MedChemComm

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/medchemcomm

MedChemComm

CONCISE ARTICLE

Property-based characterization of kinase-like ligand space for library design and virtual screening

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/medchemcomm

Dávid Bajusz, György G. Ferenczy and György M. Keserű*

A property-based desirability scoring scheme has been developed for kinase-focused library design and ligand-based pre-screening of large compound sets. Property distributions of known kinase inhibitors from the ChEMBL Kinase Sarfari database were investigated and used for a desirability function-based score. The scoring scheme is easily interpretable as it accounts for six molecular properties: topological polar surface area and the number of rotatable bonds, hydrogen bond donors, aromatic rings, nitrogen and oxygen atoms. The performance of the Kinase Desirability Score (KiDS) is evaluated on both public and proprietary experimental screening data.

Introduction

Phosphorylation is a ubiquitous signaling and regulating mechanism in living organisms. Kinases are enzymes that carry out the phosphorylation of mostly other proteins or other types of substrates. They function by transferring a phosphate group from a bound ATP-molecule to a Ser/Thr/Tyr residue on the substrate(s). There are more than 500 protein kinases encoded in the human genome[1], accounting for a total of 2% of all human genes.[2] Abnormalities in protein phosphorylation are precursors to a variety of malignancies ranging from cancer to autoimmune diseases: for many of them, small-molecule inhibition of the involved protein kinase has been shown to be an effective therapy. Consequently, protein kinases currently constitute the second most exploited drug target class after GPCRs.[3]

Since kinases have one well-defined function and share their endogenous ligand (ATP), their ATP-binding pockets are very well-conserved across the whole kinome. Thus, medicinal chemists face a great challenge in designing kinase inhibitors with sufficient selectivity towards the given target to avoid unwanted side effects. Even though the field has seen the advent of type II inhibitors in the 2000's[4], the majority of reported kinase inhibitors are still type I ligands. (Type II inhibitors bind to the inactive or "DFG-out" conformation of kinases as opposed to type I inhibitors that bind to the ATP-binding pocket in an active or "DFG-in" conformation.) Moreover, as our understanding of the mechanism of action of type II inhibitors improves, it is becoming clearer that this class of compounds is not inherently more selective than ATP-site inhibitors.[5] Thus, the predominant approach towards kinase inhibitor design is still the small-molecule targeting of the ATP-site, even more so as the majority of available structural and biochemical data refer to type I inhibitors.

Virtual screening has been proven to be a useful approach in the hit discovery of kinase targets.[6,7] However, due to the significant increase of the commercially and/or synthetically available druglike (and leadlike) chemical space, structure-based screening methods are facing capacity challenges. As a solution, less accurate but quicker filters can be applied prior to the actual virtual screening (*e.g.* docking) to derive a more focused dataset of manageable size.

Various approaches have been applied previously to assemble kinase-focused compound libraries (virtual and physical as well), including substructure-based methods[8–13] and similarity-based methods.[14–16] Most recently, Singh and coworkers explored the possibility of characterizing kinase-like ligands based on physicochemical descriptors.[17] With the increasing amount of publicly available inhibitor activity data[18], this approach becomes an attractive opportunity, since substructure- and similarity-based methods inherently retrieve molecules that are structurally similar to the reference compound(s), limiting the ability to identify inhibitors with novel scaffolds. In contrast, property-based methods do not have this limitation. The Kinase-Like Score (KLS) introduced by Singh and coworkers characterizes kinase-like ligand space on a statistical basis: it considers nine descriptors and scores them according to a formula that assumes normal distribution.

A suitable MPO (multi-parameter optimization[19]) method for compound profile optimization is the desirability function.[20,21] The essence of the underlying concept is that for each descriptor, a tailor-made scoring function is introduced, which reflects the "desirability" of the various possible values of that descriptor (*e.g.* how prevalent that descriptor value is among reference compounds). Desirability functions usually take values between 0 and 1, and generally either a sum or a product of the individual scores is calculated at the end of the

Medicinal Chemistry Research Group, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Magyar tudósok körútja 2, Budapest 1117, Hungary. E-mail: keseru.gyorgy@ttk.mta.hu

† Electronic Supplementary Information (ESI) available: Desirability functions of the descriptors applied in KiDS, Enrichment factors for the performance evaluation of KiDS, Results of the external validation of KiDS. See DOI: 10.1039/x0xx00000x

process to produce the overall desirability score. Recent examples of studies that involve desirability function-based optimizations include Cruz-Monteagudo and coworkers' paper on global QSAR studies[22], Avram and coworkers' article on the characterization of pesticide-like compounds[23], and a GPCR-focused library design implementation by our group.[24]

In this paper, we present a desirability function-based scoring scheme (Kinase Desirability Score or KiDS) using topological descriptors to screen kinase-like ligands. Based on this study, KiDS can be applied as a pre-filter for kinase-like ligands in virtual screening campaigns or alternatively it might support designing kinase-focused libraries.

Methods

We have developed and tested a desirability function-based scoring scheme (KiDS) for the quick and computationally efficient filtering of large compound collections. The study mostly involved enrichment tests on datasets, where known kinase inhibitors were mixed into a larger set of random molecules from a commercial compound database. Thorough external validation was also carried out on publicly available (Pubchem Bioassay) and proprietary (Gedeon Richter Plc.) HTS datasets. We have also examined the correlation between KiDS score and kinase promiscuity using full matrix data from the EMD Millipore Kinase Screening dataset in ChEMBL.[25] The following sections cover the applied computational methods in more detail.

Database retrieval

Structures and activity data of known kinase inhibitors (used as actives) were retrieved from the ChEMBL Kinase Sarfari database (version 17).[26] Duplicate entries were removed and the largest fragment was kept for each entry. Only those molecules with a corresponding activity measurement of type B ("Binding", such as IC_{50} or K_i in an enzyme-based assay) were kept and activity values were converted to IC_{50} where K_i or K_d were provided. For molecules with multiple activities, the lowest IC_{50} values were kept. Actives were defined as molecules that exhibit an $IC_{50} \leq 10$ μM value on at least one kinase. The Mcule Purchasable Compounds Database was utilized as the source of random molecules (identified as non-actives)[27], which were filtered to exclude any known kinase actives present in the ChEMBL Kinase Sarfari or the Pubchem test set (see below). To reduce the effect of molecular size, the input databases were focused to leadlike compounds, as defined by Teague and co-workers ($250 \leq MW \leq 350$, $logP \leq 3.5$, $rotB \leq 7$) .[28] Several datasets were compiled, where actives and non-actives were mixed in an approximately 1:9 ratio. The Training set contained 2500 actives from ChEMBL and 22803 non-actives from Mcule, while Test set 1 counted 1923 actives (ChEMBL) and 18000 non-actives (Mcule), and Test set 2 counted 730 actives (PubChem[29]) and 6300 non-actives (Mcule). Both test sets were used for external validation. An additional effort for external validation involved the exchange of the random molecules: in Test sets Z, 1Z and 2Z, the non-actives from Mcule were exchanged to 20000 randomly selected leadlike compounds from ZINC[30,31] (while the kinase actives were the same as in the Training set and Test sets 1 and 2, respectively). The open source cheminformatics platform KNIME (version 2.9.1) was used for all dataset operations.[32] Removal of counter ions and the calculation of molecular descriptors were carried out with the KNIME implementation of JChem software (version 6.3.0), using Standardizer and Calculator Plugins.[33] The KNIME workflow for the calculation of KiDS is available on our website: http://medchem.ttk.mta.hu. A quick visual reference for the calculation of KiDS is provided in Figure 1.

Desirability functions

Scoring (classifier) variables were selected from a pool of commonly used molecular descriptors: molecular weight, logP, TPSA, pK_a of the most acidic and basic centers and the numbers of hydrogen bond acceptors and donors, heavy atoms, rings, rotatable bonds, nitrogen and oxygen atoms, aromatic, aliphatic and fused rings. (For the actual descriptors that finally constituted the Kinase Desirability Score, see the "Results" section.) As a first measure of inspection, Mann-Whitney U tests were carried out to establish whether the differences in the medians of the descriptors are statistically significant. The descriptors were tested for 2500 active and 2500 non-active molecules from the Training set and the results were significant at the $p = 0.05$ level (in fact, p values approximated 0). Since there is a known trend for statistical tests to be more sensitive as the size of the sample increases, we have inspected the distributions visually as well (on categorized histograms) and preferred those descriptors for which substantial differences were detected. For statistical testing and histogram plotting, STATISTICA 12 was applied.[34]

Desirability functions as introduced by Harrington[20] and Derringer[21] were defined for a number of molecular descriptors as custom-made functions that assign a value between 0 and 1 (desirability score) to each possible descriptor value. Generally, the assigned desirability scores were higher as the prevalence of the given descriptor value was higher among actives and lower among non-actives. (For a more detailed description, see the "Results" section and Figures S1-S6.) The additive approach was taken to summarize the separate desirabilities based on the descriptors, *i.e.* the overall Kinase Desirability Score was defined as the sum of the desirability scores obtained for the descriptors independently.

Evaluation

To assess the performance of the scoring scheme, enrichment studies were carried out on the Training set and on the two independent Test sets. Enrichment factors (EF) at 0.5%, 1%, 2% and 5%, receiver operating characteristic curves (ROC) and area under the ROC curve (AUC) values were calculated to evaluate the results. Enrichment factors were defined as suggested by Jain and Nicholls[35] to provide a size-independent measure of early enrichment:

$$EF_{x\%} = (TPR_{x\%}) / x\%, \quad (1)$$

i.e. the enrichment factor is equal to the ratio of the true positive rate and the false positive rate for a given false

positive rate x% (in other words, Y / X for a specific point on the ROC curve). Conventional enrichment factors, defined as:

$$EF_{x\%} = (N_{act,\ x\%} / N_{x\%}) / (N_{act} / N) \quad (2)$$

were also calculated and included in the Supplementary information. (Here, $N_{act,\ x\%}$ and $N_{x\%}$ are the number of actives and the total number of compounds in the top x% of the ranked list (respectively), while N_{act} and N are the number of actives and the total number of compounds in the whole dataset, respectively.)

