

From Cancer To Pain Target By Automated Selectivity Inversion Of A Clinical Candidate

Samo Turk, Benjamin Merget, Sameh Eid, and Simone Fulle

J. Med. Chem., **Just Accepted Manuscript** • DOI: 10.1021/acs.jmedchem.8b00140 • Publication Date (Web): 10 May 2018

Downloaded from <http://pubs.acs.org> on May 13, 2018

Just Accepted

“Just Accepted” manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides “Just Accepted” as a service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. “Just Accepted” manuscripts appear in full in PDF format accompanied by an HTML abstract. “Just Accepted” manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are citable by the Digital Object Identifier (DOI®). “Just Accepted” is an optional service offered to authors. Therefore, the “Just Accepted” Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the “Just Accepted” Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these “Just Accepted” manuscripts.

1
2
3
4
5
6 From Cancer To Pain Target By Automated
7
8
9
10 Selectivity Inversion Of A Clinical Candidate
11
12
13

14 *Samo Turk*[‡], *Benjamin Merget*[‡], *Sameh Eid*[‡], *Simone Fulle*^{*}
15
16

17 *BioMed X Innovation Center, Im Neuenheimer Feld 515, 69120 Heidelberg, Germany*
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

ABSTRACT

Elimination of inadvertent binding is crucial for inhibitor design targeting conserved protein classes like kinases. Compounds in clinical trials provide a rich source for initiating drug design efforts by exploiting such secondary binding events. Considering both aspects, we shifted the selectivity of tozasertib, originally developed against AurA as cancer target, towards the pain target TrkA. First, selectivity-determining features in binding pockets were identified by fusing interaction-grids of several key and off-target conformations. A focused library was subsequently created and prioritized using a multi-objective selection scheme that filters for selective and highly active compounds based on orthogonal methods grounded in computational chemistry and machine learning. 18 high-ranking compounds were synthesized and experimentally tested. The top-ranked compound has 10,000-fold improved selectivity versus AurA, nanomolar cellular activity and is highly selective in a kinase panel. This was achieved in a single round of automated *in silico* optimization, highlighting the power of recent advances in computer-aided drug design to automate design and selection processes.

Keywords: automated compound optimization, machine learning, kinases, selectivity, drug design, small molecules

INTRODUCION

The focus of R&D efforts on kinases for more than two decades has resulted in many kinase inhibitors with known bioactivity and safety profiles. These compounds provide a great source for initiating new compound design efforts, as for instance many compounds inhibit more than one molecular target and can have a therapeutic impact via so far unknown or ignored mechanisms.^{1,2} The latter is particularly true for inhibitors that bind to the highly conserved ATP-binding pocket of kinases and thus, often bind unintentionally also to secondary targets. Inhibiting multiple targets can result in beneficial synergistic therapeutic effects, but also could lead to unwanted side effects.³

Hence, specifically switching off the activity against undesired targets, while maintaining or even improving the affinity to the key target, is of great importance, especially when treating non-life-threatening diseases such as inflammation or pain. Here, we report the automated design of improved inhibitors of the tropomyosin receptor kinase A (TrkA) by jointly employing a *de novo* design platform and a multi-objective selection scheme that considers selectivity and activity aspects as predicted by novel *in silico* tools. The selection strategy aimed to shift the selectivity of a particular kinase inhibitor from its original key target towards another validated kinase and takes in addition the selectivity profile against the kinome into account. The initial target–compound pair (i.e., TrkA and tozasertib) was chosen by mining the available kinase profiling data^{4–6} for kinases inhibitors that had entered clinical trials but also inhibit non-cancer targets. The corresponding protein binding sites were subsequently analyzed with respect to selectivity-determining features to further prioritize target-compound pairs (see Experimental Section for details). The original key target of tozasertib is Aurora kinase A (AurA), whose inhibition has common adverse effects such as neutropenia and hematological toxicities.⁷ In turn, the profiling data revealed that tozasertib also inhibits TrkA, which is a validated drug target for cancer and pain.⁸ Thus, inverting the

1
2
3 selectivity of tozasertib from AurA towards TrkA will likely result in a new compound series
4
5 with reduced side effects and, hence is of potential interest for pain treatment.
6
7
8
9

10 **RESULTS AND DISCUSSION**

11 *Identification of selectivity-determining features in TrkA*

12
13
14
15
16 The selectivity hot-spots in the TrkA binding site were initially identified by fusing atom-
17
18 based interaction energy grids of several TrkA, Aurora A and B structures into one content-
19
20 rich representation of target specific sub-pockets (i.e., energy grids representing the
21
22 conformational flexibility of the key and off-targets were calculated, respectively, and finally
23
24 fused to one representation via difference rules; see Experimental Section and Supplementary
25
26 Information for details including validation results). The resulting ‘selectivity grids’ highlight
27
28 three areas of interest for compound optimization: Two are favorable hydrophobic sub-
29
30 pockets for TrkA-selectivity adjacent to the gatekeeper residue and enclosed between the Asp
31
32 residue of the DFG-motif and a Phe residue of the glycine-rich loop (G-loop), respectively,
33
34 and one is an unfavorable pocket for TrkA-selectivity that overlaps with the cyclopropyl
35
36 moiety of tozasertib (**Figure 1**).
37
38
39
40

41 *Virtual compound library design*

42
43
44 The identified selectivity-determining areas subsequently guided the compound library design
45
46 where two parts of tozasertib were enumerated using commercially available drug-like
47
48 fragments and retrosynthetic rules (**Figure 2**). The resulting compound set **A** contains 7404
49
50 possible modifications of the amino-5-methylpyrazole that forms hydrogen bonds with the
51
52 hinge and extends towards the gatekeeper residue, while set **B** contains 6326 possible
53
54 modifications of the N-(4-aminothiophene)cyclopropylcarboxamide. It should be noted that in
55
56 set **A** the cyclopropylcarboxamide moiety was removed to avoid potential intramolecular
57
58
59
60

1
2
3 steric clashes, and that in set **B** modifications extend towards the DFG-motif but can
4 potentially also form interactions with the gatekeeper residue. The resulting library was
5 prioritized using a multi-objective compound selection scheme that filters for selective and
6 highly active compounds (**Figure 3**). All employed prediction tools showed good prediction
7 power in an initial validation on project-specific data (see Supplementary Information for
8 details).

17 *Selectivity and activity optimization*

19 Initially, binding poses of compounds in sets **A** and **B** were generated with Glide SP.⁹
20 Compounds were removed for reasons of either poor docking scores (> -7.5 kcal/mol), wrong
21 orientation, or lack of key interactions (Experimental Section). In the next step (**Figure 3**),
22 compounds with an unfavorable selectivity profile were filtered out. This was accomplished
23 via machine learning-based activity prediction models¹⁰ that were used I) to remove
24 promiscuous compounds (i.e., predicted to be active at IC_{50} of 500 nM on ≥ 20 kinases) and
25 II) those that are predicted to be highly active ($IC_{50} < 10$ nM) on Aurora A, B, or C kinases.
26 The selectivity filtering was complemented by a structure-based procedure employing the
27 TrkA–Aurora ‘selectivity grids’ for rescoring of docking solutions. The remaining
28 compounds in both sets (**A**: 592; **B**: 1145) were finally prioritized for highly active
29 compounds using two complementary machine learning (ML) technologies. All compounds
30 were evaluated by an ‘MMP/ML’ approach,¹¹ which is trained on fragment-based Matched
31 Molecular Pairs (MMPs), and quantifies compound activity differences. In addition,
32 compounds in the **A** set interacting with the Phe gatekeeper were additionally evaluated by a
33 hybrid QM/ML pipeline. This pipeline is trained on high-level quantum mechanical (QM)
34 calculations to quantify ligand–gatekeeper interactions and rescores the top hits in a second
35 step with fragment molecular orbital calculations, taking the entire binding pocket into
36 consideration.^{12,13}

1
2
3 The final selection of compounds was obtained in an automated fashion based on the scoring
4 ranks (i.e., without any manual selection beside synthesizability criteria). 18 of the top
5 prioritized compounds were synthesized and tested for inhibition on TrkA and AurA
6 (Supplementary **Table S1-S2**), nine each from sets **A (a1-a9)** and **B (b1-b9)**. The synthesized
7 compounds were from the top ~50 compounds from sets **A** and **B**, respectively. This yield
8 was achieved by considering synthetic pathway and building block availability in the library
9 design.
10
11
12
13
14
15
16
17
18

19 *Experimental testing on TrkA and AurA*

20
21
22 An initial screening via binding assays revealed that 4 out of 18 tested compounds (**Table 1**)
23 have an improved activity/selectivity profile compared to the starting compound tozasertib
24 (i.e. **a1**, **a4**, **b7** and **b8**; **Figure 3**) and originate from both sets **A** and **B**. Compound **a1** shows
25 the best overall improvement compared to tozasertib with slightly higher TrkA affinity and
26 almost complete loss of the AurA activity (**Table 2**). Compound **a1** was highly scored with
27 both orthogonal ML methods used for the final ranking (ranked 1 with MMP/ML and 3 with
28 QM/ML). Predicted binding modes (**Figure 1**) indicate that, in contrast to tozasertib, **a1** forms
29 an additional π - π -interaction with the gatekeeper residue Phe in TrkA via the fluorophenyl
30 moiety and does not occupy the adjacent unfavorable area (occupied by the
31 cyclopropylcarboxamide group of tozasertib). This trend is consistent with **reference 2** (i.e.,
32 tozasertib without cyclopropylcarboxamide), which shows improved Trka-AurA selectivity
33 ($\Delta\Delta pK_d = 0.9$) and thus underscores the importance of avoiding this unfavorable area (**Table**
34 **2**). Compound **a4** also has a slightly improved TrkA affinity but retains AurA binding. In
35 contrast to the cyclopentyl of **a4**, the polarized fluorophenyl of **a1** seems to make less
36 favorable interactions with the Leu gatekeeper in AurA, possibly explaining the higher
37 selectivity of compound **a1** compared to **a4**. Based on the primary screening results it is
38 apparent that in the **A** set all triazole-containing compounds are inactive on both TrkA and
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 AurA (Supplementary **Table S1**), whereas modifications containing a pyrazole hinge binder
4 in general show activity against TrkA as well as TrkA–AurA selectivity at 100 nM screening
5 concentration. Interestingly, even a small extension toward the gatekeeper, such as the *tert*-
6 butyl moiety in compound **a8** (**Table S1**), leads to a complete loss of affinity to AurA. This
7 underlines that bulky moieties are less tolerated by the AurA binding pocket and confirms the
8 vicinity of the gatekeeper as selectivity-determining region.
9
10
11
12
13
14
15
16

17 The two compounds **b7** and **b8** are highly similar and both introduce an additional methylene-
18 linker between a sulfur and a *para*-substituted phenyl ring (*para*-fluoro in **b7** and *para*-methyl
19 in **b8**). The phenyl rings in both compounds are predicted to stack with Phe521 from the G-
20 loop, potentially benefiting from a more favorable orientation provided by the additional
21 methylene-linker group compared to tozasertib. The higher activity of **b7** against TrkA
22 indicates that the electron-withdrawing fluoro-substitution of **b7** increases the strength of the
23 stacking interaction, compared to the electron-donating methyl-substitution of **b8**. The
24 primary screening results of the **B** set overall indicate a positive effect of a methylene-linker
25 between the ligand scaffold and the terminal aryl as well as a *para*-substitution on the phenyl
26 ring for TrkA–AurA selectivity (**Table S2**).
27
28
29
30
31
32
33
34
35
36
37
38
39

40 The best compound from each set was further characterized in a cell-based functional assay
41 revealing low nanomolar cellular potency for both compounds (**a1**: IC₅₀ = 26 nM; **b7**: IC₅₀ =
42 23 nM). This indicates that the compounds are able to reach the intracellular kinase domain
43 and thus permeate through the cellular membrane.
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Kinase profiling

To evaluate the overall selectivity profile of all four hit compounds, profiling on a panel of 92 wild-type kinases and 5 additional variants was conducted at 100 nM screening concentration. The kinase selectivity of compound **a1** is very high, with only six kinases (excluding TrkA) inhibited, leading to a selectivity score of $S(35\%) = 0.08$ and $S(10\%) = 0.04$ (**Figure 4; Table S3**). The remaining three hit compounds show slightly lower kinome selectivity ($S(35\%) \approx 0.2$; **Figure S1**). At a concentration of 1 μM , **a1** still inhibits only 19 kinases, which is in line with the initial filtering step via the *in silico* kinase profiling platform that has excluded promiscuous compounds (**Figure S2**). It is also noteworthy that all four hit compounds are inactive on AurB, which was part of the off-target ensemble in the selectivity optimization scoring via the ‘selectivity grids’. Compound **a1** displays the most unique binding profile and is even more selective than the starting compound tozasertib.

The kinases inhibited by compound **a1** include, like for tozasertib, FLT3 and RET (both common cancer targets), but also the two other Trk isoforms TrkB and TrkC, rendering **a1** a pan-Trk inhibitor with excellent selectivity against other kinases. Compared to **a1**, **b8** is less active on TrkB and TrkC and inactive on PDGFRb, pointing to the potential value of fusing both modifications in the future. Finally, hematological toxicities such as thrombocytopenia, anemia and neutropenia are typically associated with ABL and JAK2 inhibitors.^{14,15} Thus, it is encouraging that compared to tozasertib, compound **a1** is also inactive against all tested ABL1 variants (i.e. two mutants and the (un)phosphorylated wild type) as well as Janus kinases (**Figure 4**). Overall, the high kinome selectivity of **a1**, together with the fact that **a1** is, compared to tozasertib, inhibiting fewer kinases with target-associated toxicity (i.e. AurA and ABL1) makes **a1** an interesting candidate for further studies. Many Trk inhibitor scaffolds have been reported in publications and patent applications and are summarized in excellent reviews on Trka inhibitors.^{8,16–19} It is noteworthy that the aminopyrazolyl-

1
2
3 pyrimidine scaffold in the presented compound series has been already reported previously in
4 the pan-Trk inhibitor AZ-23.²⁰ However, AZ-23 has a significantly different substitution
5 pattern on the pyrimidine ring (i.e. amino linker at position 2 and unsubstituted position 6).
6
7
8
9

10 *Role of TrkA for pain treatment and current status of TrkA inhibitor design*

11
12
13 TrkA is a high affinity receptor for the Nerve Growth Factor (NGF) and widely expressed on
14 peripheral pain sensing neurons. Activation of TrkA can lead to pathogenesis of many
15 difficult to treat human pain conditions such as osteoarthritis and cancer-related types of pain.
16
17 Loss-of-function TrkA variants are associated with congenital insensitivity to pain.²¹
18
19 Furthermore, *in vivo* TrkA inhibition studies indicate pain suppression effects^{22,23} underlining
20 the promise of TrkA as a pain target. Other selective pan-Trk inhibitors already exist²⁴⁻²⁷ and
21 show promise for the treatment of acute and chronic pain. Studies with an allosteric TrkA
22 inhibitor indicate that TrkA specific inhibitors can be expected to have an improved safety
23 profile since pan-Trk inhibitors have a hyperphagic effect.⁸ Reaching isoform selectivity vs
24 TrkB and TrkC in the ATP-binding site is difficult as only two residues differ among the three
25 isoforms (i.e. two Arg in TrkA are replaced by Lys in TrkB and TrkC)⁸ but might be tackled
26 in future improvements with modifications of the so far unexplored N-methylpiperazine
27 moiety that form e.g. cation- π interactions with one of the arginines.²⁸
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

CONCLUSION

Starting from a compound that reached clinical trials for cancer indications, we have designed a new series of TrkA inhibitors by employing a multi-objective selection scheme that filters for selective and highly active compounds. In agreement with the design objectives, the top compound **a1** is highly active on TrkA, overall highly selective in a kinase panel, and – most importantly from a design perspective – has a 10,000-fold improved selectivity against the selected off-target AurA. Nanomolar cellular potency against TrkA further underlines the potential value of **a1** as an advanced hit compound for the treatment of acute and chronic pain or other conditions resulting from abnormal TrkA activity such as inflammation and cancer.

Promiscuous interactions between drugs and proteins can cause adverse effects.²⁹ Thus, being able to switch off undesired targets in a single round of automated *in silico* optimization underlines the power of recent advances in computer-aided drug design technologies, especially those employing machine learning techniques such as the virtual kinase profiling platform¹⁰ or the MMP/ML approach.¹¹ Noteworthy, compound **a1** was the highest-ranking compound in the filtering pipeline (Supplementary **Table S1**). The initially employed ‘selectivity grids’ pointed to selectivity-determining features in the TrkA structure and thereby guided the subsequent compound library design. Furthermore, they were used to filter for TrkA–Aurora selective compounds. Other successful *in silico*-driven compound design efforts considering pharmacological profiles include those that identified new targets^{30,31} or guided compound optimization.²⁹ The uniqueness of our study is that we navigated through the highly conserved and large target class of kinases instead of seeking predictions across more diverse protein classes. Overall, the discovery platform described herein is widely applicable for compound optimization for target classes with common selectivity issues but can be also generally employed for switching off undesired activities in order to reduce adverse effects.³²

EXPERIMENTAL SECTION

Computational Methods

Details about the employed computational methods are described below and related validation results in the Supplementary Information. Kinome tree figures were created using KinMap.³³

Selection of initial target-compound pair. Mining kinase inhibitors in clinical trials (obtained from the data sources ChEMBL and clinicaltrials.gov) pointed to TrkA as drug target for pain treatment, while mining bioactivity data of a curated ChEMBL v22⁴ kinase inhibitor panel and other kinase profiling data sets^{3,5,6} pointed to the Aurora inhibitor tozasertib as promising starting compound.

Identification of selectivity determining features in TrkA. PDB structures of TrkA, AurA, and AurB kinases in the DFG-in state were used for the generation of the selectivity grids (**Table 3**; i.e. 4 PDB structures of TrkA were used as the key target set, while 67 AurA and 2 AurB PDB structures constituted the off-target set). All kinase structures were aligned to 4pmt using PyMOL, considering only binding site residues (defined as all residues within 5 Å of any bound ligand). Atom-based energy grids were calculated for the aligned structures using AutoGrid (version 4.2.5),³⁴ a grid spacing of 0.375 Å, and covering all ligand atoms in the PDB structures. Subsequently, energy grids were fused into single key and off-target grids by taking the minimum (median) interaction energy value at each point across the respective grid ensemble for scoring (for visualization). Extreme energy values from steric clashes were capped at a maximum value corresponding to the absolute value of the most favorable interaction energy. Finally, the key and off-target fused grids were combined using fusion rules.³⁵ An aromatic carbon grid representation (A probe) was used to identify selective hydrophobic sub-pockets, while the acceptor oxygen (OA), acceptor nitrogen (NA), and donor hydrogen (HD) probes were used to identify polar sub-pockets.

1
2
3 **Virtual compound library design.** A virtual compound library was designed by modification
4 of tozasertib via systematic replacements of 3-methylpyrazole (compound set **A**) and N-(4-
5 aminothiophene)cyclopropylcarboxamide (compound set **B**), using the MolPort building
6 block database as source of fragments. For set **A**, all primary amines connected to an aromatic
7 N with one further connection were extracted and further narrowed down to fragments with \leq
8 15 heavy atoms and exactly one amino group. For set **B**, aromatic thiole containing fragments
9 were combined with acyl chloride containing fragments. Here, aromatic thiols were initially
10 filtered to keep only those with \leq 15 heavy atoms and one amino group while acyl chlorides
11 were filtered to keep only those with \leq 10 heavy atoms and no aromatic rings. The resulting
12 fragments in set **A** were combined with the remaining part of tozasertib without the
13 cyclopropylcarboxamide moiety to avoid potential intramolecular steric clashes. All virtual
14 compounds were finally filtered using the following cutoffs: Molecular weight \leq 600 Da, logP
15 \leq 5, number of hydrogen bond donors \leq 5 and acceptors \leq 12.
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30

31
32 **Prediction of binding modes and structure-based screening via ‘selectivity grids’.** The
33 four pre-aligned TrkA PDB structures (**Table 3**) were processed using the Protein Preparation
34 Wizard in the Schrödinger Suite (release 2016-4, Schrödinger, LLC, New York, NY, 2016)
35 by adjusting protonation, optimizing hydrogens, and restricted geometry optimization using
36 the OPLS3 force field. Docking grids were generated for the four protein models using the
37 Glide program,⁹ including a hydrogen bond constraint to the backbone NH of the Met592
38 hinge residue. The ensemble docking run employed the standard precision protocol (Glide SP)
39 with enhanced conformational sampling and increasing the number of initial poses to 50000.
40 For each virtual ligand, only the best pose as judged by the Glide SP score, was kept for
41 subsequent analysis along with the corresponding protein configuration. To filter the virtual
42 library further, compounds were excluded if: (1) the top pose had a docking score $>$ -7.5
43 kcal/mol, (2) it had a wrong binding orientation, e.g. the piperazine moiety is buried rather
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 than being solvent-exposed as in tozasertib, or (3) it lacked key tozasertib-TrkA interactions;
4
5 namely, Phe589 contact, H-bond to Met592, contact to Asp668 in the DFG motif, and a
6
7 favorable contact to either Arg654 or Arg673. Finally, the top poses of the remaining
8
9 compounds were re-scored by interpolating the energy values from the corresponding TrkA-
10
11 Aurora selectivity grids to estimate their selectivity propensity.
12
13

14
15 **Filtering of selective compound via in silico kinase profiling platform.** The employed
16
17 platform contains machine learning-based activity prediction models for small molecules
18
19 which were trained on chemical fingerprints and a large and diverse data set of kinase
20
21 inhibitor data.¹⁰ The technology was used in a two-step filtering procedure: (1) to remove
22
23 promiscuous compounds and (2) to filter with respect to selected off-target activity.¹⁰ In the
24
25 first step, compounds were removed that were predicted to be active on ≥ 20 kinases with a
26
27 probability of ≥ 0.7 in high-quality models (132 models with AUC values ≥ 0.8 ; activity
28
29 threshold: $IC_{50} = 500$ nM). In the second step, prediction models were trained for Aurora
30
31 kinases A, B, and C employing an activity threshold of $IC_{50} = 10$ nM to remove compounds
32
33 that are highly active on these off-targets. Compounds were finally removed that were
34
35 predicted to be active with a probability of ≥ 0.7 on either of the three Aurora models.
36
37
38
39

40
41 **Prioritization of active compounds via MMP/ML and a QM/ML pipeline.** Two
42
43 complementary machine learning technologies were employed for prioritizing active
44
45 compounds. (1) a MMP/ML approach¹¹ which is trained on Matched Molecular Pairs (MMPs)
46
47 as descriptors, and considers the entire compound for ranking of compound in sets **A** and **B**
48
49 and (2) a hybrid QM/ML pipeline which is trained on quantum mechanical calculations and
50
51 used to optimize the Phe-gatekeeper interaction of compounds in set **A**.
52
53

54
55 MMP/ML was trained by obtaining all compounds with reported IC_{50} , K_i or K_d measurements
56
57 against TrkA from ChEMBL.^{4,11} MMPs were extracted using an in-house implementation
58
59
60

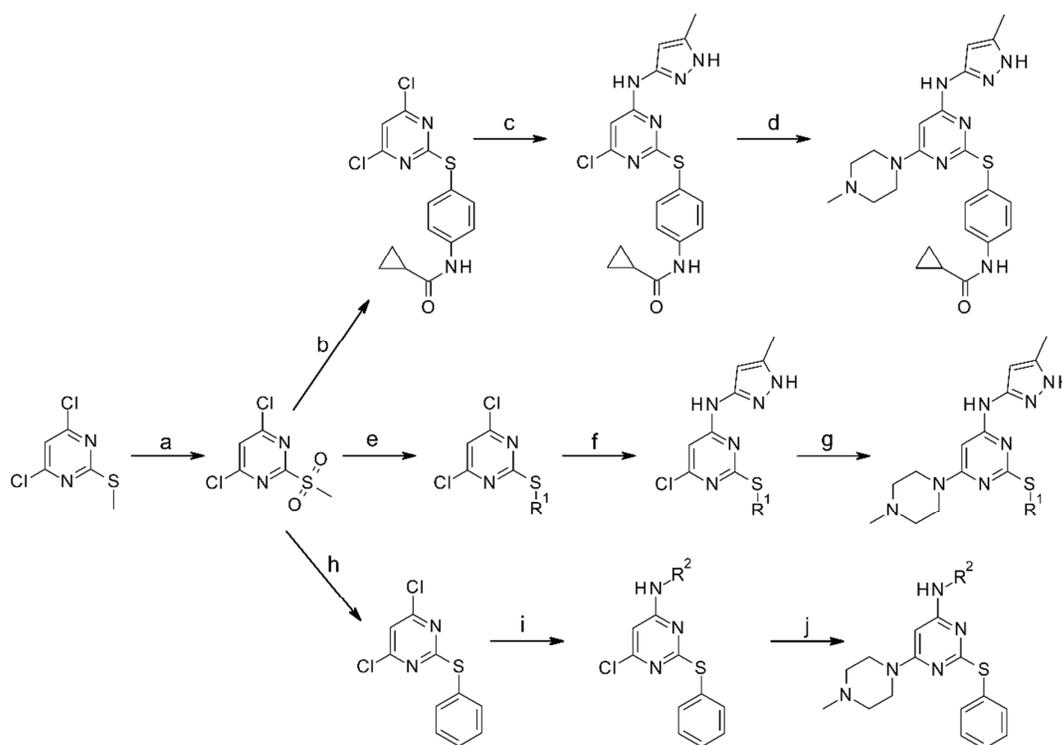
1
2
3 based on retrosynthetic rules (BRICS³⁶) and encoded as concatenated Morgan fingerprint³⁷
4 consisting of the static core that is shared between the two molecules forming an MMP and
5 both fragments F1 and F2 (describing the transformation).¹¹ A regression model was trained
6 on 23,000 fragment-based MMPs using a Deep Neural Network to predict the change in
7 activity (ΔpIC_{50}) associated with each MMP. Model validation was done using a “new
8 fragments” scenario, where fragments in the individual test sets inside a 5-fold cross
9 validation were new to the machine learning model (validation results see Supplementary
10 Information). The final prediction of change in activity was estimated by employing the
11 median ΔpIC_{50} of all five prediction models.
12
13
14
15
16
17
18
19
20
21
22

23 For the hybrid QM/ML pipeline, high-level QM calculations at the B3LYP-D3/6-31+G**
24 level were used to calculate pair interaction energies (PIE) between the gatekeeper residue
25 and a training set of compounds consisting of (1) 30 compounds selected from set **A** and (2)
26 50 benzyl-derivatives of tozasertib (**Table S4**). The benzyl-derivatives cover several
27 substituents and heterocycles (e.g., halogens, alkyls, and pyridyls) to model various
28 substituent effects. For each compound, the PIE was calculated from the gatekeeper-
29 compound dimer and the respective monomeric systems: $\text{PIE} = E_{\text{dimer}} - (E_{\text{gatekeeper}} +$
30 $E_{\text{compound}})$. Binding modes for set **A** were obtained via docking calculation, while the
31 benzyl-derivatives were modeled manually in the binding pocket starting from the tozasertib
32 binding mode and each minimized in the OPLS3 force field using Maestro from the
33 Schrödinger suite.³⁸ QM calculations were conducted using GAMESS-US.¹³ Fast Hartree–
34 Fock calculations at the HF/6-31+G** level were used to derive electronic features (dipole
35 moment, quadrupole moment, total self-consistent field energy) of the 80 ligands in the
36 training set and the compounds in set **A**. A Random Forest regression model predicting the
37 PIE from electronic ligand features was generated based on the training set using scikit-learn³⁹
38 and employed to prioritize the compounds from set **A** with respect to PIEs with the gatekeeper
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

residue. Hartree–Fock calculations and PIE predictions were applied on the interacting fragment only, i.e., not the entire ligand. The top 50 compounds were further characterized considering the entire binding pocket by fragment molecular orbital (FMO)¹² calculations at the MP2/6-31(+)-G* level (using docked binding poses and a polarizable continuum solvation water model) and sorted according to the sum of PIEs. FMO calculations were set up using Facio⁴⁰ and calculated using GAMESS-US.¹³

Chemistry

General synthetic route for tozasertib (a-d) and the compound sets **A** (a, i-j) and **B** (a, e-g) is described in Scheme S1.^a



^a Reagents and conditions: (a) 3-chloroperoxybenzoic acid, DCM, r.t., 3 h, yield: 95%; (b) N-(4-mercaptophenyl)cyclopropanecarboxamide, TEA, CH₃CN, 80 °C, 3-10 h, yield: 70%; (c) 5-methyl-1H-pyrazol-3-amine, DIPEA, DMF, 95 °C, 16 h, yield: 44%; (d) amine (1-

1
2
3 methylpiperazine or morpholine), DMF, DIPEA, 90 °C, 6-12 h, yield 43%; (e) corresponding
4
5 thiol, TEA, CH₃CN, 80 °C, 3-10 h, yields: 20-85%; (f) 5-methyl-1H-pyrazol-3-amine,
6
7 DIPEA, dioxane, 95 °C, 3-6 h, yields 50-75%; (g) amine (1-methylpiperazine or morpholine),
8
9 DMF, DIPEA, 90 °C, 6-12 h, yields: 30-70%; (h) thiophenol, TEA, THF, 50 °C, yields: 20-
10
11 85%; (i) corresponding amine, DIPEA, dioxane, 95 °C, 3-6 h, yields: 50 – 75%; (j) amine (1-
12
13 methylpiperazine or morpholine), DMF, DIPEA, 90 °C, 6-12 h, yields: 30 – 70%.

14
15
16
17 Final yields: tozasertib: 31%, reference 2: 41%, **a1**: 49%, **a2**: 37%, **a3**: 51%, **a4**: 31%, **a5**:
18
19 38%, **a6**: 51%, **a7**: 64%, **a8**: 34%, **a9**: 48%, **b1**: 58%, **b2**: 41%, **b3**: 53%, **b4**: 48%, **b5**: 51%,
20
21 **b6**: 32%, **b7**: 39%, **b8**: 41%, **b9**: 49%. All compounds were synthesized under contract by
22
23 Enamine Ltd. Purification was performed using HPLC (H₂O – MeOH; Agilent 1260 Infinity
24
25 systems equipped with DAD and mass-detectors. Waters Sunfire C18 OBD Prep Column,
26
27 100Å, 5 µm, 19 mm X 100 mm with SunFire C18 Prep Guard Cartridge, 100Å, 10 µm, 19
28
29 mm X 10 mm).

30
31
32
33 Compound characterization was done with ¹H NMR and LC/MS. ¹H NMR was performed
34
35 using Bruker AVANCE DRX 500 and Varian UNITYplus 400. LC/MS was performed using
36
37 Agilent 1100 Series LC/MSD system with DAD\ELSD and Agilent LC\MSD VL (G1956A),
38
39 SL (G1956B) mass-spectrometer and Agilent 1200 Series LC/MSD system with DAD\ELSD
40
41 and Agilent LC\MSD SL (G6130A), SL (G6140A) mass-spectrometer. Both systems used
42
43 Zorbax SB-C18 1.8 µm 4.6x15mm Rapid Resolution cartridge. Mobile phase was A –
44
45 acetonitrile, 0.1% formic acid, and B – water (0.1% formic acid) with gradient (0 min – 100%
46
47 B, 0.01 min – 100% B, 1.5 min - 0% B, 1.8 min - 0% B, 1.81 min - 100% B) and flow rate of
48
49 3 ml/min. Detection was performed using APCI ionization mode and scan range of m/z 80 -
50
51 1000. The purities of all final products of the two reference compounds (tozasertib, reference
52
53 2) and of 14 out of 18 reported compounds were found to be >95%. The remaining 4
54
55
56
57
58
59
60

1
2
3 compounds (i.e. **a2**, **b2**, **b4**, **b8**) have a purity between 90 and 95%. All reported compounds
4
5 passed common PAINS flags,⁴¹ as calculated using RDKit.
6
7

8 **Experimental Assays**

9

10 The primary screen (TrkA at 10 and 100 nM; AurA at 100 nM and 1 μ M concentration), Kd
11 measurements, profiling over a panel of 97 kinases (tozasertib, **a1**, **a4**, **b7**, **b8** at 100 nM; **a1**
12 additionally at 1 μ M concentration), and cellular assay (**a1** and **b7**) were done using the
13 KINOMEScan⁴² and PathHunter technologies from DiscoverX, respectively. Results of the
14 primary screen were reported as % of control (% Ctrl. = (test compound signal – positive
15 control signal) / (DMSO signal – positive control signal). The profiling was done via the
16 scanEDGE assay panel whereat KIT(D816V) and KIT (V559D, T670I) were replaced by
17 TrkB and TrkC. See www.discoverx.com for details on the employed profiling assays.
18 Selectivity scores (number of inhibited kinases divided by the total number of tested wild-type
19 kinases) were calculated at activity cut-offs of 35% and 10% of control, respectively.
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge via the Internet at <http://pubs.acs.org>.

It contains Figures S1-S2 with profiling results, Table S1-S4 with results of primary screens, selectivity scores of top hits, the data set used to train the QM/ML approach as well as additional information about the validation results of the employed prediction tools, about the biological assays and assay results, and analytic data. Molecular formula strings for all chemical structures mentioned (CSV) and the predicted binding pose of **a1** (PDB) are provided as data files.

AUTHOR INFORMATION

Corresponding Author

* E-mail: fulle@bio.mx

ORCID

Samo Turk: 0000-0003-2044-7670

Benjamin Merget: 0000-0002-3882-596X

Sameh Eid: 0000-0002-4689-6354

Simone Fulle: 0000-0002-7646-5889

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. ‡The authors ST, BM, and SE contributed equally to this work as a first author.

Funding Sources

The project was supported by the BioMed X Innovation Center, Heidelberg.

Notes

1
2
3 The authors declare no competing financial interest.

4
5 ACKNOWLEDGMENT

6
7 We thank Christian Tidona and Ann De Beuckelaer (BioMed X) for fruitful discussions.

8
9
10 ABBREVIATIONS

11
12 TrkA: Tropomyosin receptor kinase A; AurA: Aurora kinase A; ML: Machine Learning;

13
14 MMPs: Matched Molecular Pairs; QM: Quantum Mechanics
15
16
17
18
19

20
21 REFERENCES

- 22
23 (1) Duong-ly, K. C.; Devarajan, K.; Liang, S.; Horiuchi, K. Y.; Wang, Y.; Ma, H.;
24 Peterson, J. R.; Duong-ly, K. C.; Devarajan, K.; Liang, S.; Horiuchi, K. Y.; Wang, Y.;
25 Ma, H. Kinase Inhibitor Profiling Reveals Unexpected Opportunities to Inhibit
26 Disease-Associated Mutant Kinases. *CellReports* **2016**, *14* (4), 772–781.
27
28 (2) Oprea, T. I.; Overington, J. P. Computational and Practical Aspects of Drug
29 Repositioning. *Assay Drug Dev. Technol.* **2015**, *3* (6), 299–306.
30
31 (3) Metz, J. T.; Johnson, E. F.; Soni, N. B.; Merta, P. J.; Kifle, L.; Hajduk, P. J. Navigating
32 the Kinome. *Nat. Chem. Biol.* **2011**, *7* (4), 200–202.
33
34 (4) Bento, A. P.; Gaulton, A.; Hersey, A.; Bellis, L. J.; Chambers, J.; Davies, M.; Krüger,
35 F. A.; Light, Y.; Mak, L.; McGlinchey, S.; Nowotka, M.; Papadatos, G.; Santos, R.;
36 Overington, J. P. The ChEMBL Bioactivity Database: An Update. *Nucleic Acids Res.*
37 **2014**, *42* (D1), D1083-1090.
38
39 (5) Davis, M. I.; Hunt, J. P.; Herrgard, S.; Ciceri, P.; Wodicka, L. M.; Pallares, G.; Hocker,
40 M.; Treiber, D. K.; Zarrinkar, P. P. Comprehensive Analysis of Kinase Inhibitor
41 Selectivity. *Nat. Biotechnol.* **2011**, *29* (11), 1046–1051.
42
43 (6) Anastasiadis, T.; Deacon, S. W.; Devarajan, K.; Ma, H.; Peterson, J. R.
44 Comprehensive Assay of Kinase Catalytic Activity Reveals Features of Kinase
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- Inhibitor Selectivity. *Nat. Biotechnol.* **2011**, *29* (11), 1039–1045.
- (7) Falchook, G. S.; Bastida, C. C.; Kurzrock, R. Aurora Kinase Inhibitors in Oncology Clinical Trials: Current State of the Progress. *Semin. Oncol.* **2015**, *42* (6), 832–848.
- (8) Norman, B. H.; McDermott, J. S. Targeting the Nerve Growth Factor (NGF) Pathway in Drug Discovery. Potential Applications to New Therapies for Chronic Pain. *J. Med. Chem.* **2017**, *60* (1), 66–88.
- (9) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shaw, D. E.; Shelley, M.; Perry, J. K.; Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47*, 1739–1749.
- (10) Merget, B.; Turk, S.; Eid, S.; Rippmann, F.; Fulle, S. Profiling Prediction of Kinase Inhibitors: Toward the Virtual Assay. *J. Med. Chem.* **2017**, *60* (1), 474–485.
- (11) Turk, S.; Merget, B.; Rippmann, F.; Fulle, S. Coupling Matched Molecular Pairs with Machine Learning for Virtual Compound Optimization. *J. Chem. Inf. Model.* **2017**, *57* (12), 3079–3085.
- (12) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. Fragment Molecular Orbital Method: An Approximate Computational Method for Large Molecules. *Chem. Phys. Lett.* **1999**, *313* (3), 701–706.
- (13) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Jr, J. A. M. General Atomic and Molecular Electronic Structure System. *J. Comput. Chem.* **1993**, *14* (11), 1347–1363.
- (14) Oliveira, M. Managing Common Toxicities with New Tyrosine Kinase Inhibitors. *CancerWorld* **2015**, No. December, 31–37.
- (15) O'shea, J. J.; Kontzias, A.; Yamaoka, K.; Tanaka, Y.; Laurence, A. Janus Kinase Inhibitors in Autoimmune Diseases. *Ann. Rheum. Dis.* **2013**, *72* (2), 1–11.

- 1
2
3 (16) Wang, T.; Yu, D.; Lamb, M. L. Trk Kinase Inhibitors as New Treatments for Cancer
4 and Pain. *Expert Opin. Ther. Pat.* **2009**, *19* (3), 305–319.
5
6
7 (17) McCarthy, C.; Walker, E. Tropomyosin Receptor Kinase Inhibitors: A Patent Update
8 2009 - 2013. *Expert Opin. Ther. Pat.* **2014**, *7*, 731–744.
9
10
11 (18) Bailey, J.; Schirmacher, R.; Farrell, K.; Bernard-Gauthier, V. Tropomyosin Receptor
12 Kinase Inhibitors: An Updated Patent Review for 2010-2016 - Part I. *Expert Opin.*
13 *Ther. Pat.* **2017**, *6*, 733–751.
14
15
16 (19) Bailey, J.; Schirmacher, R.; Farrell, K.; Bernard-Gauthier, V. Tropomyosin Receptor
17 Kinase Inhibitors: An Updated Patent Review for 2010-2016 - Part II. *Expert Opin.*
18 *Ther. Pat.* **2017**, *7*, 831–849.
19
20
21 (20) Thress, K.; MacIntyre, T.; Wang, H.; Whitston, D.; Liu, Z.-Y.; Hoffmann, E.; Wang,
22 T.; Brown, J. L.; Webster, K.; Omer, C.; Zage, P. E.; Zeng, L.; Zweidler-McKay, P. A.
23 Identification and Preclinical Characterization of AZ-23, a Novel, Selective, and Orally
24 Bioavailable Inhibitor of the Trk Kinase Pathway. *Mol. Cancer Ther.* **2009**, *8* (7),
25 1818–1827.
26
27
28 (21) Bakri, F. G.; Wahbeh, A.; Abu Sneina, A.; Al Khader, A.; Obeidat, F.; AlAwwa, I.;
29 Buni, M.; Ki, C.-S.; Masri, A. Congenital Insensitivity to Pain and Anhidrosis due to a
30 Rare Mutation and That Is Complicated by Inflammatory Bowel Disease and
31 Amyloidosis: A Case Report. *Clin. Case Reports* **2016**, *4* (10), 997–1000.
32
33
34 (22) Nwosu, L. N.; Mapp, P. I.; Chapman, V.; Walsh, D. A. Blocking the Tropomyosin
35 Receptor Kinase A (TrkA) Receptor Inhibits Pain Behaviour in Two Rat Models of
36 Osteoarthritis. *Ann. Rheum. Dis.* **2016**, *57* (6), 1246–1254.
37
38
39 (23) Andrews, S. W. Allosteric Small Molecule Inhibitors of the NGF/TrkA Pathway, a
40 New Approach to Treating Inflammatory Pain. In *International Association for the*
41 *Study of Pain World Congress on Pain*; 2012.
42
43
44 (24) Wang, T.; Lamb, M. L.; Scott, D. A.; Wang, H.; Block, M. H.; Lyne, P. D.; Lee, J. W.;

- 1
2
3 Davies, A. M.; Zhang, H.-J.; Zhu, Y.; Gu, F.; Han, Y.; Wang, B.; Mohr, P. J.; Kaus, R.
4
5 J.; Josey, J. A.; Hoffmann, E.; Thress, K.; MacIntyre, T.; Wang, H.; Omer, C. A.; Yu,
6
7 D. Identification of 4-Aminopyrazolylpyrimidines as Potent Inhibitors of Trk Kinases.
8
9 *J. Med. Chem.* **2008**, *51* (15), 4672–4684.
10
11 (25) Stachel, S. J.; Sanders, J. M.; Henze, D. A.; Rudd, M. T.; Su, H.-P.; Li, Y.; Nanda, K.
12
13 K.; Egbertson, M. S.; Manley, P. J.; Jones, K. L. G.; Brnardic, E. J.; Green, A.;
14
15 Grobler, J. A.; Hanney, B.; Leitl, M.; Lai, M.-T.; Munshi, V.; Murphy, D.; Rickert, K.;
16
17 Riley, D.; Krasowska-Zoladek, A.; Daley, C.; Zuck, P.; Kane, S. A.; Bilodeau, M. T.
18
19 Maximizing Diversity from a Kinase Screen: Identification of Novel and Selective Pan-
20
21 Trk Inhibitors for Chronic Pain. *J. Med. Chem.* **2014**, *57* (13), 5800–5816.
22
23
24 (26) Burris, H. A.; Brose, M. S.; Shaw, A. T.; Bauer, T. M.; Farago, A. F.; Doebele, R. C.;
25
26 Smith, S.; Fernandes, M.; Cruickshank, S.; Low, J. A. A First-in-Human Study of
27
28 LOXO-101, a Highly Selective Inhibitor of the Tropomyosin Receptor Kinase (TRK)
29
30 Family. *J. Clin. Oncol.* **2015**, *33* (15\suppl), TPS2624-TPS2624.
31
32
33 (27) Skerratt, S. E.; Andrews, M.; Bagal, S. K.; Bilsland, J.; Brown, D.; Bungay, P. J.; Cole,
34
35 S.; Gibson, K. R.; Jones, R.; Morao, I.; Nedderman, A.; Omoto, K.; Robinson, C.;
36
37 Ryckmans, T.; Skinner, K.; Stupple, P.; Waldron, G. The Discovery of a Potent,
38
39 Selective, and Peripherally Restricted Pan-Trk Inhibitor (PF-06273340) for the
40
41 Treatment of Pain. *J. Med. Chem.* **2016**, *59* (22), 10084–10099.
42
43
44 (28) Bissantz, C.; Kuhn, B.; Stahl, M. A Medicinal Chemist's Guide to Molecular
45
46 Interactions. *J. Med. Chem.* **2010**, *53* (14), 5061–5084.
47
48
49 (29) Besnard, J.; Ruda, G. F.; Setola, V.; Abecassis, K.; Rodriguiz, R. M.; Huang, X.-P.;
50
51 Norval, S.; Sassano, M. F.; Shin, A. I.; Webster, L. a; Simeons, F. R. C.; Stojanovski,
52
53 L.; Prat, A.; Seidah, N. G.; Constam, D. B.; Bickerton, G. R.; Read, K. D.; Wetsel, W.
54
55 C.; Gilbert, I. H.; Roth, B. L.; Hopkins, A. L. Automated Design of Ligands to
56
57 Polypharmacological Profiles. *Nature* **2012**, *492* (7428), 215–220.
58
59
60

- 1
2
3 (30) Lounkine, E.; Keiser, M. J.; Whitebread, S.; Mikhailov, D.; Hamon, J.; Jenkins, J. L.;
4 Lavan, P.; Weber, E.; Doak, A. K.; Côté, S.; Shoichet, B. K.; Urban, L. Large-Scale
5 Prediction and Testing of Drug Activity on Side-Effect Targets. *Nature* **2012**, *486*
6 (7403), 361–367.
7
8
9
10
11 (31) Schneider, P.; Schneider, G. A Computational Method for Unveiling the Target
12 Promiscuity of Pharmacologically Active Compounds. *Angew. Chemie Int. Ed.* **2017**,
13 *56* (38), 11520–11524.
14
15
16
17 (32) Bowes, J.; Brown, A. J.; Hamon, J.; Jarolimek, W.; Sridhar, A.; Waldron, G.;
18 Whitebread, S. Reducing Safety-Related Drug Attrition: The Use of in Vitro
19 Pharmacological Profiling. *Nat. Rev. Drug Discov.* **2012**, *11* (12), 909–922.
20
21
22
23 (33) Eid, S.; Turk, S.; Volkamer, A.; Rippmann, F.; Fulle, S. KinMap: A Web-Based Tool
24 for Interactive Navigation through Human Kinome Data. *BMC Bioinformatics* **2017**,
25 *18*, 16.
26
27
28
29
30 (34) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.;
31 Olson, A. J. AutoDock4 and AutoDockTools4: Automated Docking with Selective
32 Receptor Flexibility. *J. Comput. Chem.* **2009**, *30* (16), 2785–2791.
33
34
35
36 (35) Reynolds, C. A.; Wade, R. C.; Goodford, P. J. Identifying Targets for Bioreductive
37 Agents: Using GRID to Predict Selective Binding Regions of Proteins. *J. Mol. Graph.*
38 **1989**, *7* (2), 103–108.
39
40
41
42 (36) Degen, J.; Wegscheid-Gerlach, C.; Zaliani, A.; Rarey, M. On the Art of Compiling and
43 Using “Drug-Like” Chemical Fragment Spaces. *ChemMedChem* **2008**, *3* (10), 1503–
44 1507.
45
46
47
48 (37) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**,
49 *50* (5), 742–754.
50
51
52
53 (38) Schrödinger Release 2017-1: Maestro, Schrödinger, LLC, New York, NY, 2017.
54
55
56 (39) Buitinck, L.; Louppe, G.; Blondel, M.; Pedregosa, F.; Mueller, A.; Grisel, O.; Niculae,
57
58
59
60

- 1
2
3 V.; Prettenhofer, P.; Gramfort, A.; Grobler, J.; Layton, R.; VanderPlas, J.; Joly, A.;
4
5 Holt, B.; Varoquaux, G. API Design for Machine Learning Software: Experiences from
6
7 the Scikit-Learn Project. In *ECML PKDD Workshop: Languages for Data Mining and*
8
9 *Machine Learning*; 2013; pp 108–122.
- 10
11 (40) Masahiko Suenaga. Development of GUI for GAMESS / FMO Calculation. *J. Comput.*
12
13 *Chem. Japan* **2008**, 7 (1), 33–54.
- 14
15 (41) Baell, J. B.; Holloway, G. A. New Substructure Filters for Removal of Pan Assay
16
17 Interference Compounds (PAINS) from Screening Libraries and for Their Exclusion in
18
19 Bioassays. *J. Med. Chem.* **2010**, 53 (7), 2719–2740.
- 20
21 (42) Fabian, M. A.; Biggs, W. H.; Treiber, D. K.; Atteridge, C. E.; Azimioara, M. D.;
22
23 Benedetti, M. G.; Carter, T. A.; Ciceri, P.; Edeen, P. T.; Floyd, M.; Ford, J. M.; Galvin,
24
25 M.; Gerlach, J. L.; Grotzfeld, R. M.; Herrgard, S.; Insko, D. E.; Insko, M. A.; Lai, A.
26
27 G.; L elias, J.-M.; Mehta, S. A.; Milanov, Z. V; Velasco, A. M.; Wodicka, L. M.; Patel,
28
29 H. K.; Zarrinkar, P. P.; Lockhart, D. J. A Small Molecule-Kinase Interaction Map for
30
31 Clinical Kinase Inhibitors. *Nat. Biotechnol.* **2005**, 23 (3), 329–336.
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 **TABLES**
4

5 **Table 1:** Summary of compound sets **A** and **B**.
6

7

Set	Number of compounds		Number of experimentally verified hits ¹	
	# calculated	# synthesized	# more selective	& TrkA < 10 nM
A	7404	9	4	2
B	6326	9	7	2

13

14 ¹ Number of compounds which are TrkA-selective vs. AurA [& which have in addition TrkA
15 activity < 10 nM in primary screen].
16
17
18

19
20
21 **Table 2:** Experimental affinity values of top hits.
22

23

ID	TrkA [pKd]	AurA [pKd]	Selectivity [Δ pKd]	Improved selectivity [$\Delta\Delta$ pKd]
Tozasertib	8.5	9.3	-0.8	--
Reference 2 ¹	7.5	7.4	0.1	0.9
a1	8.6	5.3	3.3	4.1
a4	8.8	7.0	1.8	2.6
b7	9.0	7.6	1.4	2.2
b8	8.4	7.3	1.1	1.9

32

33 ¹ Tozasertib without cyclopropylcarboxamide. **Reference 2** has an improved selectivity versus
34 tozasertib which is in line with the unoccupied unfavorable TrkA-selectivity area. Hit
35 compounds from sets **A** and **B** restore (or even improve) activity on TrkA without restoring
36 AurA activity.
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 3: PDB structures used to generate ‘selectivity grids’.

Kinase	PDB codes	# models ^a
TrkA	4aoj, <u>4pmt</u> ^b , 4yne, 4yps	6
AurA	1mq4, 1ol5, 1ol6, 1ol7, 2c6d, 2dwb, 2np8, 2w1c, 2w1d, 2w1e, 2w1f, 2w1g, 2wtw, 2x6d, 2x6e, 2xne, 2xng, 2xru, 3d14, 3d15, 3d2i, 3d2k, 3dj7, <u>3e5a</u> ^c , 3efw, 3fdn, 3h0z, 3ha6, 3k5u, 3m11, 3myg, 3nrm, 3o50, 3p9j, 3r21, 3r22, 3uo4, 3uo5, 3uod, 3up2, 3up7, 3vap, 4bn1, 4byi, 4c3p, 4c3r, 4ceg, 4dea, 4deb, 4ded, 4dee, 4dhf, 4j8m, 4j8n, 4jaj, 4jbo, 4jbp, 4o0s, 4o0u, 4o0w, 4prj, 4utd, 5aad, 5aae, 5aaf, 5aag, 5ew9	76
AurB	4af3, 4b8m	3

^a Total number of models includes individual chains and alternate models present in the PDB codes.

^b Used as reference structure for alignment.

^c PDB of co-crystallized tozasertib.

FIGURES

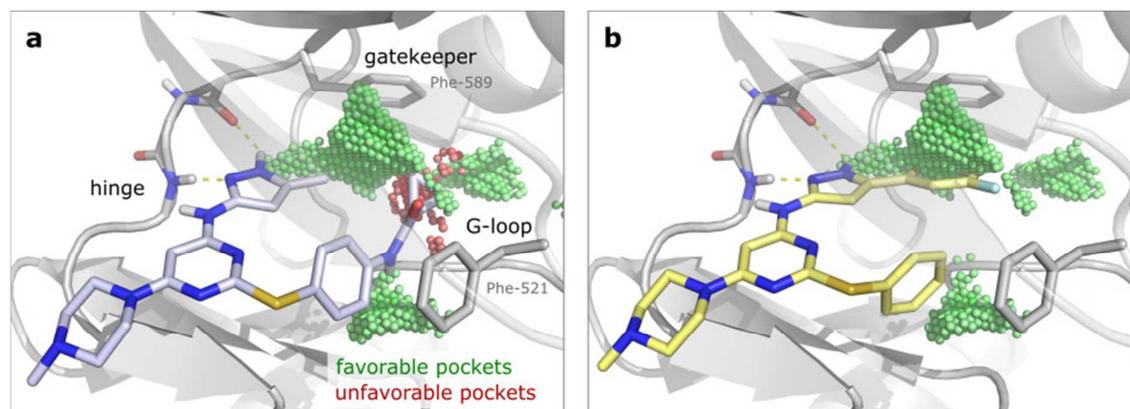


Figure 1: Selectivity hot-spot areas in TrkA binding site. Predicted binding modes of (a) tozasertib and (b) compound **a1** to TrkA (PDB code 4YNE). Compound **a1** occupies the favorable hydrophobic sub-pockets adjacent to the gatekeeper residue (green) but does not reach into the unfavorable area (red) occupied by the cyclopropyl moiety of tozasertib. The favorable sub-pockets adjacent to the Phe residue of the G-Loop is occupied by substituents in the **B** set. Note that Phe521 of the G-loop samples differing conformations in other TrkA structures and that the PDB structure 4YNE was chosen for binding mode prediction and structure-activity analysis because it resulted into the most favorable docking scores (i.e. when bound to compounds of the **B** series).

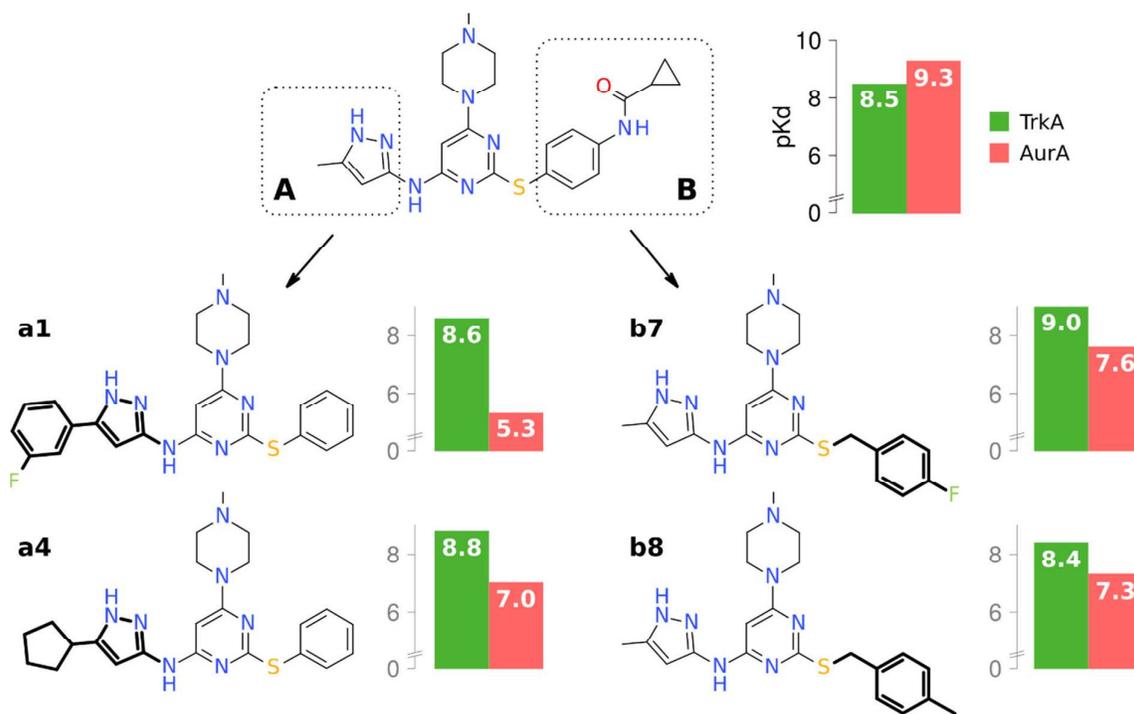


Figure 2: Summary of computationally designed hit compounds. Affinity values for tozasertib and the top hits are listed for the key target (TrkA; green) and off-target (AurA; red). Modifications compared to the starting compound are highlighted in bold lines.

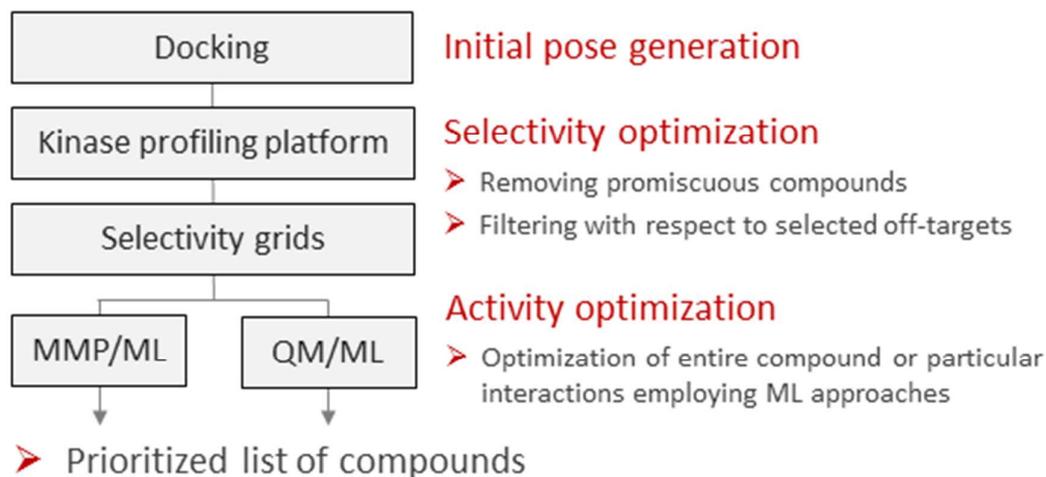


Figure 3: Multi-objective compound selection scheme. After initial binding pose generation via docking calculations, the compounds are prioritized via *in silico* methods that filter first with respect to selectivity criteria (i.e. the ligand-based ‘kinase profiling platform’ approach filters out promiscuous compounds, while the structure-based ‘selectivity grids’ approach filters for TrkA–Aurora selective compounds). Finally, highly active compounds on the key target are prioritized via two complementary machine learning technologies (i.e. using fragment-based Matched Molecular Pair descriptors and based on quantum mechanical calculations, respectively). The final selection does not rely on any manual filtering and includes in the case of the compound set **A** top ranking compounds from the MMP/ML and QM/ML technologies and in the case of the set **B** top ranking compounds from MMP/ML.

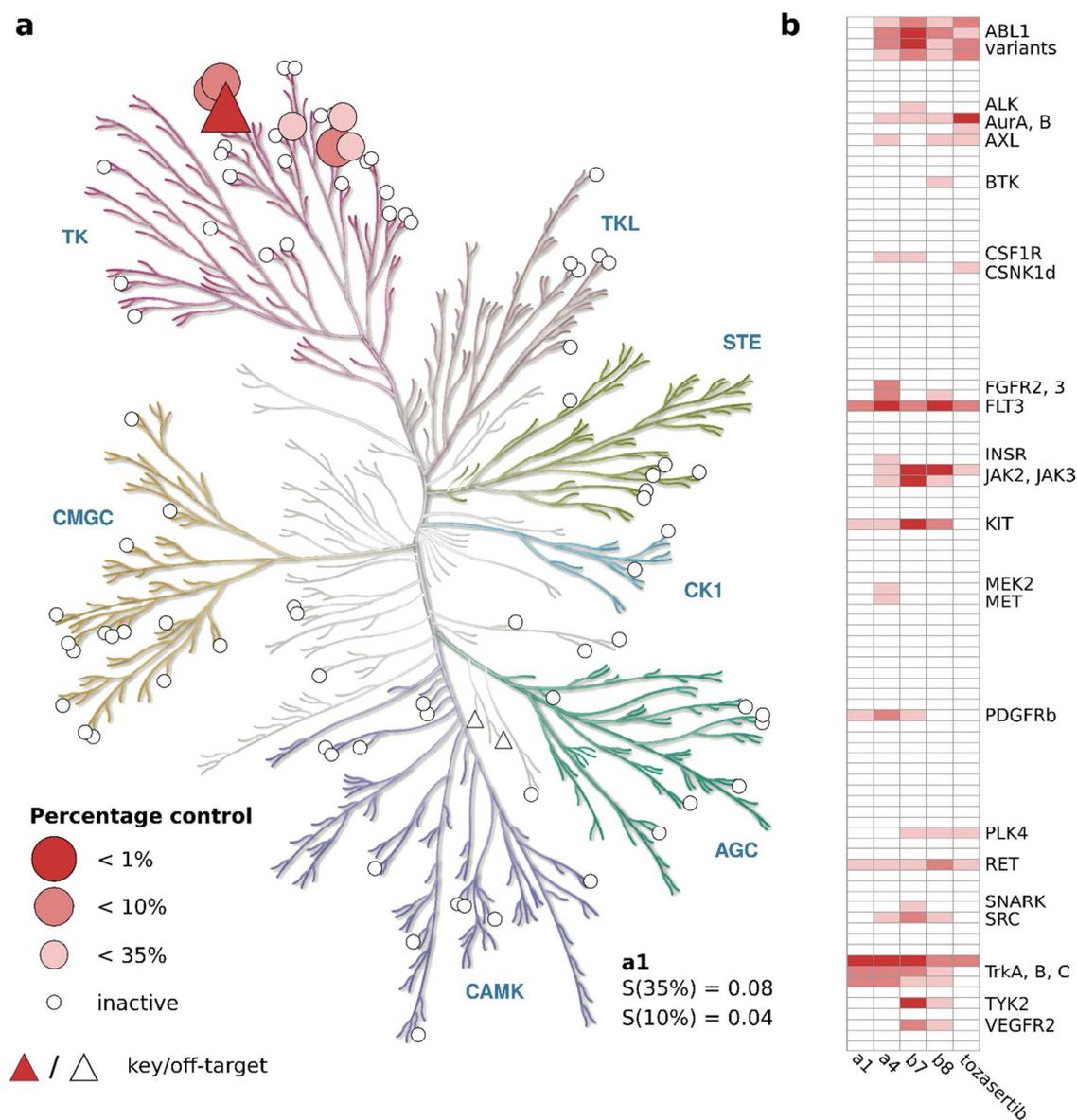
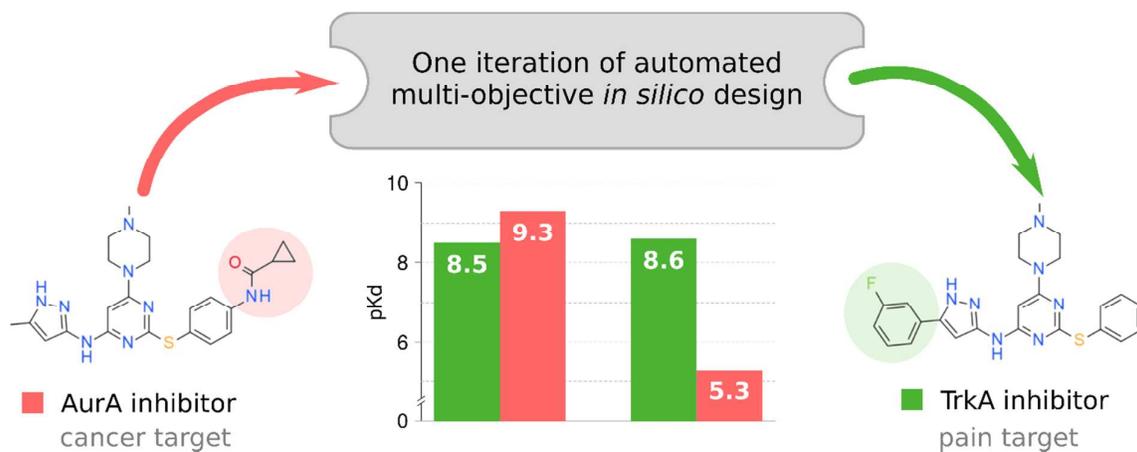
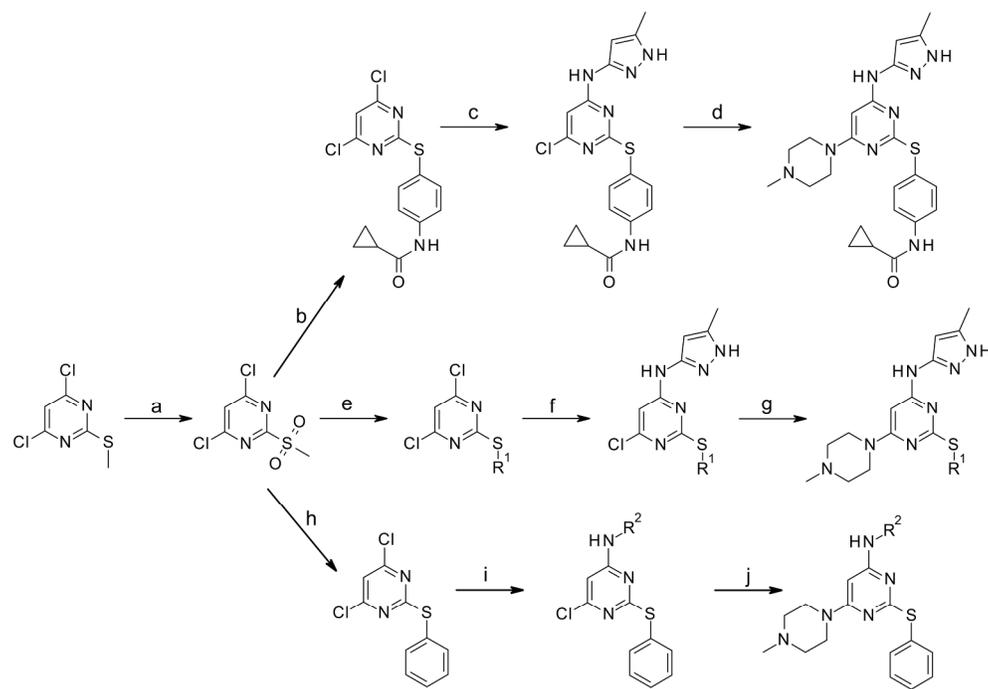
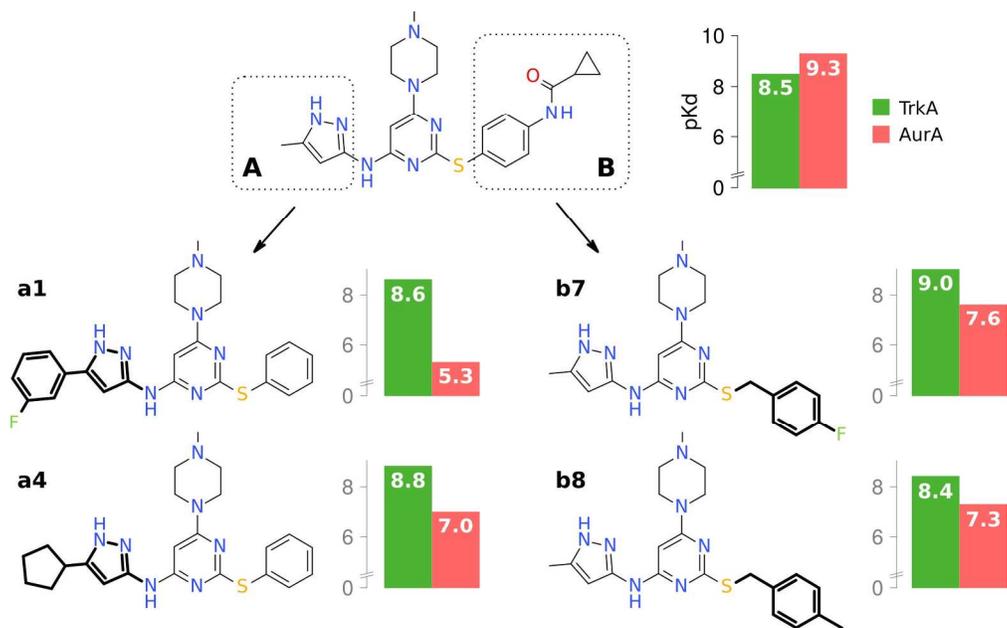


Figure 4: Profiling results. (a) Kinome profiling data of compound **a1**. With < 1% ctrl. compound **a1** is highly active only against the key target TrkA. Beside TrkA, six further kinases are inhibited at a screening concentration of 100 nM (i.e. TrkB, TrkC, FLT3, KIT, PDGFRb, RET), leading to a selectivity score of $S(35\%) = 0.08$. The compound is inactive on the off-targets AurA and AurB (white triangles). (b) Heatmap of kinase profiles of the hit compounds **a1**, **a4**, **b7**, and **b8**, as well as the starting compound tozasertib. The ABL1 variants include phosphorylated version of the ABL1 mutants E255K and T315I as well as the non-phosphorylated and phosphorylated versions of the ABL1 wild type.

TOC GRAPHIC

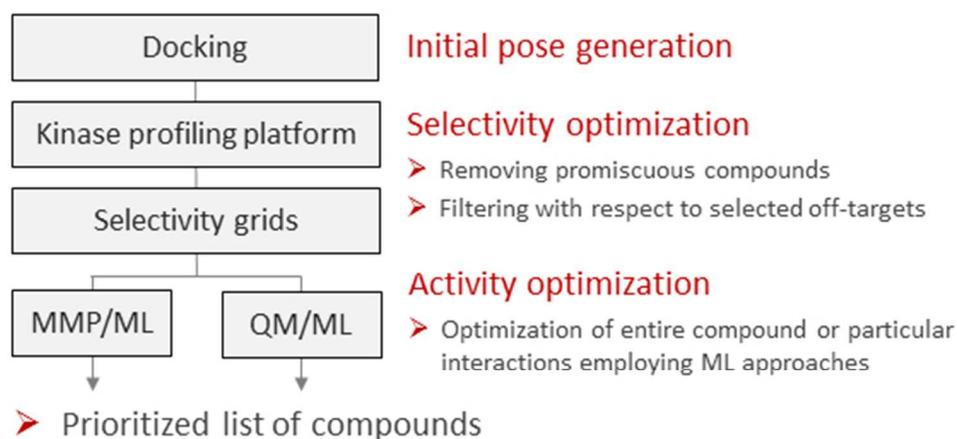






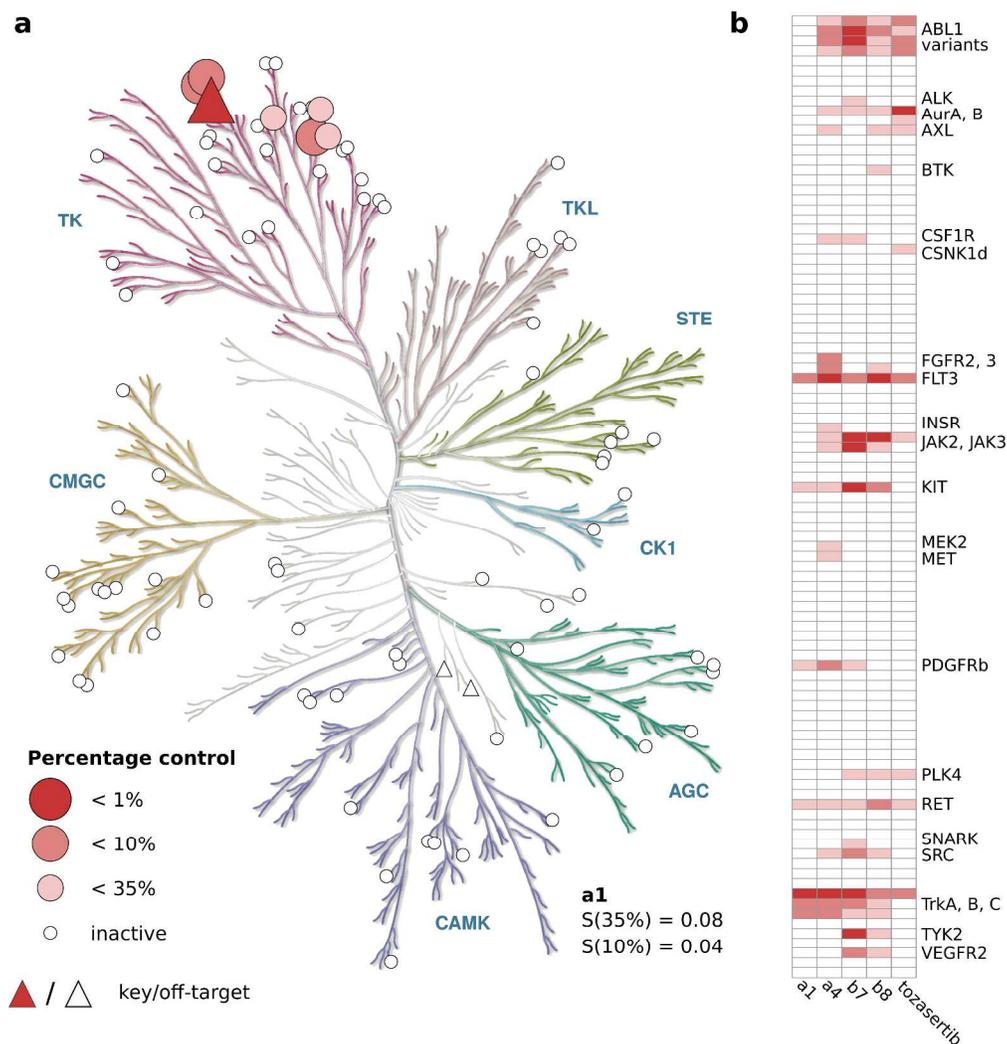
Summary of computationally designed hit compounds. Affinity values for tozasertib and the top hits are listed for the key target (TrkA; green) and off-target (AurA; red). Modifications compared to the starting compound are highlighted in bold lines.

192x119mm (300 x 300 DPI)



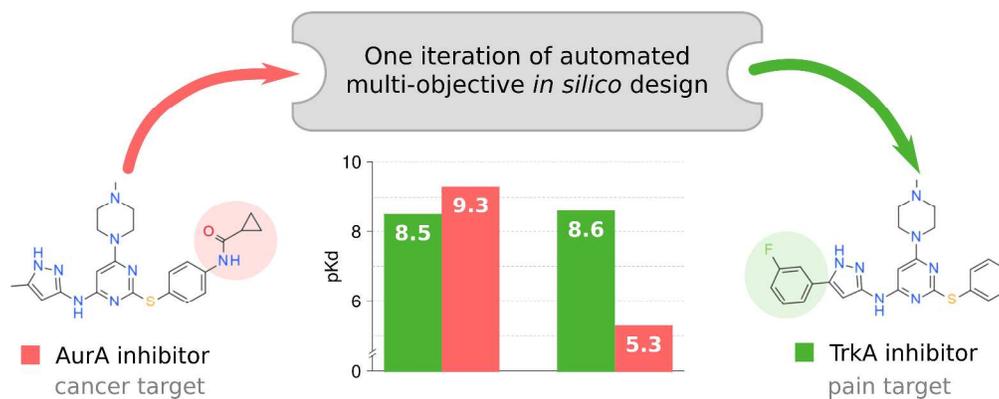
Multi-objective compound selection scheme. After initial binding pose generation via docking calculations, the compounds are prioritized via in silico methods that filter first with respect to selectivity criteria (i.e. the ligand-based 'kinase profiling platform' approach filters out promiscuous compounds, while the structure-based 'selectivity grids' approach filters for TrkA–Aurora selective compounds). Finally, highly active compounds on the key target are prioritized via two complementary machine learning technologies (i.e. using fragment-based Matched Molecular Pair descriptors and based on quantum mechanical calculations, respectively). The final selection does not rely on any manual filtering and includes in the case of the compound set A top ranking compounds from the MMP/ML and QM/ML technologies and in the case of the set B top ranking compounds from MMP/ML.

154x73mm (96 x 96 DPI)

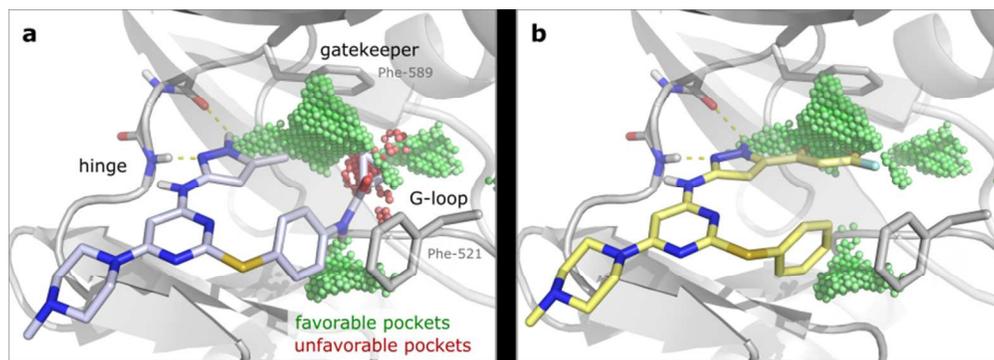


Profiling results. (a) Kinome profiling data of compound a1. With < 1% ctrl. compound a1 is highly active only against the key target TrkA. Beside TrkA, six further kinases are inhibited at a screening concentration of 100 nM (i.e. TrkB, TrkC, FLT3, KIT, PDGFRb, RET), leading to a selectivity score of $S(35\%) = 0.08$. The compound is inactive on the off-targets AurA and AurB (white triangles). (b) Heatmap of kinase profiles of the hit compounds a1, a4, b7, and b8, as well as the starting compound tozasertib. The ABL1 variants include phosphorylated version of the ABL1 mutants E255K and T315I as well as the non-phosphorylated and phosphorylated versions of the ABL1 wild type.

508x528mm (300 x 300 DPI)



471x182mm (300 x 300 DPI)



Selectivity hot-spot areas in TrkA binding site. Predicted binding modes of (a) tozasertib and (b) compound a1 to TrkA (PDB code 4YNE). Compound a1 occupies the favorable hydrophobic sub-pockets adjacent to the gatekeeper residue (green) but does not reach into the unfavorable area (red) occupied by the cyclopropyl moiety of tozasertib. The favorable sub-pockets adjacent to the Phe residue of the G-Loop is occupied by substituents in the B set. Note that Phe521 of the G-loop samples differing conformations in other TrkA structures and that the PDB structure 4YNE was chosen for binding mode prediction and structure-activity analysis because it resulted into the most favorable docking scores (i.e. when bound to compounds of the B series).

240x85mm (96 x 96 DPI)