BMC Bioinformatics

A theoretical entropy score as a single value to express inhibitor selectivity

Joost CM Uitdehaag[*] and Guido JR Zaman

Abstract

Background: Designing maximally selective ligands that act on individual targets is the dominant paradigm in drug discovery. Poor selectivity can underlie toxicity and side effects in the clinic, and for this reason compound selectivity is increasingly monitored from very early on in the drug discovery process. To make sense of large amounts of profiling data, and to determine when a compound is sufficiently selective, there is a need for a proper quantitative measure of selectivity.

Results: Here we propose a new theoretical entropy score that can be calculated from a set of IC_{50} data. In contrast to previous measures such as the 'selectivity score', Gini score, or partition index, the entropy score is non-arbitrary, fully exploits IC_{50} data, and is not dependent on a reference enzyme. In addition, the entropy score gives the most robust values with data from different sources, because it is less sensitive to errors. We apply the new score to kinase and nuclear receptor profiling data, and to high-throughput screening data. In addition, through analyzing profiles of clinical compounds, we show quantitatively that a more selective kinase inhibitor is not necessarily more drug-like.

Conclusions: For quantifying selectivity from panel profiling, a theoretical entropy score is the best method. It is valuable for studying the molecular mechanisms of selectivity, and to steer compound progression in drug discovery programs.
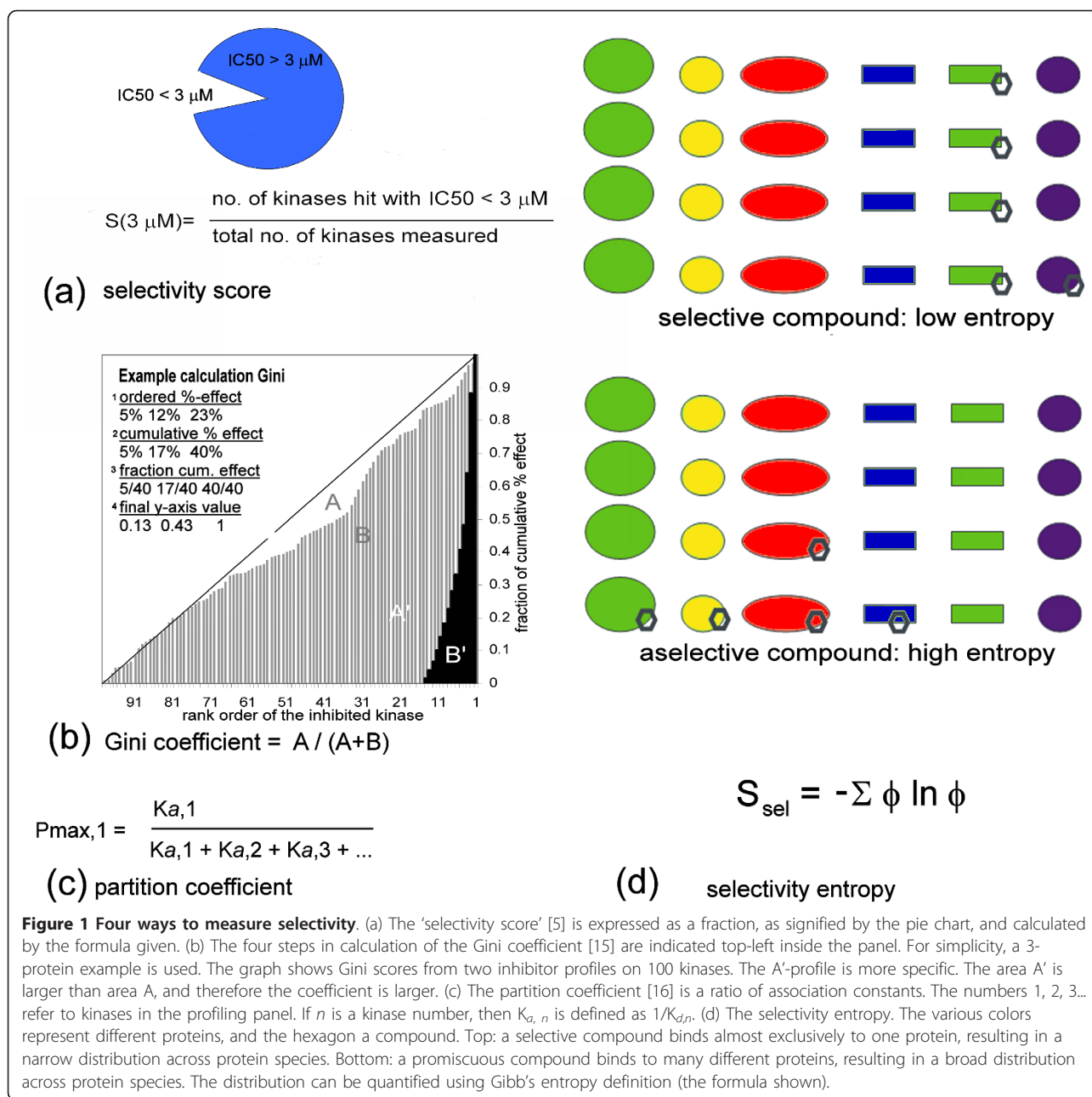
Background

In recent years, the kinase field has developed the practice of monitoring inhibitor selectivity through profiling on panels of biochemical assays [1-7], and other fields are following this example [8,9]. Such profiling means that scientists are faced with increasing amounts of data that need to be distilled into human sense. It would be powerful to have a good single selectivity value for quantitatively steering the drug discovery process, for measuring progress of series within a program, for computational drug design [10-12], and for establishing when a compound is sufficiently selective. However, in contrast to, for instance, lipophilicity and potency, where values such as logP or binding constant (K_d) are guiding, quantitative measures for selectivity are still under debate. Often graphic methods are used to give insight, for example dotting a kinome tree [13,14], heat maps [4,6], or a radius plot, but such methods only allow qualitative comparison of a limited set of compounds at a time.

To make quantitative selectivity comparisons, three notable methods have been proposed (Figure 1). The first is the 'selectivity score' [5], which simply divides the number of kinases hit at an arbitrary K_d or IC_{50} value (e.g. 3 µM) by the number of kinases tested (S (3 µM), Figure 1a). A related score is S(10x), which divides the number of kinases hit at 10 times the K_d of the target by the number of kinases tested [5]. The disadvantage of both methods is that 3 µM, or the factor 10, is an arbitrary cut-off value. For example, take two inhibitors, one that binds to two kinases with K_ds of 1 nM and 1 µM, and another with K_ds of 1 nM and 1 nM. Both are ranked equally specific by both S(3 µM) and S(10x), whereas the first compound is clearly more specific.