The ROC curve is the plot of %(true positives) vs. %(false positives) for the ranked list of objects (here, molecules). The straight diagonal line is a reference that corresponds to random classification. AUC is the area under the ROC curve which is calculated numerically. 95% confidence intervals are reported for both the AUC values and the enrichment factors as elaborated by Nicholls.[36]

Results

Development of the scoring scheme

Six descriptors were chosen to be included in the Kinase Desirability Score: topological polar surface area (TPSA) and the number of rotatable bonds (rotB), nitrogen atoms (N_N), oxygen atoms (N_O), aromatic rings (Arom) and hydrogen bond donors (HBD). For discrete descriptors (all of the above except for TPSA), desirability scores are assigned based on a simple decision matrix presented in Table 1. The score for a given property value is assigned based on robust statistical parameters (the median and the interquartile range) of that property among kinase actives and random molecules. (For example, if a property value for a compound is inside the interquartile range of that property for kinase actives, but outside of the interquartile range for random molecules, the desirability score assigned to that property is 1.) For the TPSA, the score continuously increases from 0 to 1 between the median TPSA's of random molecules and kinase actives, and decreases to 0 as it approaches the top of the upper quartile for kinase actives. (See Supplementary Figure S1). The graphical representations of the desirability functions are reported in Supplementary Figures S1-S6, while the definitions of the functions are reported in Supplementary Table S1.

From the distributions of these descriptors among kinase-like and random molecules, the following general observations can be drawn: among kinase-like compounds, less oxygen atoms and rotatable bonds, higher polar surface area, and more aromatic rings, nitrogen atoms and hydrogen bond donors are preferred than what can be observed for random compounds. These differences are reflected in the definitions of the desirability functions of KiDS.

Evaluation of the scoring scheme

Performance on the Training and Test sets. The ROC curves presented in Figure 2 display high AUC values, together with a steep initial curve that corresponds to good early enrichments (see Table 2). Early enrichments are especially important when a small portion of the top scoring functions are sought while the general character of the ROC curve and the good AUC value are substantial when a larger part of the screened dataset is selected for subsequent studies. The results suggest the applicability of KiDS for both scenarios. (Conventional enrichment factors are reported in Supplementary Table S2, while categorized histograms of the KiDS distributions are presented in Supplementary Figures S7-S9.)

External validation has been carried out on Test sets 1 and 2, and clearly the deterioration of the results (with respect to the training set) is negligible, confirming the robustness of the scoring method. An additional external validation was carried out to verify the robustness of the Kinase Desirability Score: the random compounds from Mcule (in the Training and both Test sets) were exchanged to a set of 20000 random leadlike compounds from ZINC to assess whether the scoring method is dependent on the starting dataset. (Figure 2B) Deterioration of the performance parameters was negligible, suggesting that the performance of KiDS does not depend significantly on the source of the examined database. (Enrichment factors and AUC values are reported in Supplementary Tables S3 and S4.) The active:non-active ratio on the other hand influences this performance as shown in the next section.

KiDS also outperforms the Kinase-Like Score (KLS) of Singh *et al.*[17] (presented on Figure 2 as a reference), justifying its use for the mentioned purposes. An explanation for the improved performance of KiDS relative to the Kinase-Like Score (KLS)[17] is that while KLS accounts only for the property distributions of kinase actives, KiDS considers the differences between kinase actives and random, commercially available compounds. The same can be specified as the reason for KLS being sensitive to the source of random compounds, while KiDS is not (Figure 2 B). In this context, it is worth noting that the ability to distinguish and characterize different compound databases was a key requirement during the development of KLS. While the primary purpose of KLS was to examine compound databases, KiDS was developed with the intention to provide a general tool for property-based pre-screening for structure-based virtual screens and as such, it provides a better alternative for this task than KLS.

Performance on screening datasets. As an additional measure to validate the Kinase Desirability Score, one proprietary (Gedeon Richter) and three publicly available (Pubchem Bioassay) HTS datasets were subjected to scoring and evaluation with KiDS (and also with KLS as a reference). With this calculation, we assess whether the application of KiDS as a pre-filtering step increases the chance of finding active molecules in a smaller portion of the entire HTS set (thus reducing the effective cost of finding an active molecule). Since KiDS was developed for the pre-screening of leadlike molecules, the HTS sets were first focused to this size range.[28] Table 3 summarizes the composition of these (pre-filtered) HTS sets, as well as the AUC values of their evaluation with KiDS and KLS.[17] ROC curves of the evaluations are presented in Figure 3. (Due to the very small number of confirmed actives, enrichment factors are not reported for these datasets.)

It is apparent from the results that the scoring of the screening datasets with KiDS is effective in selecting a subset enriched in kinase ligands. For example, the experimental testing of the top half of the HTS set published as AID 604 in Pubchem Bioassay (Figure 3C) would result in identifying 80% of the actives that are found during the testing of the whole dataset. Similar result is obtained for AID 524 while KiDS gave somewhat inferior results for the Gedeon Richter's HTS (60% confirmed actives in the top scored 50%) and performed better for AID 619 where over 90% of actives are identified in the top scored 50% set. (Clearly, the performance is worse than for the training and test sets presented earlier, but that can be attributed to the much lower active:inactive ratios of the Pubchem Bioassay HTS sets.) Moreover, KiDS proved to be superior to KLS in each case. These results support that the application of KiDS as a pre-filtering step can reduce the effective cost of finding active molecules in a kinase-directed high-throughput screening.

KiDS and kinase promiscuity. To examine the relationship between KiDS and the likeliness of activity on a kinase, we have calculated the KiDS scores for the EMD Millipore Kinase Screening dataset in ChEMBL.[25] The dataset contains 158 well-known kinase inhibitors, out of which 40 are leadlike.[28] Promiscuity was defined as the number of kinase targets on which a compound is active. (Actives were defined as those compounds that exhibit ≤ 50% residual activity in the screening concentration of 1 or 10 μM.) It is important to stress that the purpose of KiDS is not the selection of promiscuous compounds: correlating KiDS to the promiscuity of well characterized compounds only serves as a tool here to assess the likeliness of a given compound to be active on an arbitrary kinase. On Figure 4, a significant linear correlation can be observed between KiDS and the average number of kinases hit (with a correlation coefficient of $R^2 = 0.838$). In other words, a higher KiDS score does confer a higher chance of finding the given compound to be active on an arbitrary kinase of interest.

Conclusions

Virtual screening of large chemical databases is one of the most powerful strategies generating viable chemical starting points for kinase inhibitor discovery programs.[6,7] Structure-based methods, however, are increasingly demanding computationally as the size of the screened database increases. Although substructure- and similarity-based screening methods are faster, they are less likely to identify structurally novel hit compounds, and thus they are less suited to expand the chemical space of kinase inhibitors. To get around this problem, we identified property-based pre-screening as a useful step prior to structure-based approaches. In this study we introduced a molecular property-based scoring scheme, the Kinase Desirability Score (KiDS). The scoring scheme involves custom desirability functions based on six molecular descriptors: topological polar surface area (TPSA) and the number of rotatable bonds (rotB), nitrogen atoms (N_N), oxygen atoms (N_O), aromatic rings (Arom) and hydrogen bond donors (HBD). Scores between 0 and 1 are assigned to each of the descriptors and summed up to give the Kinase Desirability Score. KiDS is flexible in the sense that it does not impose very strict constraints regarding either of the involved molecular properties. Therefore, it allows for the identification of structurally novel kinase inhibitors.

KiDS was developed and tested using a dataset of known kinase inhibitors (ChEMBL) and random compounds from a commercial compound databases (Mcule and ZINC), and its performance was assessed with early enrichment factors, ROC curves and AUC values on Training and independent Test sets. External validation also involved testing its performance on proprietary and public HTS datasets as well as full matrix screening data. In the latter case, a significant correlation between the KiDS score and kinase promiscuity could be observed.

The good and consistent performance parameters suggest that KiDS is useful as a pre-screening step in virtual screening workflows and for kinase-focused library design, as well. It also presents a more efficient alternative for these tasks than the previously suggested Kinase-Like Score (KLS). In HTS campaigns, a KiDS-based pre-screening can reduce the effective cost of finding hit compounds.

Abbreviations

AUC, area under the (ROC) curve; EF, enrichment factor; GPCR, G-protein coupled receptor; HBD, number of hydrogen bond donors; IC_{50}, half maximal inhibitory concentration; IQR, inter-quartile range; KiDS, Kinase Desirability Score; logP, logarithm of the n-octanol/water partition coefficient; MPO, multi-parameter optimization; QSAR, quantitative structure-activity relationship; ROC, receiver operating characteristic; rotB, rotatable bond count; TPSA, topological polar surface area

Acknowledgements

The authors thank Gedeon Richter Plc (and in particular, Ákos Tarcsay) for providing the proprietary HTS dataset for validation purposes, and Károly Héberger for his valuable suggestions and explanations regarding statistical tests. This work was supported by the Hungarian Scientific Research Found (Grant No. K 116904).

Notes and references

1 G. Manning, D. B. Whyte, R. Martinez, T. Hunter and S. Sudarsanam, *Science*, 2002, **298**, 1912–34.

2 G. M. Rubin, M. D. Yandell, J. R. Wortman, G. L. Gabor Miklos, C. R. Nelson, I. K. Hariharan, M. E. Fortini, P. W. Li, R. Apweiler, W. Fleischmann, J. M. Cherry, S. Henikoff, M. P. Skupski, S. Misra, M. Ashburner, E. Birney, M. S. Boguski, T. Brody, P. Brokstein, S. E. Celniker, S. A. Chervitz, D. Coates, A. Cravchik, A. Gabrielian, R. F. Galle, W. M. Gelbart, R. A. George, L. S. Goldstein, F. Gong,

P. Guan, N. L. Harris, B. A. Hay, R. A. Hoskins, J. Li, Z. Li, R. O. Hynes, S. J. Jones, P. M. Kuehl, B. Lemaitre, J. T. Littleton, D. K. Morrison, C. Mungall, P. H. O'Farrell, O. K. Pickeral, C. Shue, L. B. Vosshall, J. Zhang, Q. Zhao, X. H. Zheng and S. Lewis, *Science*, 2000, **287**, 2204–15.

3 P. Cohen, *Nat Rev Drug Discov*, 2002, **1**, 309–315.

4 T. Schindler, W. Bornmann, P. Pellicena, W. T. Miller, B. Clarkson and J. Kuriyan, *Science*, 2000, **289**, 1938–1943.

5 Z. Zhao, H. Wu, L. Wang, Y. Liu, S. Knapp, Q. Liu and N. S. Gray, *ACS Chem. Biol.*, 2014, **9**, 1230–41.

6 C. McInnes, *Curr. Opin. Chem. Biol.*, 2007, **11**, 494–502.

7 P. D. Lyne, P. W. Kenny, D. A. Cosgrove, C. Deng, S. Zabludoff, J. J. Wendoloski and S. Ashwell, *J. Med. Chem.*, 2004, **47**, 1962–8.

8 A. M. Aronov, B. McClain, C. S. Moody and M. A. Murcko, *J. Med. Chem.*, 2008, **51**, 1214–22.

9 R. Brenk, A. Schipani, D. James, A. Krasowski, I. H. Gilbert, J. Frearson and P. G. Wyatt, *ChemMedChem*, 2008, **3**, 435–44.

10 G. Kéri, Z. Székelyhidi, P. Bánhegyi, Z. Varga, B. Hegymegi-Barakonyi, C. Szántai-Kis, D. Hafenbradl, B. Klebl, G. Muller, A. Ullrich, D. Erös, Z. Horváth, Z. Greff, J. Marosfalvi, J. Pató, I. Szabadkai, I. Szilágyi, Z. Szegedi, I. Varga, F. Wáczek and L. Orfi, *Assay Drug Dev. Technol.*, 2005, **3**, 543–51.

11 J. F. Lowrie, R. K. Delisle, D. W. Hobbs and D. J. Diller, *Comb. Chem. High Throughput Screen.*, 2004, **7**, 495–510.

12 H. Xi and E. A. Lunney, *Methods Mol. Biol.*, 2011, **685**, 279–91.

13 C. Zhang and G. Bollag, *Curr. Opin. Genet. Dev.*, 2010, **20**, 79–86.

14 F. Deanda, E. L. Stewart, M. J. Reno and D. H. Drewry, *J. Chem. Inf. Model.*, 2008, **48**, 2395–403.

15 H. Decornez, A. Gulyás-Forró, Á. Papp, M. Szabó, G. Sármay, I. Hajdú, S. Cseh, G. Dormán and D. B. Kitchen, *ChemMedChem*, 2009, **4**, 1273–1278.

16 D. Sun, C. Chuaqui, Z. Deng, S. Bowes, D. Chin, J. Singh, P. Cullen, G. Hankins, W.-C. Lee, J. Donnelly, J. Friedman and S. Josiah, *Chem. Biol. Drug Des.*, 2006, **67**, 385–94.

17 N. Singh, H. Sun, S. Chaudhury, M. D. M. AbdulHameed, A. Wallqvist and G. Tawa, *J. Cheminform.*, 2012, **4**, 4.

18. ChEMBL [https://www.ebi.ac.uk/chembl/]

19 M. D Segall, *Curr. Pharm. Des.*, 2012, **18**, 1292–1310.

20 E. C. Harrington, *Ind. Qual. Control*, 1965, **21**, 494–498.

21 G. Derringer and R. Suich, *J. Qual. Technol.*, 1980, **12**, 214–219.

22 M. Cruz-Monteagudo, F. Borges and M. N. D. S. Cordeiro, *J. Comput. Chem.*, 2008, **29**, 2445–59.

23 S. Avram, S. Funar-Timofei, A. Borota, S. R. Chennamaneni, A. K. Manchala and S. Muresan, *J. Cheminform.*, 2014, **6**, 42.

24 A. A. Kelemen, G. G. Ferenczy and G. M. Keserű, *J. Comput. Aided. Mol. Des.*, 2015, **29**, 59–66.

25 Y. Gao, S. P. Davies, M. Augustin, A. Woodward, U. A. Patel, R. Kovelman and K. J. Harvey, *Biochem. J.*, 2013, **451**, 313–28.

26. ChEMBL kinase SARfari [https://www.ebi.ac.uk/chembl/sarfari/kinasesarfari]

27 R. Kiss, M. Sándor and F. A. Szalai, *J. Cheminform.*, 2012, **4 (Suppl 1)**, 17.

28 S. J. Teague, A. M. Davis, P. D. Leeson and T. Oprea, *Angew. Chemie Int. Ed.*, 1999, **38**, 3743–3748.

29. PubChem [https://pubchem.ncbi.nlm.nih.gov/]

30 J. J. Irwin and B. K. Shoichet, *J. Chem. Inf. Model.*, 2005, **45**, 177–82.

31 J. J. Irwin, T. Sterling, M. M. Mysinger, E. S. Bolstad and R. G. Coleman, *J. Chem. Inf. Model.*, 2012, **52**, 1757–68.

32 KNIME | Konstanz Information Miner, University of Konstanz, Germany. 2015.

33 JChem 6.3.0, 2014, ChemAxon (http://www.chemaxon.com)

34. STATISTICA 12.5, StatSoft, Inc., Tulsa, OK 74104, USA, 2014.

35 A. N. Jain and A. Nicholls, *J. Comput. Aided. Mol. Des.*, 2008, **22**, 133–9.

36 A. Nicholls, *J. Comput. Aided. Mol. Des.*, 2014, **28**, 887–918.

MedChemComm

CONCISE ARTICLE

Tables

Table 1 Decision matrix for the assignment of desirability scores

	median (act.)	IQR (act.)	other (act.)
median (rand.)	–[a]	0.5	0
IQR (rand.)	1	0.5	0
other (rand.)	1	1	0, 0.2[b]

[a] No descriptors were selected where the medians of the kinase actives and random molecules coincide.

[b] In cases where a value is outside the interquartile range (IQR) for both sets, a score of 0.2 is assigned when the given value is visibly more common among kinase actives than random molecules (see Figures S1-S6).

Table 2 Performance evaluation of the Kinase Desirability Score: early enrichment factors and AUC values

Dataset	Active	Random compounds	$EF_{0.5\%}$[a]		$EF_{1\%}$[a]		$EF_{2\%}$[a]		$EF_{5\%}$[a]		AUC[a]	
			KiDS	KLS[b]	KiDS	KLS[b]	KiDS	KLS[b]	KiDS	KLS[b]	KiDS	KLS[b]
Training	2500	22803	23.2 (1.9E-2)	1.90 (5.1E-3)	14.2 (9.9E-3)	1.79 (3.5E-3)	10.6 (5.6E-3)	1.78 (2.4E-3)	7.1 (2.5E-3)	1.44 (1.3E-3)	0.786 (9.6E-3)	0.544 (0.012)
Test 1	1923	18000	22.6 (2.4E-2)	1.87 (6.5E-3)	14.0 (1.3E-2)	1.40 (3.9E-3)	10.9 (7.2E-3)	1.46 (2.8E-3)	6.9 (3.2E-3)	1.33 (1.7E-3)	0.778 (0.011)	0.537 (0.014)
Test 2	730	6300	18.9 (6.1E-2)	3.78 (2.6E-2)	14.8 (3.6E-2)	3.15 (1.7E-2)	9.9 (1.9E-2)	2.81 (1.1E-2)	6.2 (8.6E-3)	1.81 (5.3E-3)	0.757 (0.019)	0.532 (0.023)

[a] 1.96σ values (corresponding to 95% confidence intervals) are given in parentheses.[36]

[b] Performance parameters obtained for the same datasets with the KLS score of Singh *et al.* are provided as a reference.[17]

Table 3 Summary of the HTS sets applied for external validation

#[a]	AID[b]	Target	Activity threshold (μM)[c]	Confirmed active	Inactive	KiDS AUC[d]	KLS AUC[d,e]
A	GR	Undisclosed kinase target	70%[f]	28	7480	0.574 (0.110)[f]	0.397 (0.116)
B	524 (screening) 548 (confirmatory)	Protein kinase A (PKA)	60	40	22447	0.700 (0.075)	0.557 (0.086)
C	604 (screening) 644 (confirmatory)	Rho-associated protein kinase 2 (ROCK2)	10	35	20895	0.682 (0.080)	0.603 (0.083)
D	619 (screening) 785 (confirmatory)	Polo-like kinase 1 (PLK1)	50	14	30336	0.791 (0.102)	0.523 (0.131)

[a] Panel identifier on Figure 3.

[b] Pubchem Bioassay IDs (where applicable). GR: Gedeon Richter Plc. proprietary HTS dataset.

[c] IC_{50} value, below which a molecule is considered a confirmed active.

[d] 1.96σ values (corresponding to 95% confidence intervals) are given in parentheses.[36]

[e] AUC values obtained for the same datasets with the KLS score of Singh *et al.* are provided as a reference.[17]

[f] 70% inhibition at the HTS screening concentration of 10 μM. (As confirmation, single-point inhibition measurements were carried out at 10 μM in duplicate.)

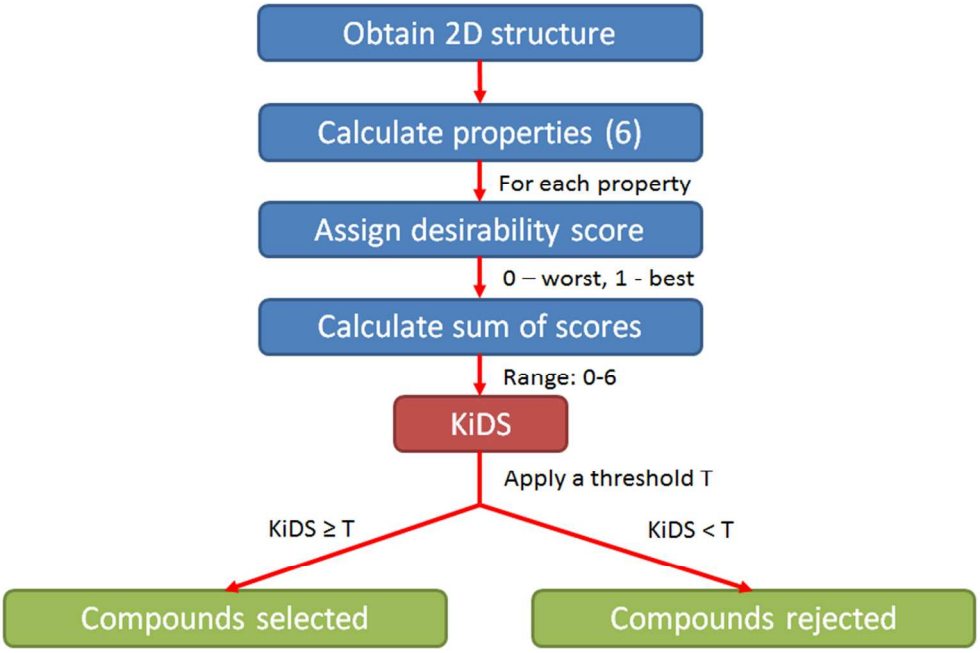
Figures

Figure 1 Workflow representation of the calculation of KiDS. The last step corresponds to the application of KiDS as a filtering criterion.

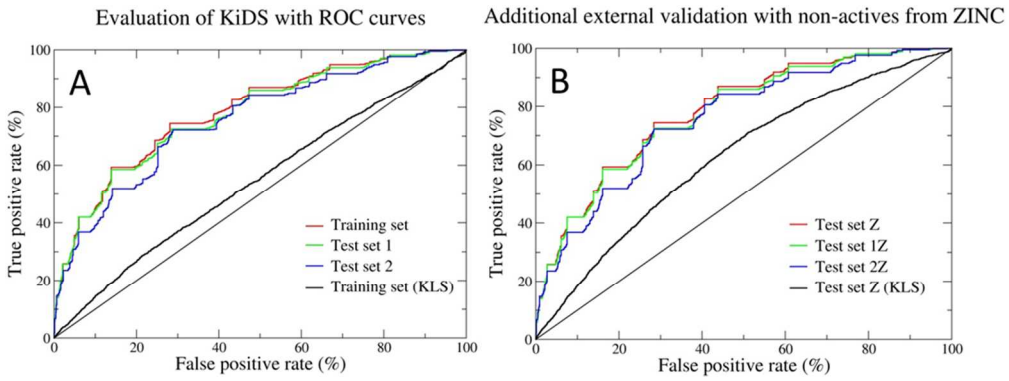
Figure 2 Evaluation of the Kinase Desirability Score. (A) ROC curves of the evaluation of the Training and Test sets with KiDS. In addition to the AUC values being close to 0.8, the initial slopes are quite high, which corresponds to good early enrichment factors (as reported in Table 2). A negligible deterioration of the results is observable for the Test sets (relative to the Training set), which suggests that the predictive power of the scoring method is sufficiently high, and thus it can be used for prospective applications. A ROC curve acquired for the Training set with the application of the Kinase-Like Score (KLS) of Singh *et al.* [17] is provided as a reference (thick black line). (B) Additional validation has been carried out with a different set of non-actives. The actives from the Training and Test sets were mixed with 20000 randomly selected leadlike molecules from the ZINC leadlike subset[30,31] to produce Test sets Z, 1Z and 2Z. The results are consistent with the curves presented in (A), confirming that no loss of performance was observed upon the exchange of the source of random compounds. A reference curve is provided once again for Test set Z with the KLS score of Singh *et al.*[17]

Figure 3 External validation of KiDS on proprietary (A) and publicly available (B-D) datasets of HTS campaigns. (See Table 3 for details) The ROC curves suggest the applicability of KiDS as a pre-screening step in HTS campaigns to reduce the necessary instrumentation (and thus, the effective cost) for finding hit compounds.

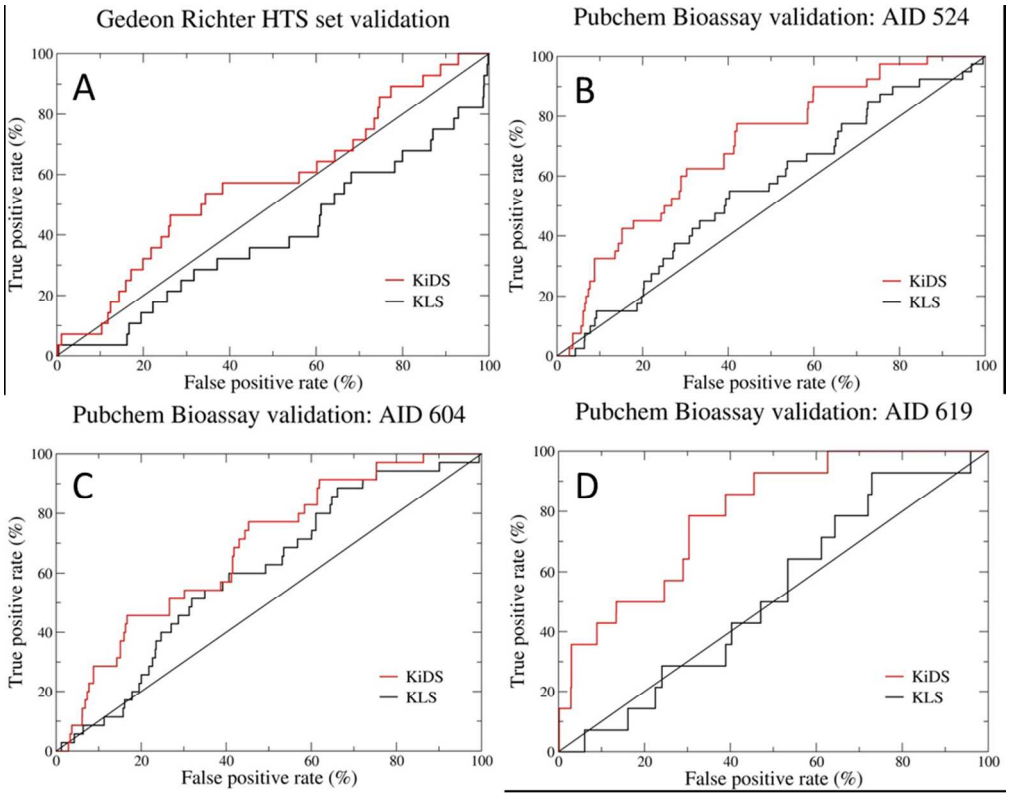
Figure 4 Plot of KiDS vs. average number of kinases hit for the EMD Millipore Kinase Screening dataset in ChEMBL.[25] For each point (X,Y), Y equals the number of kinases hit averaged over the compounds possessing a KiDS score less than or equal to X. A significant linear correlation can be observed between the KiDS score and kinase promiscuity, with $R^2 = 0.838$.



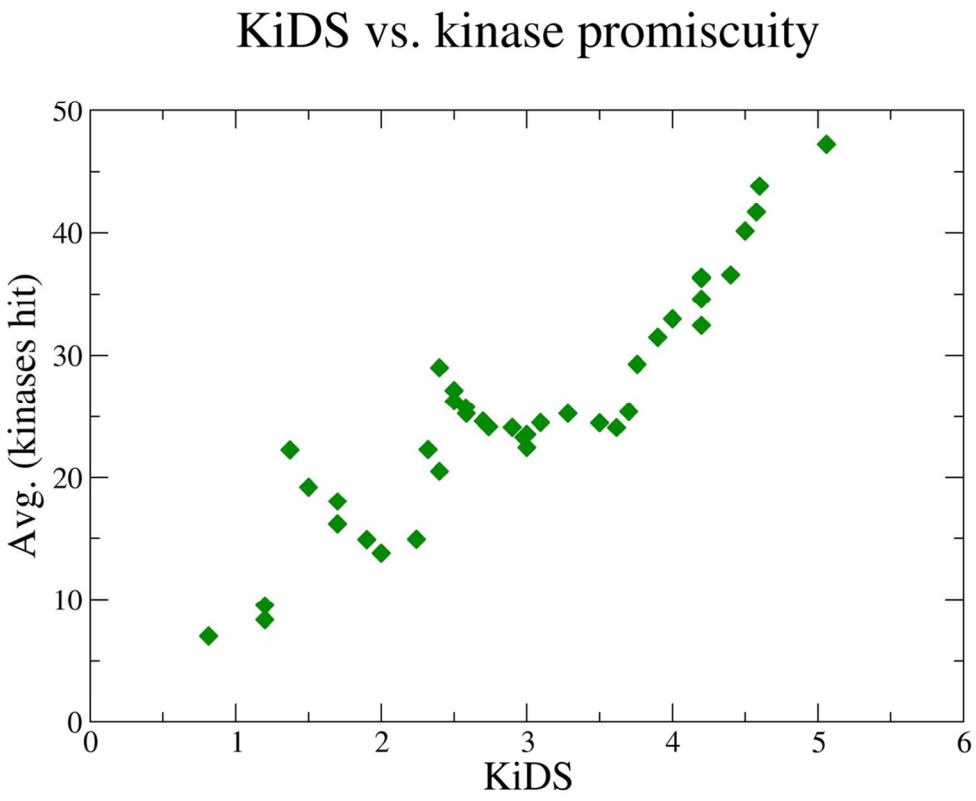
241x161mm (96 x 96 DPI)



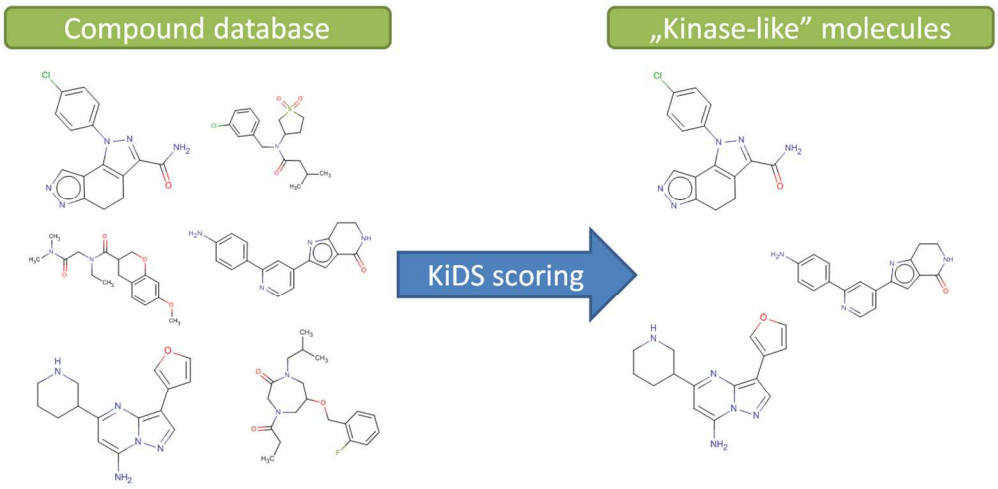
183x67mm (150 x 150 DPI)



174x137mm (150 x 150 DPI)



220x175mm (150 x 150 DPI)



366x178mm (96 x 96 DPI)