Supporting Information

Drug Discovery Maps, a Machine Learning Model That Visualizes and Predicts Kinome–Inhibitor Interaction Landscapes

Antonius P.A. Janssen[1], Sebastian H. Grimm[1], Ruud H.M. Wijdeven[2], Eelke B. Lenselink[3], Jacques Neefjes[2], Constant A. A. van Boeckel[4], Gerard J.P. van Westen[3], Mario van der Stelt*[1]*

[1] Molecular Physiology, Leiden Institute of Chemistry, Leiden University, Leiden, 2333CC, The Netherlands

[2] Department of Cell and Chemical Biology, Leiden University Medical Centre, Leiden, 2333 ZC, The Netherlands

[3] Drug and Target Discovery, Leiden Academic Centre for Drug Research, Leiden University, Leiden, 2333CC, The Netherlands

[4] Pivot Park Screening Centre, Oss, 5349 AB, The Netherlands

*****e-mail: G.J.P. van Westen: gerard@lacdr.leidenuniv.nl, M. van der Stelt: m.van.der.stelt@chem.leidenuniv.nl

Table of contents

Supplementary Tables and Figures

Table S1: Summary of large experimental (partially) public kinome screens for sets of kinase inhibitors. Numbers of inhibitors and kinases are as reported in original publications and include mutants, splicing variants and separate kinase domains from the same kinase. Coverage represents the percentage of possible kinase-inhibitor pairs for which a data point is reported.

Set	No. of inhibitors	No. of kinases	Dosage	No. of data points	Coverage
Karaman *et al.*	38	317	Dose response**	12046	100%
Metz *et al.*	3858	172	Dose response	258094	38.9%
Anastassiadis *et al.*	178	300	Single dose	52834	98.9%
Davis *et al.*	72	442	Dose response**	31824	100%
Elkins *et al.*	367*	224	Two doses	81940	99.9%

* contains duplicate molecules

** only if activity is observed at initial 10 µM single dose screen

Table S2. Overview of prediction qualities split per compound. Data corresponds to *R* = 10 and *C* = 30%.

ChEMBL ID	True positives	False positives	True negatives	False negatives	Sensitivity	Specificity	PPV	NPV	MCC
CHEMBL10	4	8	254	13	0.24	0.97	0.33	0.95	0.24
CHEMBL101253	7	41	230	1	0.88	0.85	0.15	1.00	0.32
CHEMBL103667	5	21	228	25	0.17	0.92	0.19	0.90	0.09
CHEMBL119385	2	11	265	1	0.67	0.96	0.15	1.00	0.31
CHEMBL124660	9	66	202	2	0.82	0.75	0.12	0.99	0.25
CHEMBL1336	20	28	217	14	0.59	0.89	0.42	0.94	0.41
CHEMBL1421	19	5	207	48	0.28	0.98	0.79	0.81	0.40
CHEMBL14762	0	49	227	3	0.00	0.82	0.00	0.99	-0.05
CHEMBL1721885	35	29	147	69	0.34	0.84	0.55	0.68	0.20
CHEMBL191003	20	43	174	42	0.32	0.80	0.32	0.81	0.12
CHEMBL215152	1	5	258	15	0.06	0.98	0.17	0.95	0.07
CHEMBL221959	1	46	226	6	0.14	0.83	0.02	0.97	-0.01
CHEMBL223360	26	62	187	4	0.87	0.75	0.30	0.98	0.41
CHEMBL24828	20	16	217	26	0.43	0.93	0.56	0.89	0.41
CHEMBL259084	20	65	191	3	0.87	0.75	0.24	0.98	0.37
CHEMBL261849	4	74	201	0	1.00	0.73	0.05	1.00	0.19
CHEMBL278041	2	14	249	14	0.13	0.95	0.13	0.95	0.07
CHEMBL296468	2	3	250	24	0.08	0.99	0.40	0.91	0.14
CHEMBL31965	3	3	261	12	0.20	0.99	0.50	0.96	0.29
CHEMBL388978	57	3	42	179	0.24	0.93	0.95	0.19	0.16
CHEMBL428690	2	19	228	30	0.06	0.92	0.10	0.88	-0.02
CHEMBL440084	0	12	263	4	0.00	0.96	0.00	0.99	-0.03
CHEMBL477772	22	56	185	16	0.58	0.77	0.28	0.92	0.26
CHEMBL483321	2	4	272	1	0.67	0.99	0.33	1.00	0.46
CHEMBL522892	24	31	183	41	0.37	0.86	0.44	0.82	0.24
CHEMBL535	44	30	117	89	0.33	0.80	0.59	0.57	0.14
CHEMBL553	3	3	253	20	0.13	0.99	0.50	0.93	0.23
CHEMBL554	2	0	277	0	1.00	1.00	1.00	1.00	1.00
CHEMBL558752	10	72	185	12	0.45	0.72	0.12	0.94	0.10
CHEMBL572878	2	11	199	67	0.03	0.95	0.15	0.75	-0.05
CHEMBL572881	6	16	242	15	0.29	0.94	0.27	0.94	0.22
CHEMBL573339	0	58	219	2	0.00	0.79	0.00	0.99	-0.04
CHEMBL574738	48	12	165	54	0.47	0.93	0.80	0.75	0.47
CHEMBL607707	11	45	203	20	0.35	0.82	0.20	0.91	0.14
CHEMBL608533	11	10	169	89	0.11	0.94	0.52	0.66	0.10
CHEMBL91829	1	12	240	26	0.04	0.95	0.08	0.90	-0.01
CHEMBL939	3	15	254	7	0.30	0.94	0.17	0.97	0.18
CHEMBL941	10	115	153	1	0.91	0.57	0.08	0.99	0.19
AVERAGE	12.05	29.29	211.58	26.18	0.38	0.88	0.32	0.89	0.21

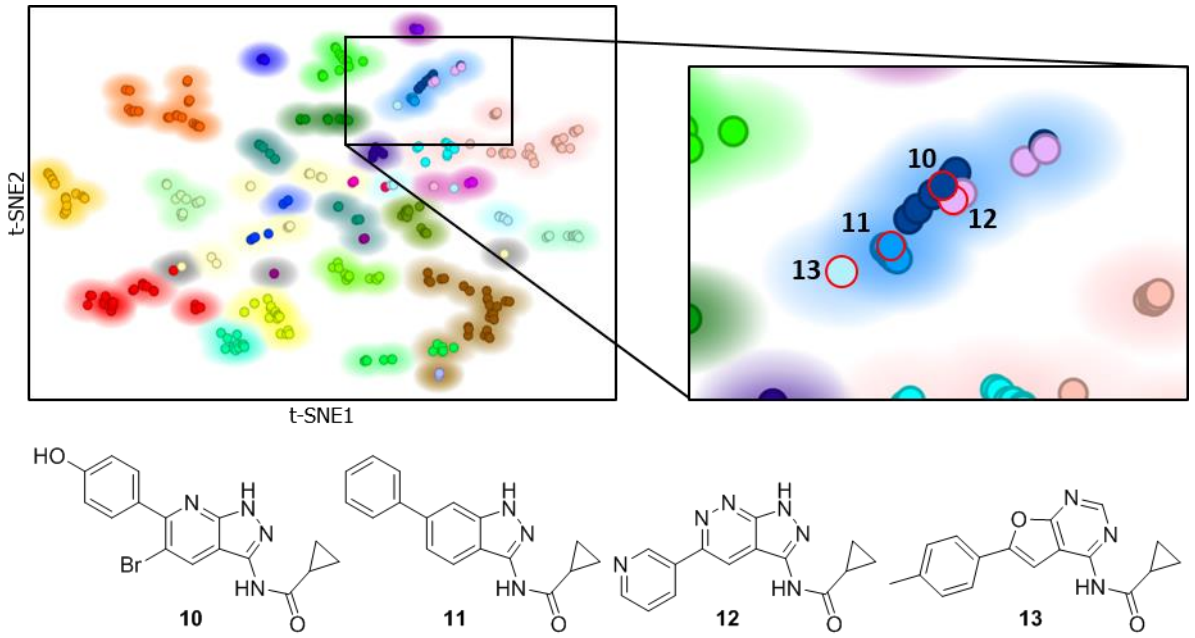


Figure S1: t-SNE embedding of the PKIS molecules. The mapping is based on their 4096-bit Morgan fingerprint (radius = 2) as in Figure 1A but including a background of the DBSCAN generated clustering, colored by the dominant chemotype in that cluster (grey are singletons or duos). t-SNE settings: perplexity = 50, learning rate = 50, iterations = 25,000. Markers are colored according to the 31 chemotypes defined by Elkins *et al.* Enlargement of the cluster dominated by 3-amino-pyrazolopyridines is shown, including four molecules of the three highly similar co-clustered chemotypes and the 'stray' sky blue compound **13**.

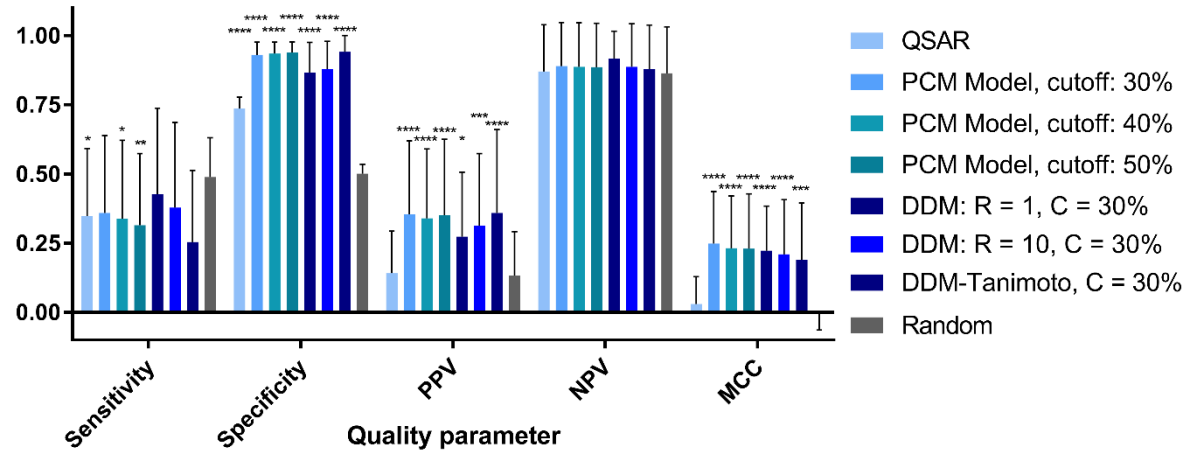


Figure S2: Statistical comparison of the new t-SNE based model with four models based on published procedures. Bars denote mean ± SD of the 38 inhibitors. Significant differences with Random have been highlighted, * P < 0.1, ** P < 0.01, *** P < 0.001, **** P < 0.0001 (Two-way ANOVA, Tukey's multiple comparison test). The relatively low PPV and high NPV for the random predictions are caused by the activity bias in the data set (13% active vs 87% inactive).

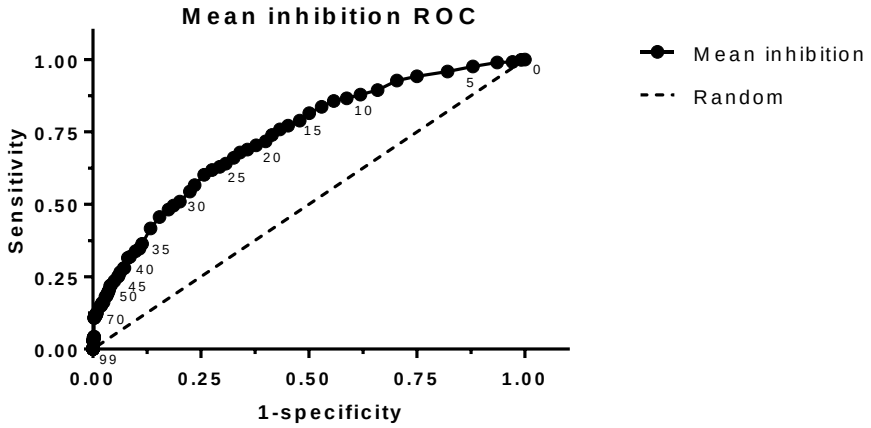


Figure S3: Receiver operating characteristic curve of the test set predictions with varying cut-off *C* values. The area under the curve is 0.76, with the maximum deviation from random at *C* values between 25 and 35%. Marker labels represent *C* values. Axes have been normalized to take into account the inability to predict all kinases based on the PKIS coverage.

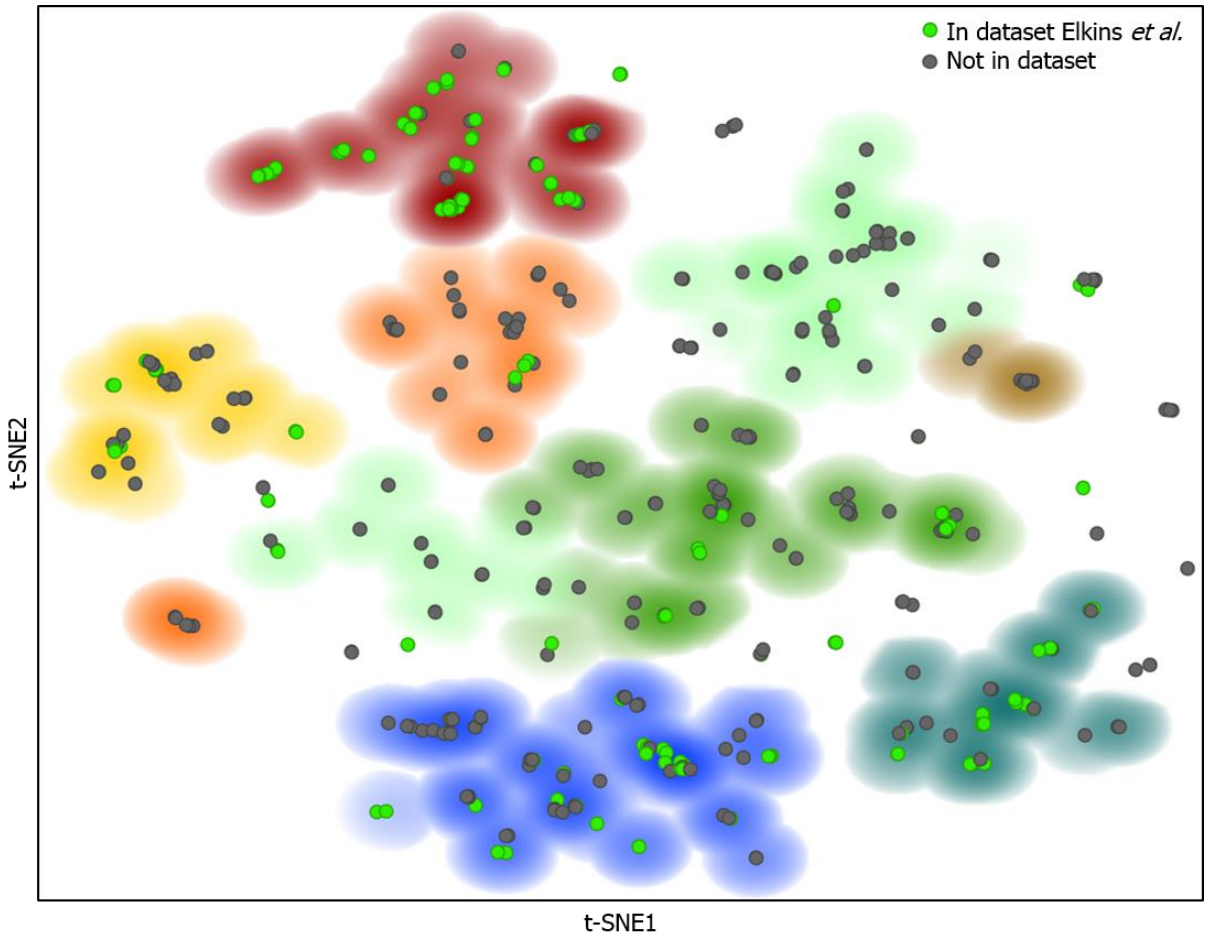


Figure S4: Inclusion of kinases in the Elkins *et al.* dataset. t-SNE embedding of physicochemical fingerprint of the kinase domains of 535 human kinases as in Figure 2. Markers are colored based on the presence (green) or absence (grey) in the Elkins *et al.* PKIS dataset.

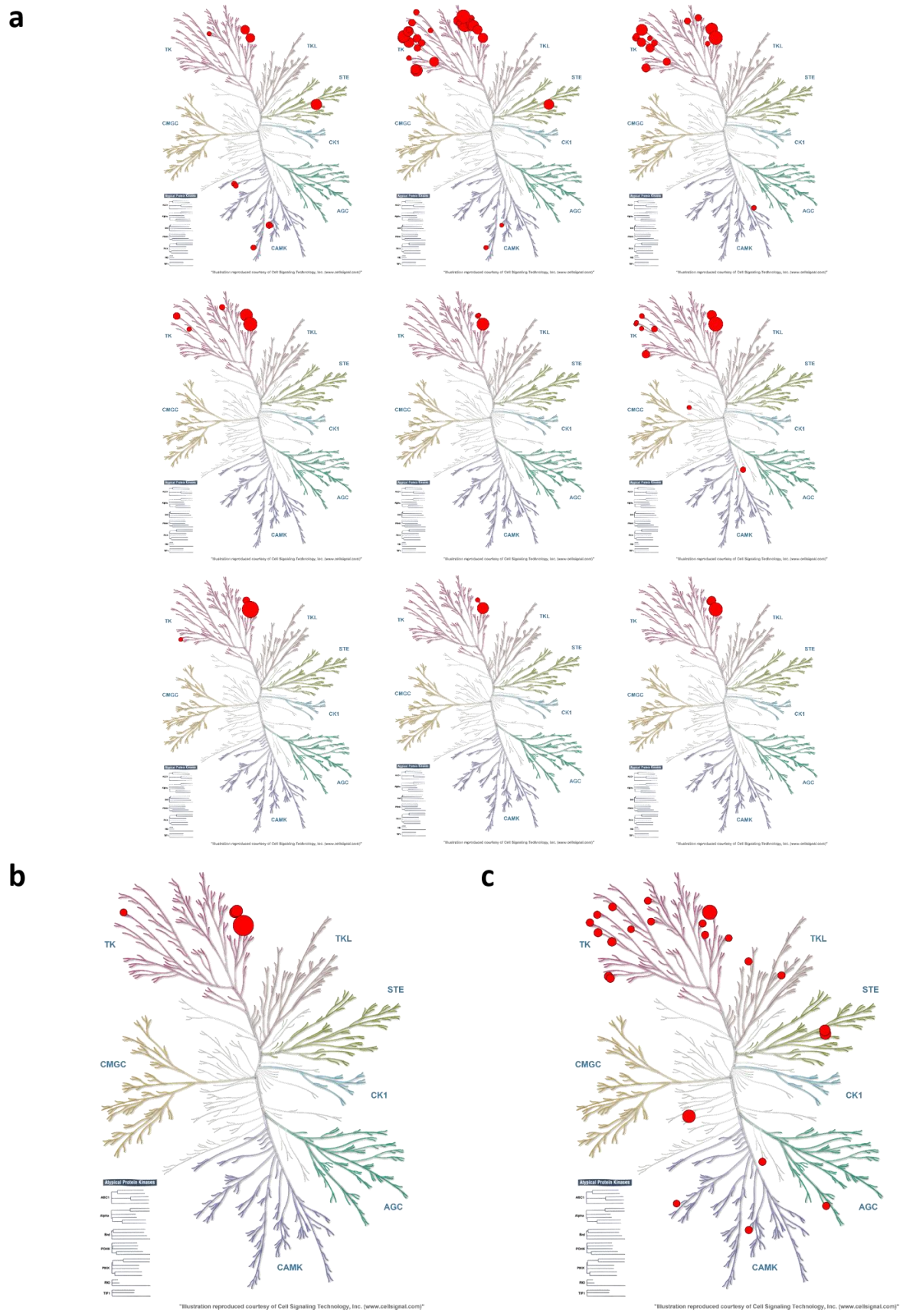


Figure S5: a) Inhibition profiles of the most similar molecules found by t-SNE, markers are scaled by the percentage of inhibition. b) Prediction inhibition profile of Erlotinib based on the mean inhibition in

a), markers are scaled by the mean percentage of inhibition. c) Actual inhibition of Erlotinib as determined by Karaman *et al*., markers are scaled by the pEC_{50}. Markers are shown if (mean) inhibition is >30% (a, b) or if pEC_{50} > 6 (c).

Table S3: Summary of the off-target screen performed for the *in situ* tested compounds **1** and **2**. Data are averages of two duplicates, and are percentages activity compared to vehicle control.

Compound 1				Compound 2			
Gene Name	**Uniprot accession**	**ChEMBL ID**	**Remaining activity (%)**	**Gene Name**	**Uniprot accession**	**ChEMBL ID**	**Remaining activity (%)**
ABL1	P00519	CHEMBL1862	26	ABL1	P00519	CHEMBL1862	23
ABL2	P42684	CHEMBL4014	12	ABL2	P42684	CHEMBL4014	28
AurA	O14965	CHEMBL4722	2	AurA	O14965	CHEMBL4722	60
AurB	Q96GD4	CHEMBL2185	2	AurB	Q96GD4	CHEMBL2185	60
AurC	Q9UQB9	CHEMBL3935	8	AurC	Q9UQB9	CHEMBL3935	91
BLK	P51451	CHEMBL2250	34	DDR1	Q08345	CHEMBL5319	7
BRK	Q13882	CHEMBL4601	61	DDR2	Q16832	CHEMBL5122	1
EGFR	P00533	CHEMBL203	96	EphA3	P29320	CHEMBL4954	9
EphA4	P54764	CHEMBL3988	72	EphA4	P54764	CHEMBL3988	28
EphA5	P54756	CHEMBL3987	52	EphA5	P54756	CHEMBL3987	7
EphB2	P29323	CHEMBL3290	14	EphA7	Q15375	CHEMBL4602	-2
EphB3	P54753	CHEMBL4901	100	EphB1	P54762	CHEMBL5072	3
ErbB2	P04626	CHEMBL1824	94	EphB2	P29323	CHEMBL3290	5
ErbB4	Q15303	CHEMBL3009	108	EphB3	P54753	CHEMBL4901	55
FER	P16591	CHEMBL3982	52	FER	P16591	CHEMBL3982	86
FES	P07332	CHEMBL5455	19	FES	P07332	CHEMBL5455	86
FGFR1	P11362	CHEMBL3650	58	FGFR1	P11362	CHEMBL3650	6
FGFR3	P22607	CHEMBL2742	48	FGFR3	P22607	CHEMBL2742	29
FGFR4	P22455	CHEMBL3973	91	FGFR4	P22455	CHEMBL3973	70
FGR	P09769	CHEMBL4454	30	FLT1	P17948	CHEMBL1868	1
FLT1	P17948	CHEMBL1868	1	FLT3	P36888	CHEMBL1974	1
FLT3	P36888	CHEMBL1974	3	FLT4	P35916	CHEMBL1955	-1
FLT4	P35916	CHEMBL1955	1	FMS	P07333	CHEMBL1844	-1
FMS	P07333	CHEMBL1844	-1	KDR	P35968	CHEMBL279	7
FYN	P06241	CHEMBL1841	9	KIT	P10721	CHEMBL1936	7
HCK	P08631	CHEMBL3234	17	MUSK	O15146	CHEMBL5684	13
KDR	P35968	CHEMBL279	5	PDGFRa	P16234	CHEMBL2007	15
KIT	P10721	CHEMBL1936	2	PDGFRb	P09619	CHEMBL1913	9
LCK	P06239	CHEMBL258	15	RET	P07949	CHEMBL2041	-2
LYN	P07948	CHEMBL3905	26	TIE2	Q02763	CHEMBL4128	1
PDGFRa	P16234	CHEMBL2007	47	TRKA	P04629	CHEMBL2815	5
PDGFRb	P09619	CHEMBL1913	46	TRKB	Q16620	CHEMBL4898	0
RET	P07949	CHEMBL2041	53	TRKC	Q16288	CHEMBL5608	-2
SRC	P12931	CHEMBL267	4				
TNK1*	Q13470	CHEMBL5334	-				
YES	P07947	CHEMBL2073	10				

*TNK1 was not available in the KinomeProfiler™

Setup of the DDM model

Trainings dataset selection

Several large datasets screening dozens of SMIs against hundreds of kinases have appeared in literature. Five datasets stand out when judged by size, these are summarized in Table S1. The first large dataset (>10,000 data points) published was that of Karaman *et al.* in 2008.[1] 38 commercially available kinase inhibitors were screened *in vitro* against 287 distinct kinases. In 2011, three large datasets were reported. Metz *et al*.[2] published a brief communication describing their analysis of more than 250,000 data points. However, a large portion of these interactions have not been made public. Later that same year, two papers were simultaneously published by Anastassiadis *et al.* and Davis *et al.*[3,4] Both studies comprise known kinase inhibitors and are screened against large portions of the kinome. The latest addition to the field is that by Elkins *et al*., who screened the Published Kinase Inhibitor Set (PKIS) by GlaxoSmithKline in an *in vitro* assay against 224 kinases, including a number of mutants.[5,6] This dataset is particularly interesting, as the PKIS molecules are available free of charge for academic research, and contains the most drug like molecules at the largest coverage. For these reasons, the dataset by Elkins *et al.* was chosen as a starting point for our data driven approach.

t-SNE for the Published Kinase Inhibitor Set

The curated dataset from ChEMBL 23[7] was used, as the original publication contained some duplicate molecules, and using the open source software package KNIME, the Morgan fingerprints (RD-Kit, 4096-bits, radius = 2) were generated.[7–9] These fingerprints were then clustered using the Python implementation of Barnes-Hut t-SNE.[10] Visual inspection of the embedding shows striking co-localization of the pre-defined chemotypes when the chemotype annotation by Elkins *et al.* is used post-hoc to color the markers (Figure 1B). To quantify the clustering quality, the embedding was analyzed using the DBSCAN algorithm, where the eps-parameter was optimized by maximizing the Silhouette Coefficient.[11] The best clustering was found for eps 0.9, for which the Adjusted Rand Index was 0.774. This unbiased clustering produced 33 clusters (Figure S1), 29 of which were significantly ($P < 0.0001$, hypergeometric test) enriched for a manually attributed chemotype. Of the 31 chemotypes annotated, 23 were fully comprised in one computationally assigned cluster. To inspect where it diverges from the human annotation, we inspected the cluster dominated by the 3-amino-pyrazolopyridines closer (Figure S1, enlargement). Inspection of the co-clustering of chemotypes shows that this cluster also contains all 3-amino-pyrazolopyridazines and the indazole-3-carboxamides, structurally very similar classes, as compounds **10**, **11** and **12** illustrate.

Nearest neighbor selection of molecules using t-SNE and initial target predictions

By appending a molecule to the PKIS dataset and regenerating the t-SNE mapping, simple Euclidian distances could be used to find its nearest neighbours. Selecting a set of neighbouring molecules should account for most chemical variation, leading to an ‘average molecule’ highly similar to our new molecule of interest.

For this set of neighboring molecules, the interaction landscapes against 200 unique non-mutant kinases have been experimentally measured, which can be averaged to yield a predicted inhibition value for the new molecule for all these kinases. For the PKIS dataset inhibition was measured at 100 nM and 1 µM. The inhibition values measured at 1 µM were used as these have the highest information density and in many clinically relevant experiments low micromolar concentrations can (locally) be reached, thus warranting a screen for off-targets at this concentration.

Expanding target prediction using t-SNE to find most similar kinases

To be able to extrapolate from the kinases measured in the PKIS set to incorporate more of the kinome, we again used a t-SNE based approach. Of the entire kinome, the amino acid sequences of the kinase domains containing the ATP-binding pocket, were aligned and expressed using a fingerprint based on physicochemical properties, derived by Heil *et al*.[12] The resulting t-SNE embedding is shown in Figure 2 of the main text, and reproduced with annotation of the inclusion in the PKIS set (Figure S4). With this similarity mapping in hand, the prediction based on the training dataset could be expanded to theoretically include the whole human kinome, by considering neighbouring kinases as plausible additional targets. The distribution of the measured kinases in the PKIS set is visualized in Figure S4 and is rather well, but certainly not homogeneous. This means that for fair parts of the kinome no truly reliable prediction can be made based only on this set, as there are no close neighbouring kinases measured. In our workflow all kinases are however still included, as this will allow any dataset to be loaded as training set, without large adaptations.

The DDM workflow

The workflow envisaged was briefly described in Figure 3, and is depicted in more detail in Figure S6: a molecule of interest is converted, via its SMILES representation, into a 4096-bit fingerprint. This fingerprint is appended to the PKIS molecules and a 2D t-SNE embedding is generated, as in Figure 1B. Based on this embedding, the closest molecules are selected and their measured bioactivity is averaged and considered as the activity prediction. Then, using the t-SNE embedding of the kinase domains (Figure 2), the kinases most similar to the predicted kinases are appended to this prediction, which is then the output of the model. Since the t-SNE algorithm is inherently stochastic, this whole process is repeated R times, after which the initial outputs are weighed and finally returned as overall output. The model thus accepts any molecular string representation and returns a list of predicted kinase targets, with a confidence parameter based on the number of repetitions in which a specific target has been found. The PKIS molecules the prediction is based upon can also be viewed, to assess with a chemical eye whether the prediction is to be trusted.

Several optimizable parameters naturally arise in this workflow: the number of considered neighbouring molecules N_m, the number of considered neighbouring kinases N_k, the number of repetitions of the whole process R, and the cut-off value C of the mean inhibition above which an inhibitor is deemed active against that kinase.

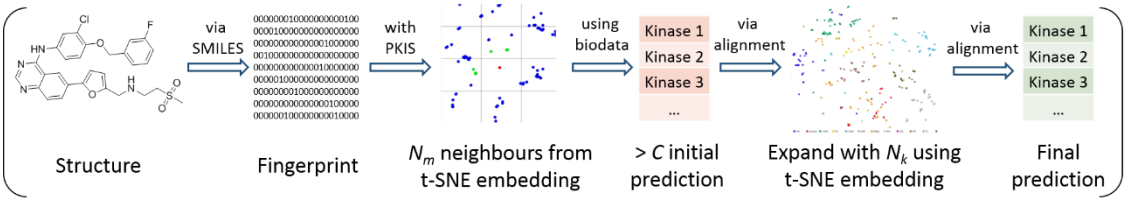


Figure S6: Schematic overview of the DDM model. A chemical structure is converted into a Morgan fingerprint using the SMILES-string, and appended to the PKIS molecules set. After t-SNE embedding the closest neighbour molecules (N_m) are identified and their biodata is averaged. A minimal average inhibition (C) is used as criterion to yield the initial prediction. The N_k kinases with the most similar kinase domain are then found in the kinase t-SNE embedding and appended to form the final prediction. This is repeated R times, weighing the final predictions to return a weighed final prediction.

Optimization of the model parameters

To optimize the model parameters and validate our model, we used the dataset generated by Karaman *et al*. as test set as this provided a comprehensive but diverse set of molecules, biochemically tested against a large set of kinases. For this test set, all K_d values below 1 µM were considered as actives, to mirror the training set, also measured at 1 µM. A multidimensional optimization was performed, initially varying the cut-off C and the two neighbouring values, N_m and N_k (Figure S7A and S7B). The number of repetitions was also optimized, as depicted in Figure S7C. The best predictions were found for 9 neighbouring molecules, 3 neighbouring kinases, and 10 repetitions. The cut-off percentage was found to be a valuable tuning parameter to either have a high sensitivity (low C) or rather a high positive predictive value (high C), which can be chosen depending on the projects specific demands. This is further illustrated by the ROC-curve in Figure S3. The mid-way value of C = 30% is recommended. The area under the ROC-curve is 0.76, scoring it as ‘fair’ according to standard criteria.

a

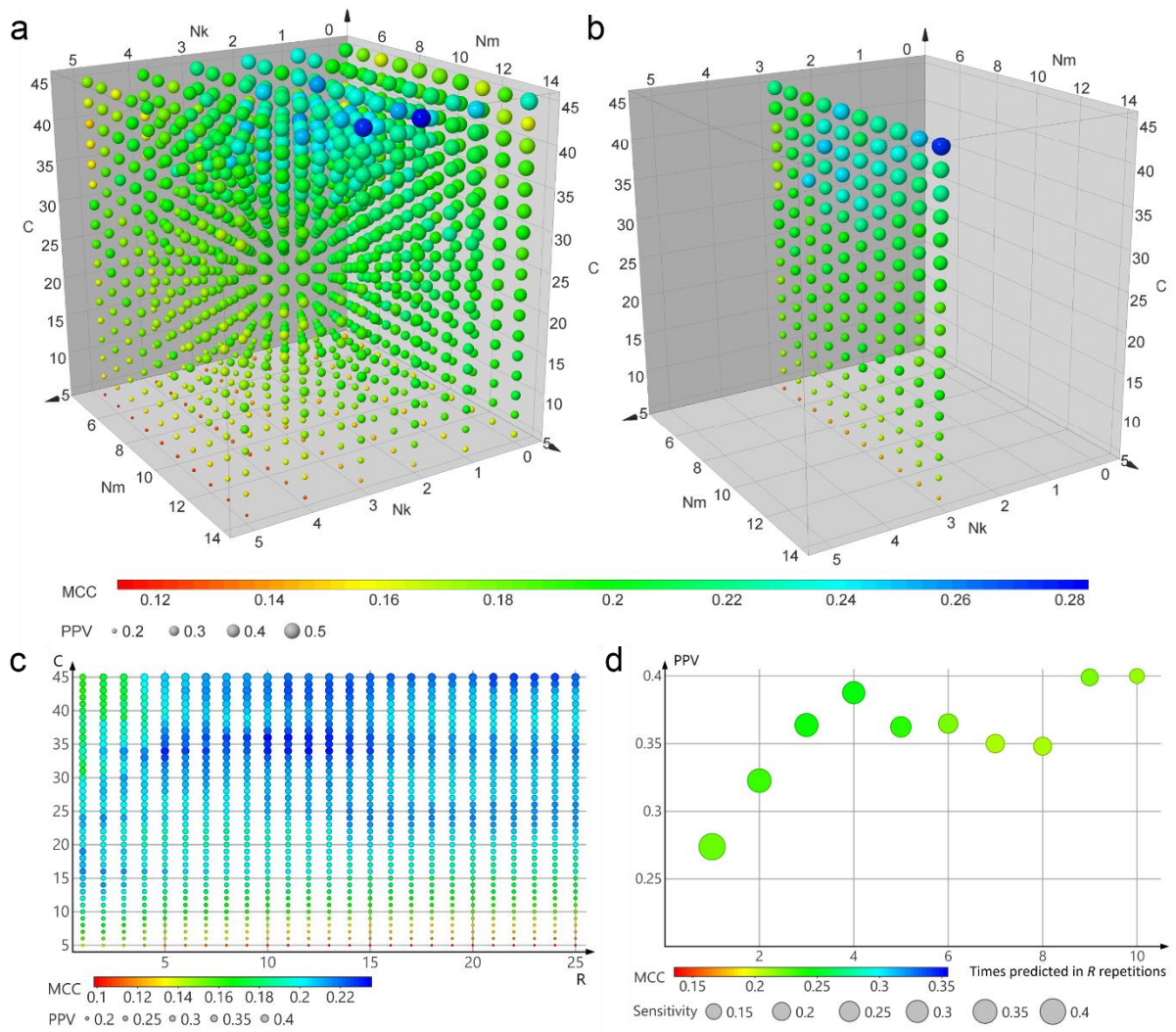


Figure S7: Optimization of the model parameters using the Karaman *et al*. dataset as test set. a) 3D-Optimisation based on the number of neighboring molecules used (N_m), the number of neighboring kinases used, N_k, and the cut-off inhibition percentage deemed active (*C*). Markers are sized based on the positive predictive value (PPV) and colored based on the Matthews Correlation Coefficient (MCC). b) As a), only the series for N_k = 3 is shown. c) 2D-optimisation of the number of repetitions R and cut-off *C* given N_k = 3 and N_m = 9. Markers are sized based on the positive predictive value (PPV) and colored based on the Matthews Correlation Coefficient (MCC). d) Visualization of the relation between the PPV and the sensitivity. Markers are sized based on the sensitivity and colored based on the Matthews Correlation Coefficient (MCC)

Comparison with state of the art methods

With this optimization of the model completed, the next step was to compare the new t-SNE based model with the state of the art in literature. To this end, we trained three proteochemometric (PCM) models and a QSAR model according to published procedures, and compared our method with those and a random model.[13,14] The result of this comparison is summarized in Figure S2, which shows that our model is significantly better than both the random model and the QSAR, and performs similarly when

compared to the PCM variants. We also included a variant of our own model where the standard Tanimoto distances were used to find the nearest molecular neighbours in the first step of our workflow. t-SNE was used in the kinase lookup step. This DDM-Tanimoto variant performed rather well, and not significantly worse than our t-SNE based approach, for this test set.

Bioactivity datasets used

Activity data for Karaman *et al.* and the Published Kinase Inhibitor Set by Elkins *et al.* was retrieved from the ChEMBL database version 23.[7,15] Data was retrieved from a local SQL install, directly from the website or through the KNIME extensions provided by the EMBL-EBI.

Kinase sequence information and bitstring

Sequence information of the kinase domains was retrieved from the KinBase situated on kinase.com, based on the paper by Manning *et al.*[16] Missing (pseudo)kinases were appended using data from the ChEMBL database version 23 and Uniprot.[17] The mapping of ChEMBL, Uniprot and KinBase identifiers was performed based on the KinBase website and extensive manual curation. The kinase domains were aligned using the online Clustal Omega tool provided by the EMBL-EBI.[18] The standard "Clustal w/o numbers" output generated was transformed to a bitstring using the amino acid fingerprints as provided in Heil *et al.*[12] with the following additions: alignment dashes (-), stops (*) and blanks (X) were all considered empty, represented by 23 0's.

t-SNE algorithm

All t-SNE embeddings were generated with the Python Scikit-learn (v. 0.19) implementation of the Barnes-Hut t-SNE algorithm, either implemented in a 'Python for KNIME' node or as part of a Python script.[10]

DBSCAN algorithm

The Python implementation of DBSCAN, available through the Scikit-learn module Cluster (v. 0.19) was used.[10] The clustering quality metrics Silhouette Coefficient and Adjusted Rand Index were calculated with the Metrics module and the former was optimized by tuning the eps-parameter in 0.05 steps. For the PKIS inhibitors a minimal cluster size of 1 was chosen, for the more disperse plotted kinases the minimal cluster size was set to 10. Comparisons to manual attributions was done manually. Statistical evaluation was performed using Microsoft Excel 2013.

Cheminformatics tools

All molecular descriptors, molecular representations (SMILES, InChIKey) and fingerprints were generated using the RDKit software, either using the KNIME extensions or as the Python implementation.[8]

QSAR and PCM models

QSAR and PCM models were trained as has been described before but for PCM using the fully aligned sequences rather than the binding site and using classification rather than regression.[13,14] PCM models were trained at an activity cut-off of 30%, 40%, and 50%. Pipeline Pilot 2016 (version 16.2.0.58 by BioVia) was used to process the data and random forests were trained in R (version 3.3, package randomforest) using 500 trees, equalizing class sizes, and randomly sampling the square root of the total present descriptors at individual splits.[19] Protein descriptors used were the first three z-scales with a mean value for each sequence for each z-scale.[20] Chemical descriptors used were circular fingerprints with a diameter of 6 bonds (FCFP_6) and physicochemical descriptors as was done previously.[14,21]

PCM models were trained on the full set of kinases and PKIs. QSAR models were trained per kinase but only if at least 5 active and 10 inactive PKIs per kinase where present (at a cut off of 30%). If no QSAR was trained missing predictions were completed to avoid bias and have the ability to compare identical prediction counts. This was obtained using a random number generator in the range of 0-1 where > 0.5 was deemed 'active' as was done previously.[14]

Out of Bag error estimates of PCM models trained on the full set were 8.47% (30% cut off), 8.15 % (40% cut off), and 7.83 % (50% cut off).

Code availability

The developed DDM workflow is available as an annotated KNIME workflow, an annotated Python script and an executable including a GUI based on the Python script. This can all be downloaded from the 4TU platform.

Docking of 1 and 2 in FLT3 crystal structure

Docking and structure based modeling was performed in the Schrödinger suite.[22] For **1** a DFG-in model was constructed on the basis of 4RT7 and 3LCD, in a similar fashion as has been done before,[23] using the knowledge based potential in prime.[24,25] Subsequently **1** was docked into this model using induced fit docking.[26] For the induced fit docking a hydrogen bond constraint on the backbone of Cys96 was used. Since **2** shares the same substructure as quizartinib, we used the crystal structure of FLT3 co-crystalized with quizartinib (4RT7)[27] as a starting point. **2** was docked using glide sp.[28] Ligand surfaces, structure renderings and 2D interaction plots were generated using Discovery Studio Visualizer v16.1.

Statistical methods

Clustering enrichment was analyzed using a hypergeometric test and was calculated using Microsoft Excel 2013. Significance was attributed only if $P < 0.0001$.

In the comparison of the quality of prediction of the various prediction models a regular 2-way ANOVA with Tukey's multiple comparison test was performed using GraphPad Prism 7. Significance is attributed according to the standard GraphPad style: * $P < 0.0332$, ** $P < 0.0021$, *** $P < 0.0002$, **** $P < 0.0001$.

High Throughput Screening FLT3

20 nL of 2 mM inhibitor solution in DMSO (row 9-48) or control (DMSO, row 1-8) was dispensed in a 1536-well plate. 2 µL assay buffer (50 mM HEPES (pH 7.5), 1 mM EGTA, 10 mM $MgCl_2$, 0.01% Tween-20, 2 mM DTT) without protein was loaded in row 5-6 as negative controls. Rows 1-4 and 7-48 were charged with 2 µL 0.75 ng/µL FLT3 in assay buffer. The plates were spun down for 30 seconds at 187.5x *g* and incubated for 30 minutes in the dark. Subsequently, 2 µL of substrate solution was added to all wells (50 mM HEPES (pH 7.5), 1 mM EGTA, 10 mM $MgCl_2$, 0.01% Tween-20, 2 mM DTT, 600 µM ATP, 12.5 nM Lance TK-peptide, 4 nM Lance anti-phosphotyrosine). The plates were spun down for 30 seconds at 187.5x *g* and incubated for 90 minutes in the dark at RT. Plates were then read on the Envision plate reader (Excitation 337 nm (laser), Emission first filter 615 nm, second filter 665 nm). Data was analysed using ActivityBase. Final assay concentrations: 10 µM inhibitor, 300 µM ATP, 0.365 ng/µL FLT3, 6.25 nM (0.5 K_M) Lance TK-peptide, 2 nM Lance anti-phosphotyrosine, 0.5% DMSO.

***In vitro* FRET based FLT3 assay**

In a 384-wells plate, 5 µL kinase+petide mix (0.06 ng/µL FLT3, 200 nM Lance TK-peptide) in assay buffer (50 mM HEPES (pH 7.5), 1 mM EGTA, 10 mM $MgCl_2$, 0.01% Tween-20, 2 mM DTT) was dispensed. Separately inhibitor solutions (10 µM – 0.1 pM) were prepared in assay buffer containing 400 µM ATP and 1% DMSO. 5 µL of these solutions was dispensed and the plate was incubated for 90 minutes in the dark. After 90 minutes the reaction was quenched by the addition of 10 µL of 20 mM EDTA containing 4 nM Lance anti-phosphotyrosine. After thorough mixing it was incubated for 60 minutes in the dark. The FRET fluorescence was measured on a Tecan Infinite M1000 Pro plate reader (excitation 320 nm, emission first filter 615 nm, second filter 665 nm). Data was processed using Microsoft Excel 2013, pIC_{50} values were fitted using GraphPad Prism 7.0. Final assay concentrations during phosphorylation: 200 µM ATP, 0.03 ng/µL FLT3, 100 nM Lance TK-peptide, 0.5% DMSO)

***In situ* testing of kinase inhibitors**

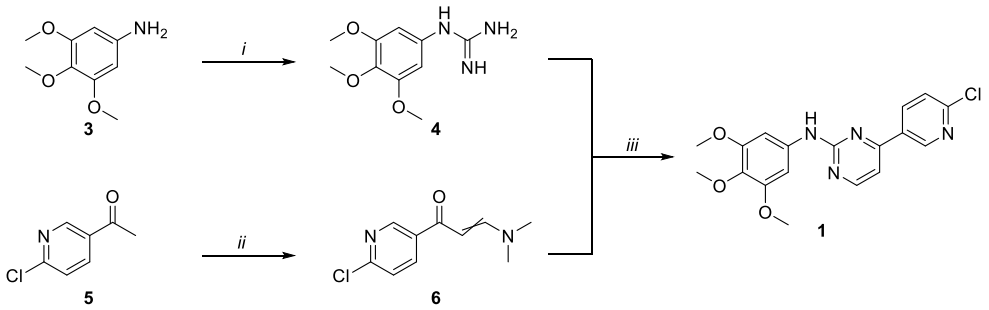
MV4-11 cells were grown in IMDM with 10% fetal-bovine-serum. U937 and Ba/F3 cells were grown in RPMI 10% fetal-bovine-serum. Ba/F3 wt cells were grown with additional IL-3 (10 ng/mL, PeproTech). All cells were cultured at 37°C under 5% CO_2. For viability assays, 10000 cells were seeded per well in a 96-wells plate and inhibitors were added at the indicated concentration. Three days later, cell viability was measured using the Cell Titer Blue viability assay (Promega), fluorescence was measured using the Clariostar (BMG Labtech). Relative survival was normalized to the untreated control and corrected for background signal.

Synthesis of *in situ* tested kinase inhibitors

General remarks

All reactions were performed using oven- or flame-dried glassware and dry solvents. Reagents were purchased from Sigma-Aldrich, Acros, and Merck and used without further purification unless noted otherwise. All moisture sensitive reactions were performed under a nitrogen atmosphere.

^{1}H and ^{13}C NMR spectra were recorded on a Bruker AV-400 (400 MHz). Used software for interpretation of NMR-data was Bruker TopSpin 1.3 and MestreNova 11.0. Chemical shift values are reported in ppm with tetramethylsilane or solvent resonance as the internal standard ($CDCl_3$: δ 7.26 for ^{1}H, δ 77.16 for ^{13}C; DMSO-*d6*: δ 2.50 for ^{1}H, δ 39.52 for ^{13}C). Data are reported as follows: chemical shifts (δ), multiplicity (s = singlet, d = doublet, dd = double doublet, td = triple doublet, t = triplet, q = quartet, quintet = quint, bs = broad singlet, m = multiplet), coupling constants *J* (Hz), and integration. Liquid chromatography was performed on a Finnigan Surveyor LC/MS system, equipped with a C18 column. Flash chromatography was performed using SiliCycle silica gel type SiliaFlash P60 (230−400 mesh). TLC analysis was performed on Merck silica gel 60/Kieselguhr F254, 0.25 mm. Compounds were visualized using $KMnO_4$ stain (K_2CO_3 (40 g), $KMnO_4$ (6 g), and water (600 mL)). Preparative HPLC (Waters, 515 HPLC pump M; Waters, 515 HPLC pump L; Waters, 2767 sample manager; Waters SFO System Fluidics Organizer; Waters Acquity Ultra Performance LC, SQ Detector; Waters Binary Gradient Module) was performed on a Waters XBridgeTM column (5 µM C18, 150 x 19 mm). Diode detection was done between 210 and 600 nm. Gradient: ACN in (H_2O + 0.2% TFA).



Scheme S1: Synthesis of screening hit **1**. Reagents and conditions: *i*) cyanamide, nitric acid, ethanol, 78 °C, 76%; *ii*) dimethylformamide diethylacetal, toluene, 80 °C, 80%; *iii*) K_2CO_3, ethanol, 78 °C, 31%.

Synthesis of 1-(3,4,5-trimethoxyphenyl)guanidine (**4**)

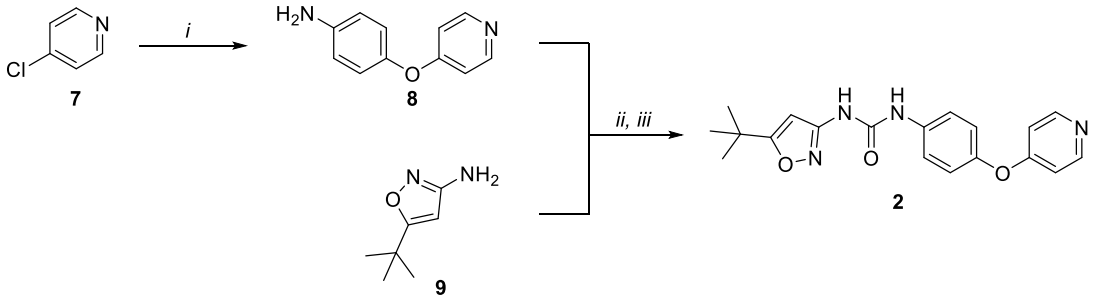
3,4,5-trimethoxyaniline **3** (500 mg, 2.73 mmol) and cyanamide (574 mg, 13.65 mmol) were dissolved in ethanol (15 mL) before nitric acid (69%$_{wt}$, 0.20 mL, 3.0 mmol) was added. The mixture was refluxed for 40 h and concentrated *in vacuo*. The resulting residue was suspended in diethylether (10 mL) and kept at 4 °C overnight. It was then filtered and air dried. Amino((3,4,5-trimethoxyphenyl)amino)-methaniminium nitrate (0.60 g, 2.1 mmol, 76%) was obtained as a dark purple solid. ^{1}H NMR (400 MHz, DMSO) δ 9.51 (bs, 1H), 8.73 – 8.12 (m, 2H), 6.55 (s, 2H), 5.44 (s, 2H), 3.77 (s, 6H), 3.65 (s, 3H). ^{13}C NMR (101 MHz, DMSO) δ 161.53, 155.86, 153.38, 130.67, 102.91, 60.06, 56.01.

Synthesis of 1-(6-chloropyridin-3-yl)-3-(dimethylamino)prop-2-en-1-one (**6**)

1-(6-chloropyridin-3-yl)ethan-1-one **5** (1.00 g, 6.43 mmol) was dissolved in toluene (20 mL) and dimethylformamide diethylacetal (1.65 mL, 9.64 mmol) was added. The mixture was stirred for 16 h at 80 °C. TLC showed near complete conversion. The mixture was cooled to 50 °C and slowly concentrated at reduced pressure. The resulting crude residue was dissolved in warm toluene and pentane was slowly added. A yellow precipitate formed which was filtered off and rinsed with pentane. After drying this yielded the title compound (1.08 g, 5.12 mmol, 80%) as a yellow solid. ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.85 (dd, *J* = 0.7, 2.5 Hz, 1H), 8.17 (dd, *J* = 2.5, 8.3 Hz, 1H), 7.86 (d, *J* = 12.2 Hz, 1H), 7.38 (dd, *J* = 0.8, 8.2 Hz, 1H), 5.63 (d, *J* = 12.1 Hz, 1H), 3.20 (s, 3H), 2.97 (s, 3H). ^{13}C NMR (101 MHz, $CDCl_3$) δ 184.90, 155.07, 153.38, 149.08, 138.10, 134.63, 124.10, 91.54, 45.46, 37.60.

Synthesis of 4-(6-chloropyridin-3-yl)-*N*-(3,4,5-trimethoxyphenyl)pyrimidin-2-amine (**1**)

To a solution of **4** (376 mg, 1.30 mmol) and **6** (250 mg, 1.19 mmol) in ethanol (20 mL) K_2CO_3 (492 mg, 3.56 mmol) was added. The resulting mixture was refluxed for 20 h. The reaction was quenched by the addition of saturated aqueous Na_2CO_3 (15 mL) and the water layer was extracted three times with EtOAc (25 mL). The combined organic layers were washed with brine, dried over $MgSO_4$ and concentrated *in vacuo*. The resulting crude was purified using silica column chromatography (50-70% EtOAc/pentane). The resulting yellow solid was crystallised from warm diethylether to yield the title compound (137 mg, 0.37 mmol, 31%) as light yellow crystals. ^{1}H NMR (400 MHz, $CDCl_3$) δ 9.07 (dd, *J* = 0.7, 2.5 Hz, 1H), 8.52 (d, *J* = 5.2 Hz, 1H), 8.32 (dd, *J* = 2.5, 8.3 Hz, 1H), 7.46 (dd, *J* = 0.7, 8.4 Hz, 1H), 7.38 (s, 1H), 7.14 (d, *J* = 5.2 Hz, 1H), 6.99 (s, 2H), 3.90 (s, 6H), 3.85 (s, 3H). ^{13}C NMR (101 MHz, $CDCl_3$) δ 161.53, 160.28, 159.11, 153.74, 153.47, 148.62, 137.11, 135.30, 131.78, 124.52, 108.15, 97.42, 61.17, 56.26.



Scheme S2: Synthesis of screening hit SG295. Reagents and conditions: *i*) 4-aminophenol, NaOH, DMSO, 100 °C, 65%; *ii*) triphosgene, DCM, 40 °C; *iii*) 1,4-dioxane, 110 °C, 44% over two steps.

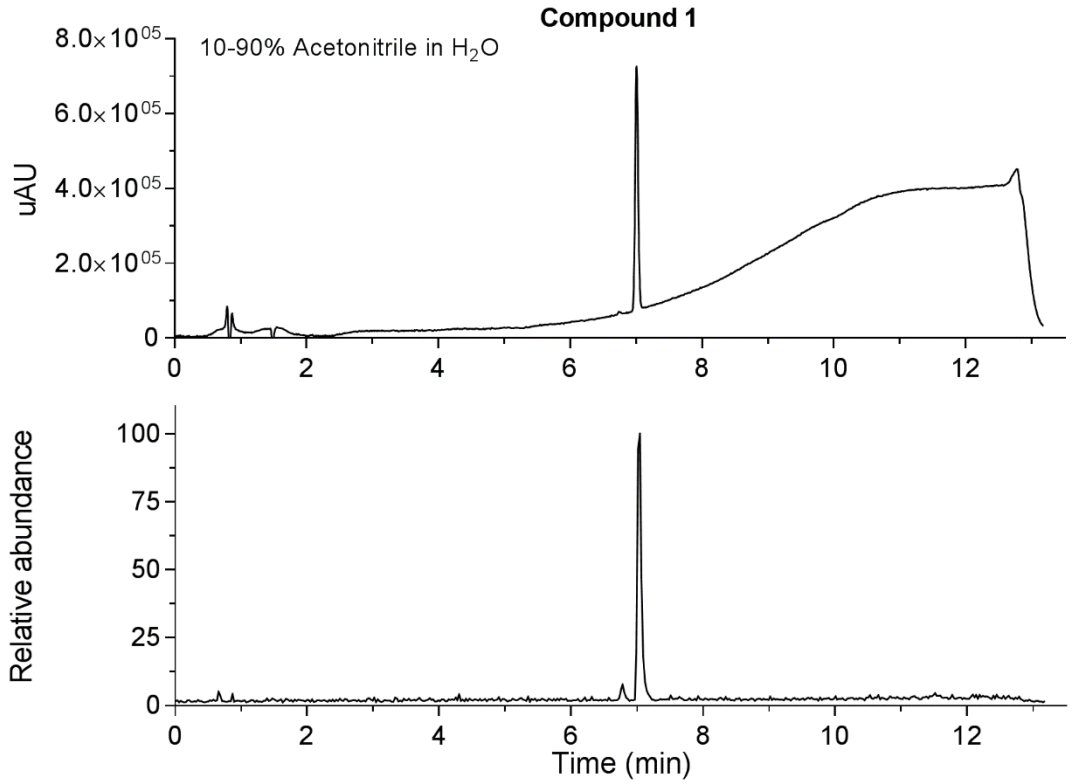
Synthesis of 4-(pyridin-4-yloxy)aniline (**8**)

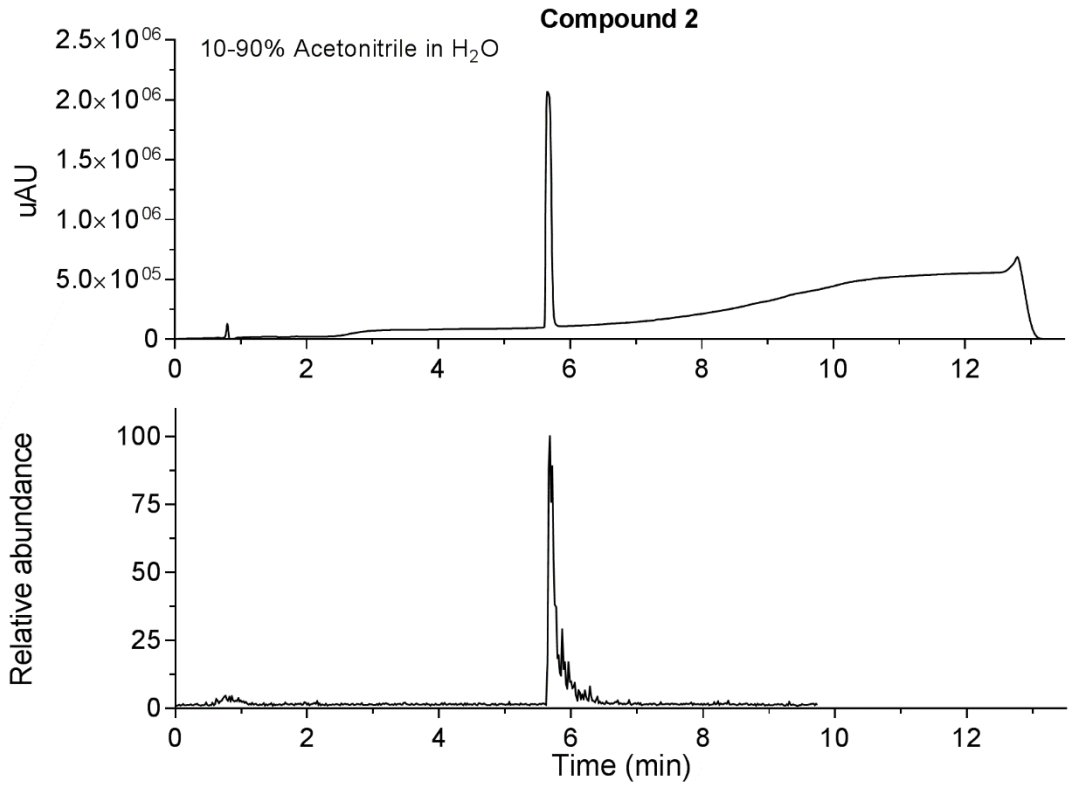
4-aminophenol (2.18 g, 20 mmol), 4-chloropyridine **7** (3.15 g, 21 mmol) and NaOH (2.0 g, 50 mmol) were dissolved in DMSO (50 mL) and heated to 100 °C for 17 h. The mixture was cooled to RT and poured into ice water (300 mL). This was extracted four times with 10% MeOH in chloroform (150 mL). The combined organic layers were washed twice with brine (150 mL), dried over Na_2SO_4 and concentrated *in vacuo*. The residue was flushed over a silica pad with pure EtOAc and concentrated to yield the title compound as a white solid (2.5 g, 13 mmol, 65%). ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.51 – 8.33 (m, 2H), 6.97 – 6.85 (m, 2H), 6.85 – 6.77 (m, 2H), 6.77 – 6.62 (m, 2H), 3.50 (s, 2H). ^{13}C NMR (101 MHz, $CDCl_3$) δ 165.89, 151.25, 145.80, 144.25, 122.12, 116.36, 111.70.

Synthesis of 1-(5-(*tert*-butyl)isoxazol-3-yl)-3-(4-(pyridin-4-yloxy)phenyl)urea (**2**)

3-amino-5-tert-butylisooxazole **9** (0.20 g, 1.4 mmol) was dissolved in DCM (14 mL) and DIPEA (0.50 mL, 2.9 mmol) was added. The mixture was cooled to 0 °C and triphosgene (0.42 g, 1.4 mmol) was added. After stirring the mixture for 19 h at RT it was refluxed for 1 h before the reaction was quenched by the addition of saturated aqueous $NaHCO_3$ (10 mL). The water layer was separated and extracted three times with DCM (10 mL). The combined organic layers were washed with brine and dried over Na_2SO_4. The solvent was removed under reduced pressure and the crude isocyanate was dissolved in 1,4-dioxane (14 mL). **8** (0.29 g, 1.6 mmol) was added to the solution and the mixture was heated to 110 °C for 2.5 h. After cooling to RT the mixture was diluted with DCM (30 mL) and saturated aqueous $NaHCO_3$ (30 mL) was added. The water layer was separated and extracted three times with DCM (30 mL). The combined organic layers were washed with brine, dried over Na_2SO_4 and concentrated *in vacuo*. The resulting yellow oil was purified by preparative HPLC to yield the title compound as a slightly yellow oil (0.22 g, 0.62 mmol, 44%). ^{1}H NMR (400 MHz, $CDCl_3$) δ 9.41 (s, 1H), 8.67 (s, 1H), 8.47 (dd, *J* = 1.5, 5.1 Hz, 2H), 7.63 – 7.54 (m, 2H), 7.10 – 7.03 (m, 2H), 6.89 – 6.81 (m, 2H), 5.95 (s, 1H), 1.36 (s, 9H). ^{13}C NMR (101 MHz, $CDCl_3$) δ 181.60, 165.38, 158.45, 152.58, 151.22, 149.95, 135.58, 122.15, 121.60, 112.17, 91.94, 33.08, 28.75.

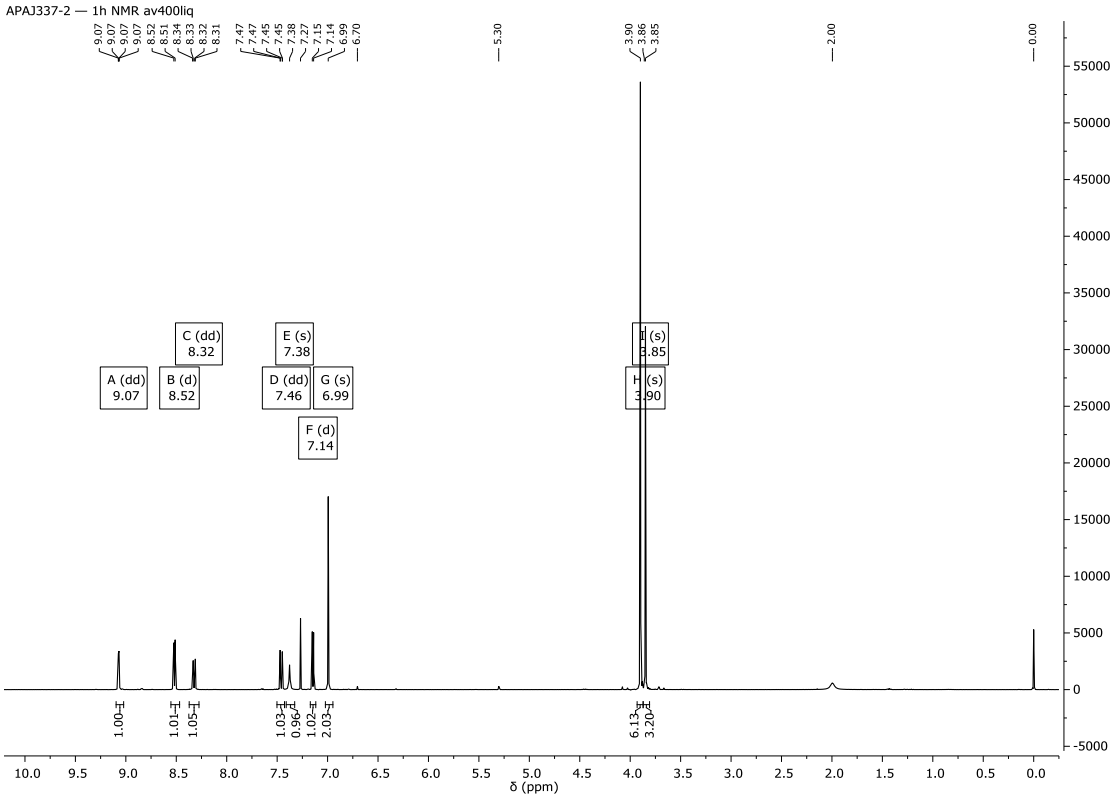
LCMS traces of *in situ* tested compounds

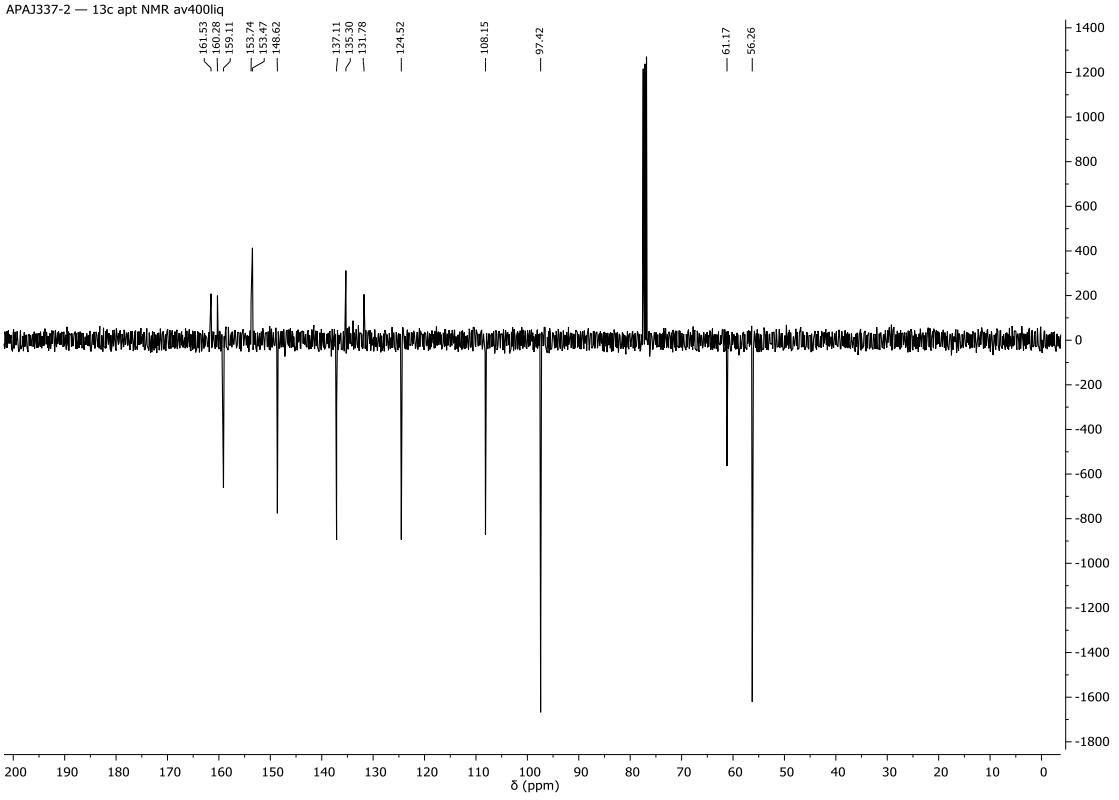




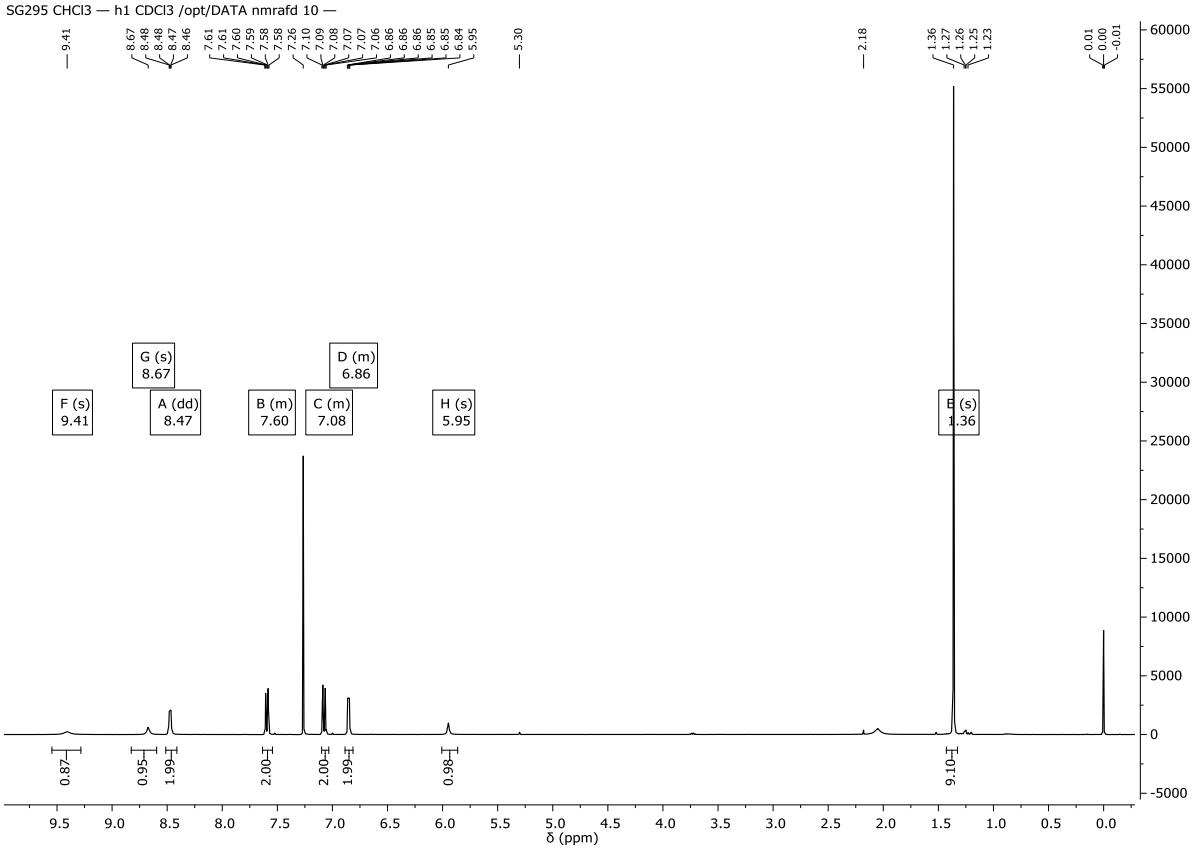
NMR spectra of *in situ* tested compounds

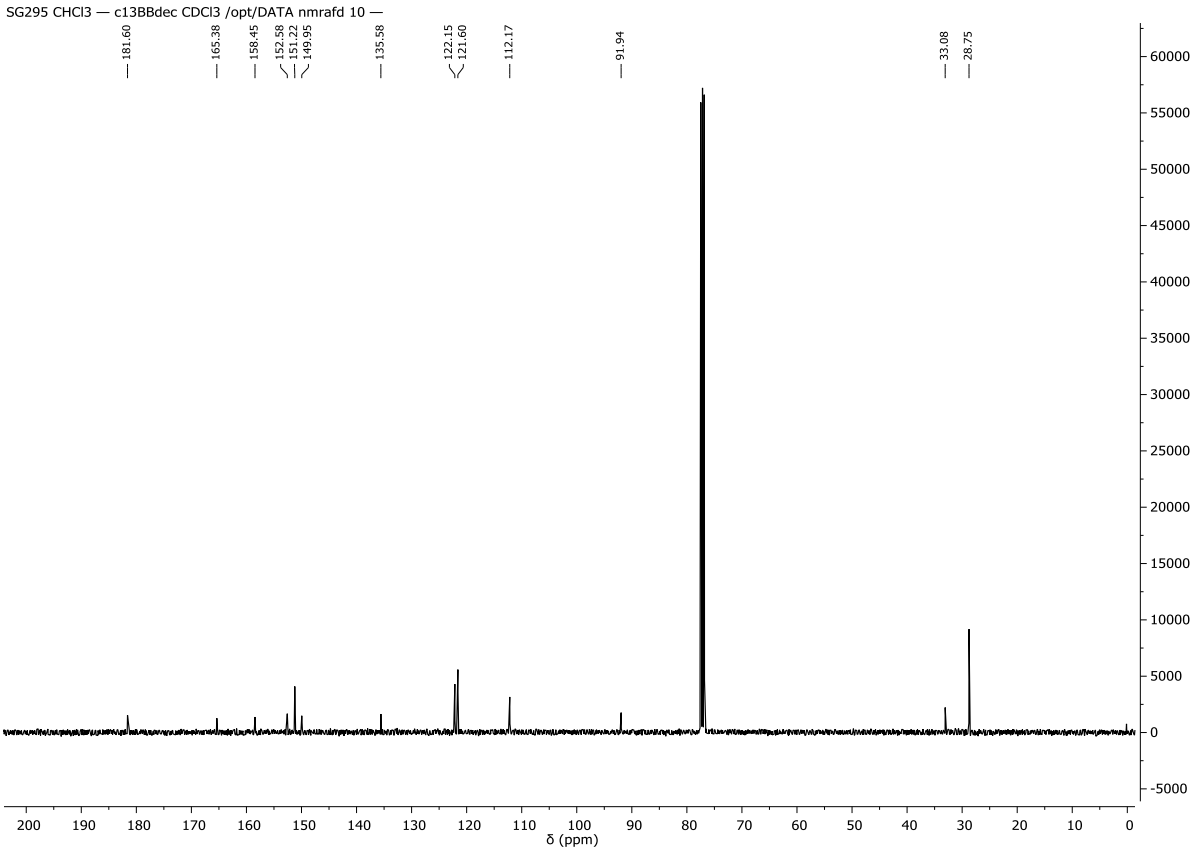
Compound 1





Compound 2


SG295 CHCl3 — h1 CDCl3 /opt/DATA nmrafd 10 —
F (s) 9.41
G (s) 8.67
A (dd) 8.47
B (m) 7.60
C (m) 7.08
D (m) 6.86
H (s) 5.95
E (s) 1.36
δ (ppm)


SG295 CHCl3 — c13BBdec CDCl3 /opt/DATA nmrafd 10 —
δ (ppm)

References

(1) Karaman, M. W.; Herrgard, S.; Treiber, D. K.; Gallant, P.; Atteridge, C. E.; Campbell, B. T.; Chan, K. W.; Ciceri, P.; Davis, M. I.; Edeen, P. T.; Faraoni, R.; Floyd, M.; Hunt, J. P.; Lockhart, D. J.; Milanov, Z. V; Morrison, M. J.; Pallares, G.; Patel, H. K.; Pritchard, S.; Wodicka, L. M.; Zarrinkar, P. P. A Quantitative Analysis of Kinase Inhibitor Selectivity. *Nat. Biotechnol.* **2008**, *26*, 127–132.

(2) Metz, J. T.; Johnson, E. F.; Soni, N. B.; Merta, P. J.; Kifle, L.; Hajduk, P. J. Navigating the Kinome. *Nat. Chem. Biol.* **2011**, *7*, 200–202.

(3) Anastassiadis, T.; Deacon, S. W.; Devarajan, K.; Ma, H.; Peterson, J. R. Comprehensive Assay of Kinase Catalytic Activity Reveals Features of Kinase Inhibitor Selectivity. *Nat. Biotechnol.* **2011**, *29*, 1039–1045.

(4) Davis, M. I.; Hunt, J. P.; Herrgard, S.; Ciceri, P.; Wodicka, L. M.; Pallares, G.; Hocker, M.; Treiber, D. K.; Zarrinkar, P. P. Comprehensive Analysis of Kinase Inhibitor Selectivity. *Nat. Biotechnol.* **2011**, *29*, 1046–1051.

(5) Elkins, J. M.; Fedele, V.; Szklarz, M.; Abdul Azeez, K. R.; Salah, E.; Mikolajczyk, J.; Romanov, S.; Sepetov, N.; Huang, X.-P.; Roth, B. L.; Al Haj Zen, A.; Fourches, D.; Muratov, E.; Tropsha, A.; Morris, J.; Teicher, B. A.; Kunkel, M.; Polley, E.; Lackey, K. E.; Atkinson, F. L.; Overington, J. P.; Bamborough, P.; Müller, S.; Price, D. J.; Willson, T. M.; Drewry, D. H.; Knapp, S.; Zuercher, W. J. Comprehensive Characterization of the Published Kinase Inhibitor Set. *Nat. Biotechnol.* **2015**, *34*, 95–103.

(6) Drewry, D. H.; Willson, T. M.; Zuercher, W. J. Seeding Collaborations to Advance Kinase Science with the GSK Published Kinase Inhibitor Set (PKIS). *Curr. Top. Med. Chem.* **2014**, *14*, 340–342.

(7) EMBL-EBI. *CHEMBL Database Release 23*; 2017.

(8) Landrum, G. RDKit: Open-Source Cheminformatics; Http://Www.Rdkit.Org.

(9) Berthold, M. R.; Cebron, N.; Dill, F.; Gabriel, T. R.; Kötter, T.; Meinl, T.; Ohl, P.; Sieb, C.; Thiel, K.; Wiswedel, B. KNIME: The Konstanz Information Miner. In *Data Analysis, Machine Learning and Applications*; Preisach C., Burkhardt H., Schmidt-Thieme L., D. R., Ed.; Springer, Berlin, Heidelberg, 2008; pp 319–326.

(10) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, É. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

(11) Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise.

(12) Heil, B.; Ludwig, J.; Lichtenberg-Frate, H.; Lengauer, T. Computational Recognition of Potassium Channel Sequences. *Bioinformatics* **2006**, *22*, 1562–1568.

(13) Christmann-Franck, S.; van Westen, G. J. P.; Papadatos, G.; Beltran Escudie, F.; Roberts, A.; Overington, J. P.; Domine, D. Unprecedently Large-Scale Kinase Inhibitor Set Enabling the Accurate Prediction of Compound–Kinase Activities: A Way toward Selective Promiscuity by Design? *J. Chem. Inf. Model.* **2016**, *56*, 1654–1675.

(14) Lenselink, E. B.; ten Dijke, N.; Bongers, B.; Papadatos, G.; van Vlijmen, H. W. T.; Kowalczyk,

W.; IJzerman, A. P.; van Westen, G. J. P. Beyond the Hype: Deep Neural Networks Outperform Established Methods Using a ChEMBL Bioactivity Benchmark Set. *J. Cheminform.* **2017**, *9*, 45.

(15) Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; Overington, J. P. ChEMBL: A Large-Scale Bioactivity Database for Drug Discovery. *Nucleic Acids Res.* **2012**, *40*, D1100–D1107.

(16) Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S. The Protein Kinase Complement of the Human Genome. *Science (80-.).* **2002**, *298*, 1912–1934.

(17) Consortium, T. U. UniProt: The Universal Protein Knowledgebase. *Nucleic Acids Res.* **2017**, *45*, D158–D169.

(18) Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; Thompson, J. D.; Higgins, D. G. Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7*, 539.

(19) Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32.

(20) van Westen, G. J.; Swier, R. F.; Cortes-Ciriano, I.; Wegner, J. K.; Overington, J. P.; IJzerman, A. P.; van Vlijmen, H. W.; Bender, A. Benchmarking of Protein Descriptor Sets in Proteochemometric Modeling (Part 2): Modeling Performance of 13 Amino Acid Descriptor Sets. *J. Cheminform.* **2013**, *5*, 42.

(21) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742–754.

(22) Schrödinger Suite: Maestro. Schrödinger LLC: New York, NY 2017.

(23) Ke, Y.-Y.; Singh, V. K.; Coumar, M. S.; Hsu, Y. C.; Wang, W.-C.; Song, J.-S.; Chen, C.-H.; Lin, W.-H.; Wu, S.-H.; Hsu, J. T. A.; Shih, C.; Hsieh, H.-P. Homology Modeling of DFG-in FMS-like Tyrosine Kinase 3 (FLT3) and Structure-Based Virtual Screening for Inhibitor Identification. *Sci. Rep.* **2015**, *5*, 11702.

(24) Jacobson, M. P.; Pincus, D. L.; Rapp, C. S.; Day, T. J. F.; Honig, B.; Shaw, D. E.; Friesner, R. A. A Hierarchical Approach to All-Atom Protein Loop Prediction. *Proteins Struct. Funct. Bioinforma.* **2004**, *55*, 351–367.

(25) Jacobson, M. P.; Friesner, R. A.; Xiang, Z.; Honig, B. On the Role of the Crystal Environment in Determining Protein Side-Chain Conformations. *J. Mol. Biol.* **2002**, *320*, 597–608.

(26) Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R. Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects. *J. Med. Chem.* **2006**, *49*, 534–553.

(27) Smith, C. C.; Zhang, C.; Lin, K. C.; Lasater, E. A.; Zhang, Y.; Massi, E.; Damon, L. E.; Pendleton, M.; Bashir, A.; Sebra, R.; Perl, A.; Kasarskis, A.; Shellooe, R.; Tsang, G.; Carias, H.; Powell, B.; Burton, E. A.; Matusow, B.; Zhang, J.; Spevak, W.; Ibrahim, P. N.; Le, M. H.; Hsu, H. H.; Habets, G.; West, B. L.; Bollag, G.; Shah, N. P. Characterizing and Overriding the Structural Mechanism of the Quizartinib-Resistant FLT3 "Gatekeeper" F691L Mutation with PLX3397. *Cancer Discov.* **2015**, *5*, 668–679.

(28) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; 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.