A less arbitrary parameter for selectivity is the Gini score [15]. This uses %-inhibition data at a single inhibitor concentration. These data are rank-ordered, summed and normalized (Figure 1b) to arrive at a cumulative

* Correspondence: joost.uitdehaag@merck.com
Merck Research Laboratories, Department of Molecular Pharmacology and DMPK PO Box 20, 5340 BH, Oss, The Netherlands

Figure 1 Four ways to measure selectivity. (a) The 'selectivity score' [5] is expressed as a fraction, as signified by the pie chart, and calculated by the formula given. (b) The four steps in calculation of the Gini coefficient [15] are indicated top-left inside the panel. For simplicity, a 3-protein example is used. The graph shows Gini scores from two inhibitor profiles on 100 kinases. The A'-profile is more specific. The area A' is larger than area A, and therefore the coefficient is larger. (c) The partition coefficient [16] is a ratio of association constants. The numbers 1, 2, 3... refer to kinases in the profiling panel. If n is a kinase number, then $K_{a,n}$ is defined as $1/K_{d,n}$. (d) The selectivity entropy. The various colors represent different proteins, and the hexagon a compound. Top: a selective compound binds almost exclusively to one protein, resulting in a narrow distribution across protein species. Bottom: a promiscuous compound binds to many different proteins, resulting in a broad distribution across protein species. The distribution can be quantified using Gibb's entropy definition (the formula shown).

fraction inhibition plot, after which the score is calculated by the relative area outside the curve (Figure 1b). Though this solves the problem with the selectivity score, it leaves other disadvantages. One is that the Gini score has no conceptual or thermodynamic meaning such as a K_d value has. Another is that it performs suboptimally with smaller profiling panels [16]. In addition, the use of %-inhibition data makes the value more dependent on experimental conditions than a K_d-based score [15]. For instance, profiling with 1 μM inhibitor concentration results in higher percentages inhibition than using 0.1 μM of inhibitor. The 1 μM test therefore

yields a more promiscuous Gini value, requiring the arbitrary 1 μM to be mentioned when calculating Gini scores. The same goes for concentrations of ATP or other co-factors. This is confusing and limits comparisons across profiles.

A recently proposed method is the partition index [16]. This selects a reference kinase (usually the most potently hit one), and calculates the fraction of inhibitor molecules that would bind this kinase, in an imaginary pool of all panel kinases (Figure 1c). The partition index (P_{max}) is a K_d-based score with a thermodynamical underpinning, and performs well when test panels are

smaller [16]. However, this score is still not ideal, since it doesn't characterize the complete inhibitor distribution in the imaginary kinase mixture, but just the fraction bound to the reference enzyme. Consider two inhibitors: A binds to 11 kinases, one with a K_d of 1 nM and ten others at 10 nM. Inhibitor B binds to 2 kinases, both with K_ds of 1 nM. The partition index would score both inhibitors as equally specific ($P_{max} = 0.5$), whereas the second is intuitively more specific. Another downside is the necessary choice of a reference kinase. If an inhibitor is relevant in two projects, it can have two different P_{max} values. Moreover, because the score is relative to a particular kinase, the error on the K_d of this reference kinase dominates the error in the partition index. Ideally, in panel profiling, the errors on all K_ds are equally weighted.

Here we propose a novel selectivity metric without these disadvantages. Our method is based on the principle that, when confronted with multiple kinases, inhibitor molecules will assume a Boltzmann distribution over the various targets (Figure 1d). The broadness of this distribution can be assessed through a theoretical entropy calculation (it is not actually measuring entropy). We show the advantages of this method and some applications. Because it can be used with any activity profiling dataset, it is a universal parameter for expressing selectivity.

Results and discussion
Theory

Imagine a theoretical mixture of all protein targets on which selectivity was assessed. No competing factors are present such as ATP. To this mixture we add a small amount of inhibitor, in such a way that approximately all inhibitor molecules are bound by targets, and no particular binding site gets saturated. A selective inhibitor will bind to one target almost exclusively and have a narrow distribution (low entropy, Figure 1d). A promiscuous inhibitor will bind to many targets and have a broad distribution (high entropy, Figure 1d). The broadness of the inhibitor distribution on the target mixture reflects the selectivity of the compound.

The binding of one inhibitor molecule to a particular protein can be seen as a thermodynamical state with an energy level determined by K_d (through $\Delta G = RT\ln K_d$). For simplicity we use the term K_d to represent both K_d and K_i. The distribution of molecules over these energy states is given by the Boltzmann law. As the broadness of a Boltzmann distribution is measured by entropy, the selectivity implied in the distributions of Figure 1d can be captured in an entropy.

A similar insight is given by information theory. It is well-established that information can be quantified using entropy [17]. A selective kinase inhibitor can be

seen as containing more information about which active site to bind than a promiscuous inhibitor. The selectivity difference between the inhibitors can therefore be quantified by information entropy.

The distribution of a compound across energy states is given by the Boltzmann formula [18]:

$$\phi_1 = e^{-\Delta G_1/kT} / \sum_i e^{-\Delta G_i/kT} \tag{1}$$

Where φ_1 is the fraction of molecules occupying state 1, and ΔG_1 is the free energy of occupying state 1 when the inhibitor comes from solution. In order to arrive at a fraction, the denominator in equation (1) contains the summation of occupancies of all states, which are labelled i, with free energies ΔG_i.

In general, entropy can be calculated from fractions of all l states using the Gibbs formula [18]:

$$S_{sel} = -\sum_l \phi_l \ln \phi_l \tag{2}$$

S_{sel} is shorthand for selectivity entropy. Compared to the original Gibbs formulation, equation (2) contains a minus sign on the right hand to ensure that S_{sel} is a positive value. Now, we need to evaluate equation (2) from a set of measurements. For this we need

$$\Delta G_i = RT \ln K_{d,i} = -RT \ln K_{a,i} \tag{3}$$

Where $K_{a,i}$ is the *association* constant of the inhibitor to target (or state) i, which is the inverse of the binding constant $K_{d,i}$ (which is a *dissociation* constant). In short: $K_{a,i} = 1/K_{d,i}$. If we express the free energy in units of 'per molecule' rather than 'per mole', equation (3) becomes

$$\Delta G_i = -kT \ln K_{a,i} \tag{4}$$

and equation (1) can be rewritten as

$$\phi_1 = e^{kT \ln K_{a,1}/kT} / \sum_i e^{kT \ln K_{a,i}/kT}$$
$$\Leftrightarrow e^{\ln K_{a,1}} / \sum_i e^{\ln K_{a,i}} = K_{a,1} / \sum_i K_{a,i} \tag{5}$$

Using this result in equation (2) gives

$$S_{sel} = -\sum_l \left(K_{a,l} / \sum_i K_{a,i}\right) \ln\left(K_{a,l} / \sum_i K_{a,i}\right) \tag{6}$$

Simplifying notation gives

$$S_{sel} = -\Sigma_a (K_a/\Sigma K) \ln (K_a/\Sigma K) \tag{7}$$

Equation (7) defines how a selectivity entropy can be calculated from a collection of association constants K_a. Here ΣK is the sum of all association constants.

It is most simple to apply equation (7) to directly measured binding constants or inhibition constants.

Also IC_{50}s can be used, but this is only really meaningful if they are related to K_d. Fortunately, for kinases it is standard to measure IC_{50} values at $[ATP] = K_{M,ATP}$. Ideally, such IC_{50}s equal 2 times K_d, according to the Cheng-Prusoff equation [19,20]. The factor 2 will drop out in equation (7), and we therefore can use data of the format IC_{50}-at-$K_{M, ATP}$ directly as if they were K_d.

Protocol for calculating a selectivity entropy

From the above, it follows that a selectivity entropy can be quickly calculated from a set of profiling data with the following protocol:

1. Generate K_a values by taking $1/K_d$ or $1/IC_{50}$
2. Add all K_a values to obtain ΣK
3. For every K_a, calculate $K_a/\Sigma K$
4. For every K_a, evaluate $(K_a/\Sigma K) \ln (K_a/\Sigma K)$
5. Sum all terms and multiply by -1

This process can be easily automated for use with large datasets [21] or internal databases.

Examples

The selectivity entropy is based on calculating the entropy of the hypothetical inhibitor distribution in a protein mixture. To give more insights into the properties of this metric, some examples are useful.

An inhibitor that only binds to a single kinase with a K_d of 1 nM ($K_a = 10^9$ M^{-1}) has $K_a/\Sigma K_a = 1$. Then $S_{sel} = -[1 \ln 1] = 0$, which is the lowest possibly entropy.

An inhibitor that binds to two kinases (X and Y) with a K_d of 1 nM has $K_x/\Sigma K_a = K_y/\Sigma K_a = 0.5$ and a selectivity entropy of $-[0.5 \ln 0.5 + 0.5 \ln 0.5] = 0.69$. Thus lower selectivity results in higher entropy.

If we modify the compound such that it still inhibits kinase X with a K_d of 1 nM, but inhibits less strongly kinase Y with a K_d of 1 μM, then the new inhibitor is more specific. Now $K_x/\Sigma K_a = 10^9/(10^9+10^6)$ and $K_y/\Sigma K_a = 10^6/(10^9+10^6)$, resulting in $S_{sel} = -[0.999 \ln 0.999 + 0.001 \ln 0.001] = 0.0079$. This is less than 0.69. This shows that the selectivity entropy can distinguish in the case where the selectivity scores S(3 μM) and S(10x) cannot (see above).

A less selective inhibitor that binds three targets with K_ds of 1 nM, has $S_{sel} = -3 \cdot [0.3 \ln 0.3] = 1.08$, and an even more promiscuous inhibitor that binds 5 targets, of which 3 at 1 nM, and 2 at 1 μM, has $\Sigma K = 3 \cdot 10^9 + 2 \cdot 10^6 = 3.002 \cdot 10^9$ and $S_{sel} = -3 \cdot [1 \cdot 10^9/3 \cdot 10^9 \ln 1 \cdot 10^9/3 \cdot 10^9] + 2 \cdot [1 \cdot 10^6/3 \cdot 10^9 \ln 1 \cdot 10^6/3 \cdot 10^9] = 3.07$. Thus S_{sel} gradually increases when more targets are more potently hit.

If we take the inhibitors A and B that were mentioned earlier, then A (with an inhibition profile of 1 nM, and ten times 10 nM), has $\Sigma K = 1 \cdot 10^9 + 10 \cdot 10^8 = 2 \cdot 10^9$ and $S_{sel} = -[1 \cdot 10^9/2 \cdot 10^9 \ln 1 \cdot 10^9/2 \cdot 10^9] + 10 \cdot [1 \cdot 10^8/2 \cdot 10^9 \ln 1 \cdot 10^8/2 \cdot 10^9] = 1.84$. This is a more aselective value than

inhibitor B with an inhibition profile of twice 1 nM, which has $S_{sel} = 0.69$ (see above). Thus the selectivity entropy can distinguish in a case where the partition coefficient P_{max} cannot.

Comparison to other methods

Having defined the entropy, we next investigated its performance relative to the most widely-used methods, on a public profiling dataset of 38 inhibitors on 290 non-mutant kinases [5] (Table 1 and Additional file 1). The values for Gini score, S(3 μM), S(10x) and partition coefficient, were taken from earlier work [16]. To this we added a K_a-Gini value and the selectivity entropy. The K_a-Gini is a Gini score directly calculated on K_as, without reverting to %-inhibition values (see below). From each of these scores we determined an inhibitor selectivity ranking, and a rank order difference compared to the entropy method (Uitdehaag_S1). In addition, to get an overview of the profiling raw data [5], we appended an activity-based heat map (Uitdehaag_S1).

From the rankings it is apparent that each of the earlier methods such as the classic Gini score, S(3 μM) and S(10x) generate considerable ranking differences compared to all other methods. This was observed earlier [16]. For the Gini score, this is related to the conversion from IC_{50} to %-inhibition, because the K_a-Gini gives more consistent rankings. For the S(3 μM) and the S (10x), the use of a cut-off is likely too coarse an approach. For instance in the case of S(10x), there are six inhibitors with a score of 0, making it impossible to distinguish between those highly specific compounds.

The newer methods such as P_{max}, K_a-Gini, and the selectivity entropy, give a more consistent ranking between them. For example, all three methods have PI-103, CI-1033, GW2580, VX-745 and gefitinib in their selectivity top five. There are differences however, most strikingly illustrated by the inhibitor SB-431542. This is ranked by P_{max} as 31st most selective, but by K_a-Gini and the selectivity entropy as 15th and 14th (Uitdehaag_S1). Also S(3 μM) ranks this ALK5 inhibitor [22] as selective. However, SB-431542 hits four kinases with very similar IC_{50}s between 100-300 nM, which leads to a broad partitioning over these kinases, resulting in a very promiscuous P_{max} of 0.14. The partition coefficient therefore ranks SB-431542 as almost equally selective to sunitinib ($P_{max} = 0.11$, rank 33). Nevertheless, sunitinib inhibits 181 kinases below 3 μM, and SB-431542 only 5. Therefore we think that K_a-Gini and the selectivity entropy are a better 'general' measure of selectivity in this case.

Another inhibitor scored differently is MLN-518 [23], which ranks 26st by P_{max}, but 14th and 15th by K_a-Gini and the selectivity entropy (Table 1 and Uitdehaag_S1). Again, these differences arise because this inhibitor hits

Table 1 Selectivity metrics calculated for the Ambit kinase profiling dataset

code	Gini	S3uM	S10x	Pmax	Ka-Gini	entropies
PI-103	0.41	0.00	0.00	0.97	0.99	0.05
CI-1033	0.72	0.15	0.00	0.88	0.99	0.17
GW-2580	0.54	0.01	0.00	0.92	0.99	0.26
VX-745	0.51	0.03	0.00	0.92	0.99	0.28
Gefitinib	0.65	0.07	0.00	0.89	0.99	0.44
Lapatinib	0.57	0.01	0.00	0.70	0.99	0.70
Erlotinib	0.75	0.15	0.00	0.67	0.99	0.88
EKB-569	0.78	0.18	0.00	0.67	0.99	0.89
Imatinib	0.79	0.07	0.02	0.65	0.99	1.05
CP-724714	0.49	0.02	0.01	0.33	0.99	1.10
CP-690550	0.66	0.03	0.01	0.50	0.99	1.11
BIRB-796	0.81	0.16	0.00	0.47	0.99	1.15
PTK-787	0.71	0.03	0.02	0.44	0.98	1.54
SB-431542	0.42	0.02	0.01	0.14	0.98	1.56
MLN-518	0.74	0.06	0.01	0.26	0.98	1.58
LY-333531	0.76	0.16	0.01	0.42	0.98	1.68
CHIR-258	0.76	0.33	0.01	0.44	0.97	1.80
AZD1152HQPA	0.78	0.10	0.02	0.34	0.98	1.91
ABT-869	0.82	0.16	0.01	0.32	0.98	1.93
MLN-8054	0.75	0.13	0.01	0.36	0.97	1.95
Roscovitine	0.33	0.03	0.03	0.12	0.98	2.02
SU-14813	0.70	0.51	0.02	0.18	0.97	2.04
Sunitinib	0.67	0.57	0.02	0.11	0.97	2.05
GW-786034	0.80	0.21	0.02	0.33	0.97	2.06
SB-202190	0.76	0.09	0.02	0.35	0.97	2.08
Sorafenib	0.82	0.18	0.10	0.46	0.97	2.15
SB-203580	0.77	0.10	0.02	0.24	0.97	2.26
AMG-706	0.82	0.09	0.04	0.22	0.97	2.36
SNS-032	0.80	0.13	0.06	0.30	0.97	2.44
CHIR265	0.74	0.13	0.12	0.36	0.96	2.47
Flavopiridol	0.77	0.19	0.17	0.42	0.95	2.50
AST-487	0.73	0.45	0.04	0.11	0.95	2.74
ZD-6474	0.79	0.27	0.26	0.28	0.94	2.88
Staurosporine	0.35	0.87	0.50	0.02	0.93	2.91
VX-680	0.76	0.38	0.03	0.15	0.93	3.09
Dasatinib	0.79	0.28	0.10	0.03	0.93	3.21
PKC-412	0.71	0.47	0.08	0.10	0.88	3.72
JNJ-7706621	0.74	0.37	0.08	0.10	0.87	3.73

Values for various selectivity metrics. In the Additional file 1, this table is extended with compound rank ordering and rank differences, and colour coded for clarity.

4 kinases with roughly equal potencies between 2-10 nM, leading to a promiscuous P_{max} (0.26). However, MLN-518 only hits 10 kinases below 3 μM, making it intuitively more selective than e.g. ZD-6474 [24] (P_{max} = 0.28, ranked 25th by P_{max}), which hits 79 kinases below 3 μM. These cases illustrate the earlier point that P_{max} underscores inhibitors that only hit a few kinases at comparable potencies. The Gini score and selectivity entropy assign a higher selectivity to these cases.

Finally, any selectivity score should be in line with the visual ranking from a heat map. The Additional file 1 shows that, generally, compounds with a higher entropy indeed have a busier heat map. A few exceptions stand out, which by eye appear more promiscuous than their entropy ranking indicates, for instance SU-14813, sunitinib and staurosporin. However, these compounds have extreme low K_ds on selected targets (SU-14813: 0.29 nM on PDGFRβ, sunitinib: 0.075 nM on PDGFRβ, staurosporin: 0.037 nM on LOK and 0.024 nM on SLK). Therefore they are relatively selective over activities in the 1-100 nM range, whereas these activities still fall within the highlighted ranges in Uitdehaag_S1. In a sense, the large dynamic range of the data limits visual assessment through a heat map.

Consistency across profiling methods

As a next step we selected 16 compounds from the public profile (Ambit) [5], and measured activity data on these using a different profiling service (Millipore, data available as Additional file 2). The 16 compounds represent a diversity of molecular scaffolds, promiscuity and target classes (Table 2). Also for these new data, we calculated the selectivity metrics (Uitdehaag_S2). In the ideal case, the selectivity values are similar irrespective of profiling technology (in the same way that a K_d value

Table 2 16 compounds used to check the robustness of the selectivity metrics

code	name	structure
ABT-869 / AL-39324	linifanib	
AZD-1152HQPA	barasertib	
AZD-6474	vandetanib	
BAY43-9006	sorafenib	
BIRB-796	doramapimod	
BMS-345825	dasatinib	
CP-690550	tasocitinib	
MK-0457 / VX-680	tozasertib	
MLN-518 / CT-53518	tandutinib	
MLN-8054	-	
PI 103	-	
RAF-265 / CHIR-265	-	
SB 203580	-	
STI-571	imatinib	
SU-11248	sunitinib	
VX-745	-	

is ideally independent of laboratory and assay format). The data of both methods are plotted in Figure 2.

All metrics except the entropy and P_{max} tend to be quite unevenly distributed. For instance all K_a-Gini scores fall between 0.93 and 1.00, where they can theoretically range from 0 to 1. If we nevertheless calculate the correlation statistics between both datasets, the R-square from linear regression and the correlation indicate that the selectivity entropy, S(3 μM) and K_a-Gini are the most robust methods (Figure 2).

It would be ideal if the absolute value of the metrics could also be compared between datasets. This means that a specificity of e.g. 1.2 in the first profile, would also score 1.2 in the second profile. To get insight in this, we calculated the best fit to a 1:1 correlation (the diagonal line in Figure 2), using normalized data. The K_a-Gini score was rescaled to its useful range of 0.93-1.00 (see legend to Figure 2), and then fitted. The S(3 μM) and the selectivity entropy have the best fit. The fact that here the K_a-Gini performs poorer is probably caused by the use of cumulative inhibition values (Figure 1b), which leads to the accumulation of errors (as pointed out in ref. 16).

In all fits, the P_{max} and S(10x) scores show worse fits and more scatter, indicating that these methods generate more error in their final value. For S(10x) and for P_{max}, this is because both methods make use of a reference value, usually the most potent IC_{50}, and errors in this reference value propagate more than errors in other IC_{50}s. Ideally, for S(10x) and P_{max}, the reference value specifically would have to be more accurately established.

If all analyses are taken together, the selectivity entropy avoids many pitfalls of the other methods (see above), shows consistent compound ranking (Table 1, Uitdehaag_S1), and is among the most robust methods across profiling datasets (Figure 2). For this reason, we propose the entropy method as the best metric for general selectivity.

Defining average selectivity

Quantification of selectivity helps to define when a compound is selective or promiscuous. Because of its consistency, the entropy method is ideally suited for benchmarking selectivity values. In the 290-kinase profiling dataset, the entropies are monomodally distributed, with an average of 1.8 (median of 1.9) and a standard deviation (σ) of 1.0 (not shown). Based on the correlation in Figure 2, it is expected that these statistics will be conserved in other profiling sets. Therefore, in general, a kinase compound with an entropy less than about 2 can be called selective, and more than 2 promiscuous. This provides a first quantitative definition of kinase selectivity.

Selectivity of allosteric inhibitors

It is generally thought that allosteric kinase inhibitors (known as type II, type III, or DFG-out inhibitors) are more selective [25,26]. The selectivity entropy now allows quantitative testing of this idea. We identified, from literature, which inhibitors in the profiling datasets are type II and III, based on X-ray structures. Sorafenib induces the kinase DFG-out conformation in B-RAF [27], nilotinib and gleevec in Abl [28], GW-2580 in Fms [29] and BIRB-796 in p38α [30]. Lapatinib induces a C-helix shift in EGFR [31]. PD-0325901 [32] and AZD-6244 induce a C-helix shift in MEK1 [32]. All other kinase inhibitors in the profile were labelled type I. Comparing the entropy distributions in both samples shows that type II/III inhibitors have significantly lower entropies (Figure 3a). Although other factors, such as the time at which a compound was developed, could influence the entropy differences, the correlation between low entropy and allostery strongly supports the focus on allostery for developing specific inhibitors [25,26].

Among the specific inhibitors in the type I category, 3D-structures of PI-103, CI-1033 and VX-745 bound to their targets have not been determined. Therefore,
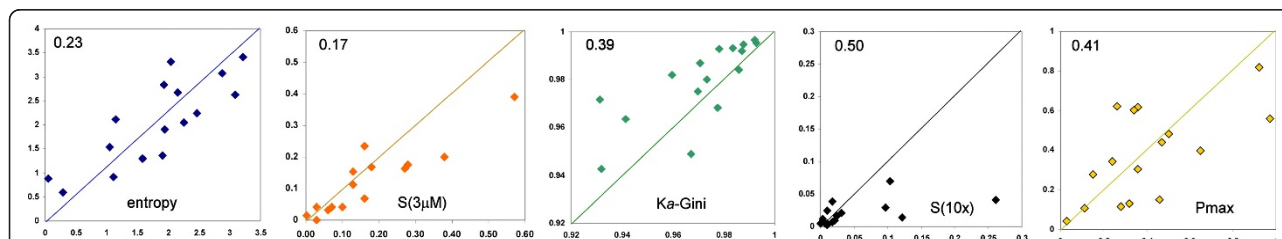

Figure 2 Correlation between specificity values calculated from different datasets. All x-axes: scores from binding data from the Ambit kinase dataset [5]. All y-axes: scores from activity data measured on the same compounds at Millipore. We calculated (in Microsoft Excel, for panels from left to right) the R-squares from linear regression as: 0.93, 0.92, 0.99, 0.54, 0.81 and the correlation coefficients as: 0.81, 0.90, 0.75, 0.57, 0.63. The straight line represents the ideal case of specificity values being insensitive to profiling method. The total squared distance of (normalized) data points to the straight line is given in the top left corner of each panel. For this latter calculation, data were normalized by dividing all values by the highest value in their set. Because the K_a-Gini values are very unevenly distributed, the lowest value 0.93 was first subtracted from all data in this set. Irrespective of statistical method, the selectivity entropy, S(3 μM) and K_a-Gini are the most robust metrics.
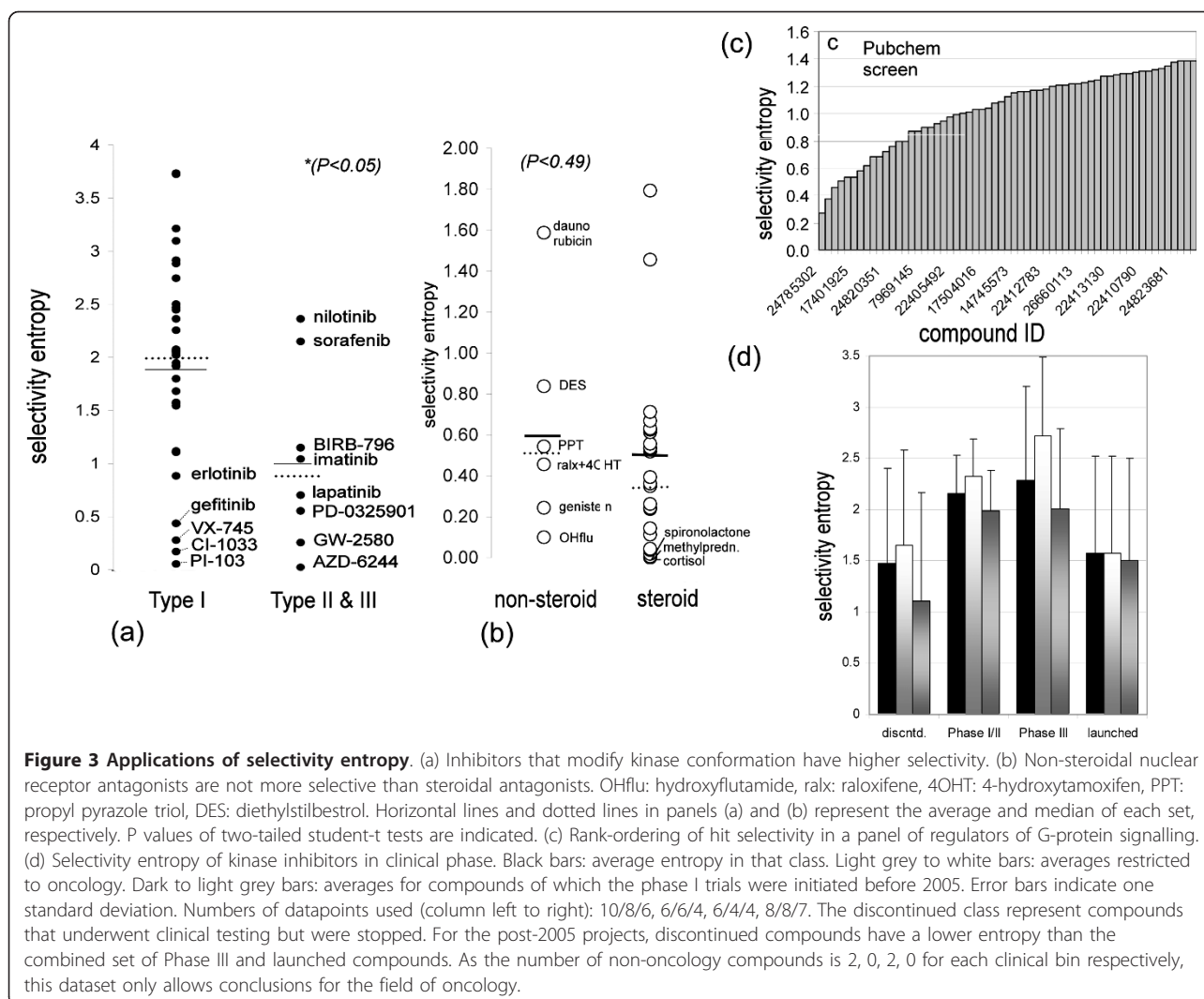
Figure 3 Applications of selectivity entropy. (a) Inhibitors that modify kinase conformation have higher selectivity. (b) Non-steroidal nuclear receptor antagonists are not more selective than steroidal antagonists. OHflu: hydroxyflutamide, ralx: raloxifene, 4OHT: 4-hydroxytamoxifen, PPT: propyl pyrazole triol, DES: diethylstilbestrol. Horizontal lines and dotted lines in panels (a) and (b) represent the average and median of each set, respectively. P values of two-tailed student-t tests are indicated. (c) Rank-ordering of hit selectivity in a panel of regulators of G-protein signalling. (d) Selectivity entropy of kinase inhibitors in clinical phase. Black bars: average entropy in that class. Light grey to white bars: averages restricted to oncology. Dark to light grey bars: averages for compounds of which the phase I trials were initiated before 2005. Error bars indicate one standard deviation. Numbers of datapoints used (column left to right): 10/8/6, 6/6/4, 6/4/4, 8/8/7. The discontinued class represent compounds that underwent clinical testing but were stopped. For the post-2005 projects, discontinued compounds have a lower entropy than the combined set of Phase III and launched compounds. As the number of non-oncology compounds is 2, 0, 2, 0 for each clinical bin respectively, this dataset only allows conclusions for the field of oncology.

potentially, these inhibitors could also derive their specificity from a form of undiscovered induced fit. Indeed, VX-745-related compounds induce a peptide flip near Met109/Gly110 in P38α [33]. Of the five most selective compounds in Table 1, only gefitinib so far is undoubtedly a type I inhibitor [34], making this EGFR inhibitor an interesting model for the structural biology of non-allosteric specificity.

Use of selectivity measures in nuclear receptor profiling

Selectivity profiling is most advanced in the kinase field, but is emerging in other fields. To illustrate that selectivity metrics such as the entropy can also be used with other target families, we investigated a long-standing question in the nuclear receptor field: are non-steroidal ligands more selective than steroidals? [35]. For this, we calculated the entropies of a published profile of 35 antagonists on a panel of 6 steroid receptors [9] (the androgen receptor, estrogen receptor α, estrogen

receptor β, mineralocorticoid receptor, glucocorticoid receptor, and progesterone receptor). This shows that there are no statistically significant selectivity differences between steroidals and non-steroidals (Figure 3b). A more important determinant for selectivity could be, in parallel to kinase inhibitors, if a ligand induces a conformational change. Indeed, many nuclear receptor agonists are known to induce a transformation from a flexible receptor to a rigid agonistic form [36-40], or a heterodimer form [41,42]. In contrast, antagonists are know to displace helix 12 specifically from the agonistic form [36]. Thus, the large role of induced fit in ligand binding to nuclear receptors might explain the relative high selectivity of these ligands [9,36,43,44].

Use in hit prioritization

Aside from solving questions in the structure-function area, the selectivity entropy can be used during drug discovery. Previously it has been shown that selectivity

metrics can be used in lead optimization projects to classify compounds, set targets, and rationalize improvement [16]. In addition, metrics such as the entropy are useful in evaluating screening data, especially now screening larger compound collections in parallel assays is increasingly popular.

We downloaded PubChem data of 59 compounds tested in a panel of four assays for regulators of G protein signalling (RGS) [21]. These data were selected because they were publicly available and were neither a kinase nor a nuclear receptor panel. In addition the data were dose-response, were all in a similar assay format, and were ran in the same lab with the same compound set.

We calculated the compound entropies across the RGS panel, and used them for ranking, which immediately distinguishes the scaffolds that are specific (Figure 3c). The best are ID 24785302, a pyrazole-phenoxy derivative, and ID 24834029, a bicyclo-octane derivative, which are likely to be better lead optimization starting points than more promiscuous scaffolds. Triaging compounds by entropy is a far more time-efficient and unbiased way than manual evaluation of four parallel columns of data. Indeed, listing of the selectivity entropy in public databases of screening data would provide users with immediate information on scaffold promiscuity.

Selectivity and clinical outcome

Finally, the selectivity entropy can be used to study clinical success. Selective compounds are generated because they are thought to be less toxic and therefore better doseable to effective ranges [45]. To test the hypothesis that clinically approved inhibitors are more selective, we binned the compounds in the public kinase profile [5] according to their clinical history, and calculated their average entropies (Figure 3d, Additional file 3). Compared to the average discontinued compound, the average marketed kinase inhibitor is not more selective, and the average Phase III compound is even significantly more aselective. To exclude therapy area effects, we also performed the analysis for compounds in the oncology area, which is the only therapeutic area with a statistically significant amount of projects. This leads to a similar conclusion (Figure 3d). To exclude effects of time from this analysis (more recently invented kinase inhibitors might be more selective, because of advances in the kinase field), we repeated the analysis for compounds that entered clinical phase I before 2005. This shows even more clearly that more succesful compounds are, if anything, more broadly selective (Figure 3d).

Behind such statistics lies the success of, for instance, the spectrum selective drugs dasatinib, sorafenib and sunitinib (an average entropy of 3.13), and the failure of the highly selective MEK-targeted drugs PD-0325901 and CI-1040 (an average entropy of 0.32). Because 66-100% of the analysed compounds in each clinical bin are (or were) developed for oncology, our conclusion is primarily valid for oncology, until more kinase inhibitors enter the clinic for other indications. Nevertheless, the finding that a selective kinase inhibitor has fewer chances of surviving early clinical trials fuels the notion that polypharmacology is sometimes required to achieve effect (in oncology) [45-47].

Conclusions

In order to quantify compound selectivity as a single value, based on data from profiling in parallel assays, we have presented a selectivity entropy method, and compared this to other existing methods. The best method should avoid artifacts that obscure compound ranking, and show consistent values across profiling methods. Based on these criteria, the selectivity entropy is the best method.

A few cautionary notes are in order. First, the method is labelled an entropy in the sense of information theory [17], which is different to entropy in the sense of vibrational modes in enzyme active sites. Whereas these vibrations can form a physical basis for selectivity [39,48,49], our method is a computational metric to condense large datasets.

Secondly, any selectivity metric that produces a general value does not take into account the specific importance of individual targets. Therefore, the entropy is useful for generally characterizing tool compounds and drug candidates, but if particular targets need to be hit, or avoided, the K_ds on these individual targets need to be monitored. It is possible to calculate an entropy on any particular panel of all-important targets, or to assign a weighing factor to every kinase, as suggested for P_{max} [16] and calculate a weighted entropy. However, the practicality of this needs to be assessed.

Next, it is good custom to perform profiling in biochemical assays at [ATP] = K_{M-ATP}, because this generates IC_{50}s that are directly related to the ATP-independent K_d value. However, in a cellular environment, there is a constant high (~5 mM) ATP concentration and therefore a biochemically selective inhibitor will act with different specificity in a cell. If the inhibitor has a specificity for a target with a $K_{M,ATP}$ above the panel average, then that inhibitor will act even more specifically in a cell and vice versa ($K_{M,ATP}$ values can generally be found on websites of profiling research organizations). Selectivity inside the cell is also determined by factors such as cellular penetration, compartimentalization and metabolic activity [39]. Therefore,

selectivity from biochemical panel profiling is only a first step in developing selective inhibitors.

Another point is that any selectivity metric is always associated with the assay panel used, and the entropy value will change if an inhibited protein is added to the panel. Adding a protein that does not bind inhibitor will not affect the entropy value. In this way the discovery of new inhibitor targets by e.g. pulldown experiments, can change the idea of inhibitor selectivity, and also the entropy value. A good example is PI-103, the most selective inhibitor in Table 1, which in the literature is known as a dual PI3-kinase/mTOR inhibitor [50], and which appears specific in Table 1 because PI3-kinase is not incorporated in the profiling panel.

In addition, an inhibitor that hits 2 kinases at 1 nM from a panel of 10 has the same selectivity entropy as an inhibitor that inhibits 2 kinases at 1 nM in a panel of 100. However, intuitively, the second inhibitor is more specific (the 'selectivity score' differentiates in this case). This illustrates that it is important to compare entropy scores on similar panels. At the same time, when results from different panels are weighed, as in the example, it should not be assumed for the first inhibitor, that it is inactive against all 90 other kinases in the second panel. It would be better to assign an average K_d where measurements are missing. In that case the first inhibitor would score a more promiscuous entropy compared to the second inhibitor.

Finally it must be stressed that the selectivity entropy could be applied in many more fields. It could, for instance, be a useful metric in the computational studies that attempt to link compound *in vitro* safety profiles to compound characteristics [51-53]. Currently, that field uses various forms of 'promiscuity scores' which bear similarity to the selectivity score. A more robust and non-arbitrary metric such as the selectivity entropy could be of help in building more detailed pharmacological models of compound activity-selectivity relationships [51-53].

In summary, the selectivity entropy is a very useful tool for making sense of large arrays of profiling data. We have demonstrated its use in characterizing tool compounds and drug candidates. Many more applications are imaginable in fields where an array of data is available and the selectivity of a response needs to be assessed. In that sense, the selectivity entropy is a general aid in the study of selectivity.

Methods
Calculation of other selectivity scores
For comparisons between currently used methods, we calculated the selectivity scores S(3 μM) and S(10x) as outlined above and in ref. 5. The partition coefficient P_{max} was calculated as originally proposed [16], by taking

the K_a value of the most potently hit kinase, and dividing it by ΣK_a. It is worth to note that the partition coefficient is the same as φ_l in our entropy equation (eq. 2).

The Gini score was calculated from data on %-inhibition [15]. In Figure 1b, these data were extracted from K_d values using the Hill expression: %-inhibition = 100/$(1+10^{-(pK_d - pconc)})$, where pK_d = -log (K_d) and pconc = -log (inhibitor concentration evaluated). In addition, to work more directly with K_ds, we also introduce a K_a-Gini score, in which association constants are used for rank-ordering the kinase profile. From this K_a-rank ordering, a cumulative effect is calculated and normalized, after which the areas are determined, in the same way as for the original Gini score [15]. All calculations were done in Microsoft Excel.

Sources of existing and new data
For our comparative rank-ordering (Table 1, Uitdehaag_S1) we used the publicly available dataset released by Ambit http://www.ambitbio.com, which contains binding data (K_ds) of 38 inhibitors on 290 kinases (excluding mutants), and which is currently the largest single profiling set available [5].

For comparing profiles across methods (Figure 2), we selected 16 kinase inhibitors of the Ambit profile (Table 2) and submitted these to the kinase profiling service from Millipore (http://www.millipore.com/drug-discovery/svp3/kpservices, data available as Additional file 2). Both profiling methods are described earlier [3,5,14] and differ (among other variations) in the following way: Ambit uses a competitive binding setup in absence of ATP on kinases from T7 or HEK293 expression systems [14]. Millipore uses a radioactive filter binding activity assay, with kinases purified from *Escherichia coli* or baculovirus expression systems [3]. All Millipore profiling was done on 222 human kinases at [ATP] = $K_{M,ATP}$.

For comparing inhibitors with an allosteric (actually: induced fit) profile (Figure 3a), we used data from the Ambit profile [5], supplemented with Millipore profiling data on nilotinib, PD-0325901 and AZD6244, because these important inhibitors were lacking in the Ambit dataset (data available in Additional file 2).

For comparing nuclear receptor data (Figure 3b), we used the published profiling dataset of 35 inhibitors on a panel consisting of all six steroid hormone receptors [9] The data we used were EC_{50}s in cell-based assays.

For evaluation of a screening dataset (Figure 3c), we selected data from the PubChem initiative, determined at the University of New Mexico on regulators of G protein signalling (isoforms 4, 19, 7 and 16. Assay identifiers: 1872, 1884, 1888 and 1869) [21].

For evaluating clinical success (Figure 3d), we tracked the clinical status of each compound in the

Ambit profile using the Thompson Pharma® database (status February 2011, analysis availabe as Additional file 3).

Additional material

Additional file 1: Selectivity metrics, heat maps, and selectivity rank ordering for the Ambit profiling dataset (an extension of Table 1).

Additional file 2: EC50 values and selectivity metrics from an activity based profiling of 16 reference inhibitors.

Additional file 3: Selectivity entropy and status of clinically tested kinase inhibitors.

Authors' contributions
JU conceived the entropy principle and drafted the manuscript. GZ organized the kinase profiling data and helped to draft the manuscript. All authors read and approved the final manuscript.

References
1. Davies SP, Reddy H, Caivano M, Cohen P: **Specificity and mechanism of some commonly used protein kinase inhibitors.** *Biochem J* 2000, **351(Pt1)**:95-105.
2. Vieth M, Higgs RE, Robertson DH, Shapiro M, Gragg EA, Hemmerle H: **Kinomics - structural biology and chemogenomics of kinase inhibitors and targets.** *Biochem Biophys Acta* 2004, **1697(1-2)**:243-257.
3. Bain J, Plater L, Elliott M, Shapiro N, Hastie CJ, McLauchlan H, Klevernic I, Arthur JSC, Alessi DR, Cohen P: **The selectivity of protein kinase inhibitors: a further update.** *Biochem J* 2007, **408(3)**:297-315.
4. Fedorov O, Marsden B, Pogacic V, Rellos P, Müller S, Bullock AN, Schwaller J, Sundström M, Knapp S: **A systematic interaction map of validated kinase inhibitors with Ser/Thr kinases.** *Proc Natl Acad Sci USA* 2007, **104(51)**:20523-20528.
5. Karaman MW, Herrgard S, Treiber DK, Gallant P, Atteridge CE, Campbell BT, Chan KW, Ciceri P, Davis MI, Edeen PT, Faraoni R, Floyd M, Hunt JP, Lockhart DJ, Milanov ZV, Morrison MJ, Pallares G, Patel HK, Pritchard S, Wodicka LM, Zarrinkar PP: **A quantitative analysis of kinase inhibitor selectivity.** *Nature Biotechnol* 2008, **26(1)**:127-132.
6. Bamborough P, Drewry D, Harper G, Smith GK, Schneider K: **Assessment of chemical coverage of kinome space and its implications for kinase drug discovery.** *J Med Chem* 2008, **51(24)**:7898-7914.
7. Smyth LA, Collins I: **Measuring and interpreting the selectivity of protein kinase inhibitors.** *J Chem Biol* 2009, **2(3)**:131-151.
8. Heilker R, Wolff M, Tautermann CS, Bieler M: **G-protein-coupled receptor focused drug discovery using a target class platform approach.** *Drug Discov Today* 2009, **14(5-6)**:231-240.
9. Wilkinson JM, Hayes S, Thompson D, Whitney P, Bi K: **Compound profiling using a panel of steroid hormone receptor cell-based assays.** *J Biomol Screen* 2008, **13(8)**:755-765.
10. Sciabola S, Stanton RV, Wittkop S, Wildman S, Moshinsky D, Potluri S, Xi H: **Predicting kinase selectivity profiles using free-wilson QSAR analysis.** *J Chem Inf Model* 2008, **48(9)**:1851-1867.
11. Sheridan RP, Nam K, Maiorov VN, McMasters DR, Cornell WD: **QSAR models for predicting the similarity in binding profiles for pairs of protein kinases and the variation of models between experimental data sets.** *J Chem Inf Model* 2009, **49(8)**:1974-1985.
12. Brandt P, Jensen AJ, Nilsson J: **Small kinase assay panels can provide a measure of selectivity.** *BioOrg Med Chem Letters* 2009, **19(20)**:5861-5863.
13. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S: **The protein kinase complement of the human genome.** *Science* 2002, **298(5600)**:1912-1934.
14. Fabian MA, Biggs WH III, Treiber DK, Atteridge CE, Azimioara MD, Benedetti MG, Carter TA, Ciceri P, Edeen PT, Floyd M, Ford JM, Galvin M, Gerlach JL, Grotzfeld RM, Herrgard S, Insko DE, Insko MA, Lai AG, Lélias JM, Mehta SA, Milanov ZV, Velasco AM, Wodicka LM, Patel HK, Zarrinkar PP, Lockhart DJ: **A small molecule-kinase interaction map for clinical kinase inhibitors.** *Nature Biotechnol* 2005, **23(3)**:329-336.
15. