSO) δ 12.77 (s, 1H), 8.07 (d, *J* = 8.1 Hz, 1H), 7.45 (dd, *J* = 7.9, 1.6 Hz, 1H), 7.28 (ddd, *J* = 8.5, 7.5, 1.6 Hz, 1H), 7.13 (dd, *J* = 8.3, 1.3 Hz, 1H), 7.05 – 6.96 (m, 2H), 6.79 (d, *J* = 7.5 Hz, 1H), 4.70 (s, 2H), 2.39 (s, 3H). ¹³C NMR (101 MHz, DMSO) δ 169.21, 165.27, 153.48, 136.61, 136.38, 130.87, 130.04, 128.14, 126.29, 125.36, 122.35, 122.01, 116.93, 114.85, 68.77, 21.90. HRMS *m/z* [M + H]⁺ calcd for C₁₆H₁₅ClNO₄ 320.0684, found: 320.0684. mp: 254.0-258.8°C. UPLC (method A): *t*_R = 2.28 min., purity: 96 %.

4-Chloro-2-(2-(2-chlorophenoxy)acetamido)-*N*-(cyclopropylsulfonyl)benzamide (9). A solution of **CBA** (0.29 mmol; 0.10 g), Oxyma (0.44 mmol; 0.06 g), DIC (0.44 mmol; 0.07 mL) and collidine (0.44 mmol; 0.06 mL) was added to a solution of the cyclopropanesulfonamide (0.88 mmol; 0.11 mg) in DCM (4 mL). The reaction mixture was stirred at RT for 2h30 until the reaction was complete. The reaction mixture was concentrated under reduced pressure and the obtained crude was further purified by RP-HPLC (gradient from 0% to 80 % 9:1 acetonitrile/water 0.1 % TFA in

40 minutes) to give a pure batch of the compound and two other impure fractions that were recrystallized in ethyl acetate to give a white lyophilisate (0.02 g; 17%). ¹H NMR (400 MHz, DMSO) δ 11.94 (s, 1H), 8.76 (d, *J* = 2.1 Hz, 1H), 8.01 (d, *J* = 8.6 Hz, 1H), 7.48 (dd, *J* = 7.9, 1.6 Hz, 1H), 7.31 – 7.26 (m, 3H), 7.17 (ddd, *J* = 14.0, 8.3, 1.2 Hz, 2H), 7.03 (td, *J* = 7.7, 1.4 Hz, 2H), 4.89 (s, 2H), 2.91 (tt, *J* = 8.1, 4.8 Hz, 1H), 1.11 – 0.88 (m, 3H). ¹³C NMR (101 MHz, DMSO) δ 168.42, 167.36, 152.69, 140.99, 138.50, 132.93, 130.20, 128.35, 123.21, 122.74, 119.39, 115.27, 114.43, 114.28, 68.19, 30.39, 5.27. HRMS *m/z* [M + Na]⁺ calcd for C₁₈H₁₆Cl₂N₂O₅SNa 465.0049, found: 465.0040. UPLC (method A): *t_R* = 2.34 min., purity: 97 %.

Methyl 4-bromo-2-(2-(2-chlorophenoxy)acetamido)benzoate (10). Acid **2** (21.4 mmol; 4.00 g) was dissolved in 16 mL toluene and SOCl₂ (107 mmol; 7.82 mL) was added. The mixture was refluxed overnight, and the solvent was removed on high vacuum to give the crude acyl chloride. This product was dissolved in a small amount of THF and added to a solution of methyl 2-amino-4-bromobenzoate (19.4 mmol; 4.48 g) and K₂CO₃ (38.9 mmol; 5.38 g) in THF (115 ml) stirred for 15 min at RT. The reaction mixture was stirred at RT overnight until the reaction was complete. The reaction was diluted with water (30 mL) and extracted with EtOAc (3 x 30 mL). The combined organics were washed with brine, dried with Na₂SO₄ and evaporated to dryness to give the pure product as a brown solid (6.40 g; 83 %). ¹H NMR (400 MHz, CDCl₃) δ 11.88 (s, 1H), 9.02 (d, *J* = 1.9 Hz, 1H), 7.89 (d, *J* = 8.5 Hz, 1H), 7.44 (dd, *J* = 7.7, 1.5 Hz, 1H), 7.29 (dd, *J* = 8.5, 2.0 Hz, 1H), 7.23 (dd, *J* = 7.8, 1.5 Hz, 1H), 7.06 – 6.96 (m, 2H), 4.71 (s, 2H), 3.90 (s, 3H). ¹³C NMR (101 MHz, CDCl₃) δ 167.51, 167.21, 153.28, 141.04, 132.13, 130.94, 129.40, 128.06, 126.76, 123.96, 123.91, 123.41, 115.12, 114.84, 69.26, 52.67. HRMS *m/z* [M + Na]⁺ calcd for C₁₆H₁₃BrClNO₄Na 419.9609, found: 419.9601. mp: 127.1-130.2°C.

Methyl 2-(2-(2-chlorophenoxy)acetamido)-4-(3,5-dimethylisoxazol-4-yl)benzoate & Ethyl 2-(2-(2-chlorophenoxy)acetamido)-4-(3,5-dimethylisoxazol-4-yl)benzoate (11). **10** (0.38 mmol; 0.15 g), (3,5-dimethylisoxazol-4-yl)boronic acid (0.56 mmol; 0.08 g), K₂CO₃ (1.13 mmol; 0.16 g) and palladium tetrakis (0.02 mmol; 0.02 g) were dissolved in a mixture of dioxane/water/ethanol (2 mL;

4:1:10) degassed under argon. The reaction mixture was stirred at 80°C for 2h until the reaction was complete. The reaction mixture was concentrated under reduced pressure and taken back in water and diethyl ether. The organic phase was washed with brine, dried with Na₂SO₄ and concentrated under reduced pressure. The obtained crude was further purified by automated flash chromatography on 12 g of silica gel (gradient from 80 % 20 % Hexane/EtOAc to 100 % EtOAc) to give a yellow oil corresponding to a 50/50 mixture of both esters (0.10 g; 63 %). ¹H NMR (400 MHz, CDCl₃) δ 11.97 (s, 1H), 11.89 (s, 1H), 8.76 (t, *J* = 1.4 Hz, 2H), 8.11 (dd, *J* = 8.2, 5.3 Hz, 2H), 7.44 (ddd, *J* = 8.0, 2.3, 1.7 Hz, 2H), 7.23 (dd, *J* = 5.6, 3.9 Hz, 1H), 7.06 (dt, *J* = 8.2, 1.8 Hz, 2H), 7.01 (tt, *J* = 8.0, 2.2 Hz, 4H), 4.72 (s, 4H), 4.39 (q, *J* = 7.1 Hz, 2H), 3.93 (s, 3H), 2.50 (d, *J* = 0.9 Hz, 6H), 2.36 (d, *J* = 0.7 Hz, 6H), 1.39 (t, *J* = 7.1 Hz, 3H). ¹³C NMR (101 MHz, CDCl₃) δ 167.31, 167.24, 166.39, 153.35, 140.56 (d, *J* = 8.7 Hz), 136.84 (d, *J* = 14.7 Hz), 131.46 (d, *J* = 6.6 Hz), 130.94 (d, *J* = 1.7 Hz), 128.06 (d, *J* = 2.4 Hz), 123.94, 123.46 (dd, *J* = 17.7, 5.3 Hz), 121.37 (d, *J* = 7.3 Hz), 115.59, 115.35, 114.85 (d, *J* = 2.3 Hz), 69.29, 61.71, 52.62, 14.37, 12.05, 11.19.

2-(2-(2-chlorophenoxy)acetamido)-4-(3,5-dimethylisoxazol-4-yl)benzoic acid (12). A 1.4 M aqueous solution of lithium hydroxide (1.40 mmol; 1.00 mL) was added to a solution of **11** (0.23 mmol; 0.10 g) in 2 mL dioxane and stirred at RT for 1 h until the reaction was complete. The reaction mixture was quenched with 0.5 mL of a 1M HCl solution and concentrated under reduced pressure. The crude was first filtrated on a silica pad with ethyl acetate acidified with acetic acid then triturated in dry methanol and *tert*-butylmethylether to give a yellow solid (0.03 g; 33 %). ¹H NMR (400 MHz, DMSO) δ 8.61 (s, 1H), 8.06 (d, *J* = 8.1 Hz, 1H), 7.47 (dd, *J* = 7.9, 1.4 Hz, 1H), 7.34 – 7.24 (m, 1H), 7.15 (dd, *J* = 12.2, 4.5 Hz, 2H), 7.05 – 6.97 (m, 1H), 4.83 (s, 2H), 2.45 (s, 3H), 2.27 (s, 3H). ¹³C NMR (101 MHz, DMSO) δ 168.63, 166.69, 165.61, 157.94, 153.15, 140.22, 131.61, 130.14, 128.26, 122.67, 122.47, 121.84, 119.43, 115.49, 114.49, 68.48, 11.48, 10.62. HRMS *m/z* [M + H]⁺ calcd for C₂₀H₁₈ClN₂O₅ 401.0899, found: 401.0899. mp: 196.6-204.2°C. UPLC (method A): *t*_R = 2.48 min., purity: > 99 %.

Methyl 2-(2-(2-chlorophenoxy)acetamido)-4-(thiophen-3-yl)benzoate (13). **10** (0.44 mmol; 0.18 g), 3-thienylboronic acid (0.66 mmol; 0.08 g), K_2CO_3 (1.32 mmol; 0.18 g) and palladium tetrakis (0.02 mmol; 0.02 g) were dissolved in a mixture of dioxane/water/methanol (2 mL; 4:1:10) degased under argon. The reaction mixture was stirred at 80°C for 2h until the reaction was complete. The reaction mixture was concentrated under reduced pressure and taken back in water and diethyl ether. The formed precipitate was filtered to give the pure product as a brown solid (0.13 g; 72 %). 1H NMR (400 MHz, $CDCl_3$) δ 11.90 (s, 1H), 9.08 (d, $J = 1.7$ Hz, 1H), 8.06 (d, $J = 8.3$ Hz, 1H), 7.66 – 7.63 (m, 1H), 7.48 (dd, $J = 5.0, 1.4$ Hz, 1H), 7.44 (dd, $J = 7.8, 1.5$ Hz, 1H), 7.41 (dd, $J = 5.1, 3.0$ Hz, 1H), 7.38 (dd, $J = 8.3, 1.8$ Hz, 1H), 7.23 (dd, $J = 7.9, 2.1$ Hz, 1H), 7.06 – 6.95 (m, 2H), 4.73 (s, 2H), 3.91 (s, 3H). ^{13}C NMR (101 MHz, $CDCl_3$) δ 167.85, 167.21, 153.46, 141.56, 141.10, 140.81, 131.67, 130.92, 128.05, 126.74, 126.51, 123.98, 123.32, 122.83, 121.27, 118.68, 114.92, 114.87, 69.43, 52.45. HRMS m/z $[M + Na]^+$ calcd for $C_{20}H_{16}ClNO_4SNa$ 424.0381, found: 424.0384. mp: 137.1-138.8°C.

2-(2-(2-chlorophenoxy)acetamido)-4-(thiophen-3-yl)benzoic acid (14). A 1.4 M aqueous solution of lithium hydroxide (1.87 mmol; 1.33 mL) was added to a solution of **13** (0.31 mmol; 0.13 g) in 2.5 mL dioxane and stirred at RT for 2 h until the reaction was complete. The reaction mixture was quenched with 0.6 mL of a 1M HCl solution and concentrated under reduced pressure. The crude was first filtrated on a silica pad with ethyl acetate acidified with acetic acid then triturated in ethanol to give a yellow solid (0.03 g; 23 %). 1H NMR (400 MHz, DMSO) δ 13.63 (s, 1H), 11.91 (s, 1H), 9.01 (d, $J = 1.7$ Hz, 1H), 8.02 (d, $J = 8.3$ Hz, 1H), 7.98 (dd, $J = 2.9, 1.3$ Hz, 1H), 7.72 (dd, $J = 5.0, 2.9$ Hz, 1H), 7.58 – 7.51 (m, 2H), 7.48 (dd, $J = 7.9, 1.6$ Hz, 1H), 7.35 – 7.28 (m, 1H), 7.20 (dd, $J = 8.3, 1.2$ Hz, 1H), 7.03 (td, $J = 7.7, 1.3$ Hz, 1H), 4.89 (s, 2H). ^{13}C NMR (101 MHz, DMSO) δ 168.92, 167.04, 152.82, 140.61, 140.29, 140.24, 131.89, 130.20, 128.34, 127.86, 125.99, 123.39, 122.67, 121.78, 120.92, 117.23, 114.99, 114.41, 68.28. HRMS m/z $[M + H]^+$ calcd for $C_{19}H_{15}ClNO_4S$ 388.0405, found: 388.0404. mp: 240.5-268.1°C. UPLC (method A): $t_R = 2.64$ min., purity: 99 %.

2-(2-(2-chlorophenoxy)acetamido)-4-(1-methyl-1H-pyrazol-4-yl)benzoic acid & 2-(2-(2-chlorophenoxy)acetamido)-4-(1-methyl-1H-pyrazol-5-yl)benzoic acid (15). 10 (0.63 mmol; 0.25 g), 1-methyl-4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)pyrazole (0.94 mmol; 0.20 g), K_2CO_3 (1.88 mmol; 0.26 g) and palladium tetrakis (0.03 mmol; 0.04 g) were dissolved in a mixture of dioxane/water/methanol (4 mL; 4:1:10) degassed under argon. The reaction mixture was stirred at 80°C overnight until the reaction was complete. The reaction mixture was concentrated under reduced pressure and taken back in water and diethyl ether. The formed precipitate corresponding to the ester was filtered and the aqueous phase is freeze-dried. The obtained lyophilisate was filtrated on a silica pad with ethyl acetate acidified with acetic acid then triturated in ethyl acetate and *tert*-butylmethylether to give a yellow solid (0.02 g; 10 %; products ratio 87:13). 1H NMR (400 MHz, DMSO) δ 15.33 (s, 1H), 8.66 (d, $J = 1.7$ Hz, 1H), 8.04 (s, 1H), 7.92 (d, $J = 8.0$ Hz, 1H), 7.73 (d, $J = 0.5$ Hz, 1H), 7.44 (dd, $J = 7.9, 1.6$ Hz, 1H), 7.31 – 7.24 (m, 1H), 7.13 (ddd, $J = 9.5, 8.2, 1.5$ Hz, 2H), 6.98 (td, $J = 7.7, 1.4$ Hz, 1H), 4.77 (s, 2H), 3.85 (s, 3H). ^{13}C NMR (101 MHz, DMSO) δ 168.78, 165.75, 153.66, 140.65, 135.85, 133.56, 131.57, 130.01, 128.10, 127.82, 123.79, 122.08, 122.04, 121.84, 118.34, 114.72, 114.51, 68.70, 38.55. HRMS m/z $[M + H]^+$ calcd for $C_{19}H_{17}ClN_3O_4$ 386.0902, found: 386.0900. HRMS m/z $[M + Na]^+$ calcd for $C_{19}H_{16}ClN_3O_4Na$ 408.0722, found: 408.0717. UPLC (method A): $t_R = 2.10$ & 2.48 min., purity: 99 %.

Methyl 2-(2-(2-chlorophenoxy)acetamido)-4-(oxazol-2-yl)benzoate (16). 10 (0.50 mmol; 0.20 g), tributyl(oxazol-2-yl)stannane (1.05 mmol; 0.22 mL) and palladium tetrakis (0.40 mmol; 0.46 g) were dissolved in THF (50 mL) degassed under argon. The reaction mixture was stirred at 65°C overnight until the reaction was complete. The reaction mixture was concentrated under reduced pressure and taken back in EtOAc/DCM/Hexane. The formed precipitate was filtered, and the filtrate was concentrated under reduced pressure. The obtained crude was further purified by automated flash chromatography on 15 g of silica gel (isocratic eluent of 2:1 Hexane/EtOAc) to give a brown solid (0.07 g; 34 %). 1H NMR (400 MHz, $CDCl_3$) δ 11.88 (s, 1H), 9.43 (d, $J = 1.5$ Hz, 1H), 8.14 (d, $J = 8.3$ Hz, 1H), 7.86 (dd, $J = 8.3, 1.6$ Hz, 1H), 7.78 (s, 1H), 7.44 (dd, $J = 8.3, 1.5$ Hz,

1H), 7.30 (s, 1H), 7.24 (dd, $J = 8.1, 1.3$ Hz, 1H), 7.02 (d, $J = 7.8$ Hz, 2H), 4.75 (s, 2H), 3.93 (s, 3H). ^{13}C NMR (101 MHz, CDCl_3) δ 167.50, 167.18, 160.89, 153.35, 140.64, 139.68, 132.55, 131.68, 130.93, 129.18, 128.07, 123.90, 123.34, 121.19, 118.70, 117.55, 114.80, 69.30, 52.71. HRMS m/z $[\text{M} + \text{H}]^+$ calcd for $\text{C}_{19}\text{H}_{16}\text{ClN}_2\text{O}_5$ 387.0742, found: 387.0745. mp: 145-146°C.

2-(2-(2-chlorophenoxy)acetamido)-4(oxazol-2-yl)benzoic acid (17). A 1.4 M aqueous solution of lithium hydroxide (1.00 mmol; 0.72 mL) was added to a solution of **16** (0.17 mmol; 0.06 g) in 2 mL dioxane and stirred at RT for 1 h until the reaction was complete. The reaction mixture was quenched with 0.3 mL of a 1M HCl solution and concentrated under reduced pressure. The crude was first filtrated on a silica pad with ethyl acetate acidified with acetic acid then triturated in ethanol to give a first pure batch of the compound. The trituration filtrate was concentrated under reduced pressure and recrystallized in $\text{H}_2\text{O}/\text{ACN}$ to give grey crystals that are gathered with the first batch (0.02 g; 24 %). ^1H NMR (400 MHz, DMSO) δ 12.05 (s, 1H), 9.35 (d, $J = 1.5$ Hz, 1H), 8.32 (s, 1H), 8.14 (d, $J = 8.3$ Hz, 1H), 7.78 (dd, $J = 8.3, 1.6$ Hz, 1H), 7.49 (d, $J = 1.5$ Hz, 1H), 7.47 (s, 1H), 7.35 – 7.28 (m, 1H), 7.20 (dd, $J = 8.3, 1.1$ Hz, 1H), 7.03 (td, $J = 7.8, 1.3$ Hz, 1H), 4.90 (s, 2H). ^{13}C NMR (101 MHz, DMSO) δ 168.63, 167.14, 159.78, 152.80, 141.03, 140.41, 132.17, 131.20, 130.18, 129.04, 128.32, 122.69, 121.80, 120.30, 118.26, 117.04, 114.46, 68.31. HRMS m/z $[\text{M} + \text{H}]^+$ calcd for $\text{C}_{18}\text{H}_{14}\text{ClN}_2\text{O}_5$ 373.0586, found: 373.0581. mp: 253.3-253.4°C. UPLC (method A): $t_R = 2.17$ min., purity: 96 %.

Methyl 2-(2-(2-chlorophenoxy)acetamido)-4-(thiazol-2-yl)benzoate (18). **10** (0.50 mmol; 0.20 g), tributyl(thiazol-2-yl)stannane (1.05 mmol; 0.33 mL) and palladium tetrakis (0.40 mmol; 0.46 g) were dissolved in THF (50 mL) degased under argon. The reaction mixture was stirred at 65°C overnight until the reaction was complete. The reaction mixture was concentrated under reduced pressure and the obtained crude was further purified by automated flash chromatography on 50 g of silica gel (isocratic eluent of 2:1 Hexane/EtOAc) to give a yellow solid (0.17 g; 86 %). ^1H NMR (400 MHz, CDCl_3) δ 11.91 (s, 1H), 9.39 (d, $J = 1.7$ Hz, 1H), 8.13 (d, $J = 8.3$ Hz, 1H), 7.94 (d, $J = 3.2$ Hz, 1H), 7.84 (dd, $J = 8.4, 1.8$ Hz, 1H), 7.45 (dd, $J = 6.3, 2.1$ Hz, 2H), 7.25 – 7.22 (m, 1H), 7.01

(dd, $J = 12.0, 4.5$ Hz, 2H), 4.75 (s, 2H), 3.93 (s, 3H). ^{13}C NMR (101 MHz, CDCl_3) δ 167.56, 167.28, 166.97, 153.39, 144.31, 140.81, 138.67, 131.85, 130.94, 128.07, 123.97, 123.37, 121.09, 120.54, 119.16, 117.21, 114.90, 69.36, 52.66. HRMS m/z $[\text{M} + \text{H}]^+$ calcd for $\text{C}_{19}\text{H}_{16}\text{ClN}_2\text{O}_4\text{S}$ 403.0514, found: 403.0506. HRMS m/z $[\text{M} + \text{Na}]^+$ calcd for $\text{C}_{19}\text{H}_{15}\text{ClN}_2\text{O}_4\text{SNa}$ 425.0333, found: 425.0324. mp: 71.4-71.5°C.

2-(2-(2-chlorophenoxy)acetamido)-4-(thiazol-2-yl)benzoic acid (19). A 1.4 M aqueous solution of lithium hydroxide (2.51 mmol; 1.79 mL) was added to a solution of **18** (0.42 mmol; 0.17 g) in 4 mL dioxane and stirred at RT for 1 h until the reaction was complete. The reaction mixture was quenched with 0.8 mL of a 1M HCl solution and concentrated under reduced pressure. The crude was first filtrated on a silica pad with ethyl acetate acidified with acetic acid then recrystallized in acetonitrile to give impure crystals that were triturated in *tert*-butylmethylether to give a yellow solid (0.02 g; 14 %). ^1H NMR (400 MHz, DMSO) δ 13.17 (s, 1H), 9.29 (d, $J = 1.4$ Hz, 1H), 8.09 (d, $J = 8.2$ Hz, 1H), 7.99 (d, $J = 3.2$ Hz, 1H), 7.87 (d, $J = 3.2$ Hz, 1H), 7.69 (dd, $J = 8.2, 1.6$ Hz, 1H), 7.47 (dd, $J = 7.9, 1.4$ Hz, 1H), 7.33 – 7.26 (m, 1H), 7.18 (dd, $J = 8.2, 0.9$ Hz, 1H), 7.02 (td, $J = 7.8, 1.2$ Hz, 1H), 4.87 (s, 2H). ^{13}C NMR (101 MHz, DMSO) δ 168.48, 166.86, 166.26, 153.09, 144.21, 140.58, 140.58, 136.16, 136.16, 132.14, 130.15, 128.28, 122.51, 121.83, 121.46, 120.18, 116.86, 114.49, 68.45. HRMS m/z $[\text{M} + \text{H}]^+$ calcd for $\text{C}_{18}\text{H}_{14}\text{ClN}_2\text{O}_4\text{S}$ 389.0357, found: 389.0364. mp: 228.1-228.3°C. UPLC (method A): $t_{\text{R}} = 2.30$ min., purity: 95 %.

Acknowledgment. This work was supported financially by the Swiss National Science Foundation, NCCR TransCure.

Competing interest statement. The authors declare not competing interests.

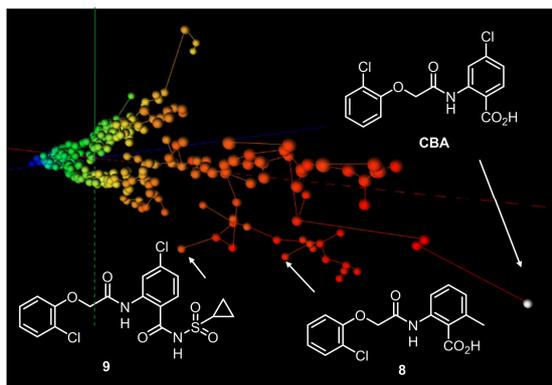
References

1. Ozthail, L. C.; Delalande, C.; Bianchi, B.; Nemeth, G.; Kappel, S.; Thomet, U.; Ross-Kaschitza, D.; Simonin, C.; Rubin, M.; Gertsch, J.; Lochner, M.; Peinelt, C.; Reymond, J. L.; Abriel, H. Identification of Potent and Selective Small Molecule Inhibitors of the Cation Channel Trpm4. *Br. J. Pharmacol.* **2018**, *175*, 2504-2519.
2. Wang, C.; Naruse, K.; Takahashi, K. Role of the Trpm4 Channel in Cardiovascular Physiology and Pathophysiology. *Cells* **2018**, *7*.
3. Hantute-Ghesquier, A.; Haustrate, A.; Prevarskaya, N.; Lehen'kyi, V. Trpm Family Channels in Cancer. *Pharmaceuticals (Basel)* **2018**, *11*, 58.
4. Guinamard, R.; Hof, T.; Del Negro, C. A. The Trpm4 Channel Inhibitor 9-Phenanthrol. *Br. J. Pharmacol.* **2014**, *171*, 1600-1613.
5. Simonin, C.; Awale, M.; Brand, M.; van Deursen, R.; Schwartz, J.; Fine, M.; Kovacs, G.; Häfliger, P.; Gyimesi, G.; Sithampari, A.; Charles, R.-P.; Hediger, M. A.; Reymond, J.-L. Optimization of Trpv6 Calcium Channel Inhibitors Using a 3d Ligand-Based Virtual Screening Method. *Angew. Chem. Int. Ed.* **2015**, *54*, 14748-14752.
6. Kilchmann, F.; Marcaida, M. J.; Kotak, S.; Schick, T.; Boss, S. D.; Awale, M.; Gönczy, P.; Reymond, J.-L. Discovery of a Selective Aurora a Kinase Inhibitor by Virtual Screening. *J. Med. Chem.* **2016**, *59*, 7188-7211.
7. Awale, M.; Probst, D.; Reymond, J. L. Webmolcs: A Web-Based Interface for Visualizing Molecules in Three-Dimensional Chemical Spaces. *J Chem Inf Model* **2017**, *57*, 643-649.
8. Morgan, H. L. The Generation of a Unique Machine Description for Chemical Structures-a Technique Developed at Chemical Abstracts Service. *J. Chem. Doc.* **1965**, *5*, 107-113.
9. Penny, R. H. A Connectivity Code for Use in Describing Chemical Structures. *J. Chem. Doc.* **1965**, *5*, 113-117.
10. Probst, D.; Reymond, J. L. A Probabilistic Molecular Fingerprint for Big Data Settings. *ChemRxiv* **2018**, doi.org/10.26434/chemrxiv.7176350.v1.
11. Hagadone, T. R. Molecular Substructure Similarity Searching: Efficient Retrieval in Two-Dimensional Structure Databases. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 515-521.

12. Geppert, H.; Vogt, M.; Bajorath, J. Current Trends in Ligand-Based Virtual Screening: Molecular Representations, Data Mining Methods, New Application Areas, and Performance Evaluation. *J. Chem. Inf. Model.* **2010**, *50*, 205-216.
13. Owen, J. R.; Nabney, I. T.; Medina-Franco, J. L.; López-Vallejo, F. Visualization of Molecular Fingerprints. *J. Chem. Inf. Model.* **2011**, *51*, 1552-1563.
14. Scior, T.; Bender, A.; Tresadern, G.; Medina-Franco, J. L.; Martinez-Mayorga, K.; Langer, T.; Cuanalo-Contreras, K.; Agrafiotis, D. K. Recognizing Pitfalls in Virtual Screening: A Critical Review. *J. Chem. Inf. Model.* **2012**, *52*, 867-881.
15. Riniker, S.; Landrum, G. A. Open-Source Platform to Benchmark Fingerprints for Ligand-Based Virtual Screening. *J. Cheminform.* **2013**, *5*, 26.
16. Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742-754.
17. Mysinger, M. M.; Carchia, M.; Irwin, J. J.; Shoichet, B. K. Directory of Useful Decoys, Enhanced (Dud-E): Better Ligands and Decoys for Better Benchmarking. *J. Med. Chem.* **2012**, *55*, 6582-6594.
18. Nguyen, K. T.; Blum, L. C.; van Deursen, R.; Reymond, J.-L. Classification of Organic Molecules by Molecular Quantum Numbers. *ChemMedChem* **2009**, *4*, 1803-1805.
19. van Deursen, R.; Blum, L. C.; Reymond, J. L. A Searchable Map of Pubchem. *J. Chem. Inf. Model.* **2010**, *50*, 1924-1934.
20. Schwartz, J.; Awale, M.; Reymond, J.-L. Smifp (Smiles Fingerprint) Chemical Space for Virtual Screening and Visualization of Large Databases of Organic Molecules. *J. Chem. Inf. Model.* **2013**, *53*, 1979-1989.
21. Awale, M.; Reymond, J.-L. Atom Pair 2d-Fingerprints Perceive 3d-Molecular Shape and Pharmacophores for Very Fast Virtual Screening of Zinc and Gdb-17. *J. Chem. Inf. Model.* **2014**, *54*, 1892-1907.
22. Probst, D.; Reymond, J. L. Fun: A Framework for Interactive Visualizations of Large, High-Dimensional Datasets on the Web. *Bioinformatics* **2018**, *34*, 1433-1435.

23. Raghavendra, A. S.; Maggiora, G. M. Molecular Basis Sets - a General Similarity-Based Approach for Representing Chemical Spaces. *J. Chem. Inf. Model.* **2007**, *47*, 1328-1240.
24. Medina-Franco, J. L.; Maggiora, G. M.; Giulianotti, M. A.; Pinilla, C.; Houghten, R. A. A Similarity-Based Data-Fusion Approach to the Visual Characterization and Comparison of Compound Databases. *Chem. Biol. Drug. Des.* **2007**, *70*, 393-412.
25. Awale, M.; Reymond, J. L. Similarity Mapplet: Interactive Visualization of the Directory of Useful Decoys and ChEMBL in High Dimensional Chemical Spaces. *J. Chem. Inf. Model.* **2015**, *55*, 1509-1516.
26. Wassermann, A. M.; Wawer, M.; Bajorath, J. Activity Landscape Representations for Structure-Activity Relationship Analysis. *J. Med. Chem.* **2010**, *53*, 8209-8923.
27. Gaspar, H. A.; Baskin, I. I.; Marcou, G.; Horvath, D.; Varnek, A. Chemical Data Visualization and Analysis with Incremental Generative Topographic Mapping: Big Data Challenge. *J. Chem. Inf. Model.* **2014**, *55*, 84-94.
28. Sander, T.; Freyss, J.; von Korff, M.; Rufener, C. Datawarrior: An Open-Source Program for Chemistry Aware Data Visualization and Analysis. *J. Chem. Inf. Model.* **2015**, *55*, 460-473.
29. Ballatore, C.; Huryn, D. M.; Smith Iii, A. B. Carboxylic Acid (Bio)Isosteres in Drug Design. *ChemMedChem* **2013**, *8*, 385-395.
30. Scola, P. M.; Sun, L. Q.; Wang, A. X.; Chen, J.; Sin, N.; Venables, B. L.; Sit, S. Y.; Chen, Y.; Cocuzza, A.; Bilder, D. M.; D'Andrea, S. V.; Zheng, B.; Hewawasam, P.; Tu, Y.; Friberg, J.; Falk, P.; Hernandez, D.; Levine, S.; Chen, C.; Yu, F.; Sheaffer, A. K.; Zhai, G.; Barry, D.; Knipe, J. O.; Han, Y. H.; Schartman, R.; Donoso, M.; Mosure, K.; Sinz, M. W.; Zvyaga, T.; Good, A. C.; Rajamani, R.; Kish, K.; Tredup, J.; Klei, H. E.; Gao, Q.; Mueller, L.; Colonna, R. J.; Grasela, D. M.; Adams, S. P.; Loy, J.; Levesque, P. C.; Sun, H.; Shi, H.; Sun, L.; Warner, W.; Li, D.; Zhu, J.; Meanwell, N. A.; McPhee, F. The Discovery of Asunaprevir (Bms-650032), an Orally Efficacious Ns3 Protease Inhibitor for the Treatment of Hepatitis C Virus Infection. *J. Med. Chem.* **2014**, *57*, 1730-1752.
31. Law, V.; Knox, C.; Djoumbou, Y.; Jewison, T.; Guo, A. C.; Liu, Y.; Maciejewski, A.; Arndt, D.; Wilson, M.; Neveu, V.; Tang, A.; Gabriel, G.; Ly, C.; Adamjee, S.; Dame, Z. T.; Han, B.; Zhou, Y.; Wishart, D. S. Drugbank 4.0: Shedding New Light on Drug Metabolism. *Nucleic Acids Res.* **2014**, *42*, D1091-D1097.

Graphics for the Table of Contents:



In this paper we propose interactive MHFP6 chemical space 3D-maps as an alternative to R-group variation tables for designing, interpreting and communicating the results of SAR studies, and exemplify the method at the example of a structure-activity relationship study of TRPM4 inhibitors.

ACCEPTED MANUSCRIPT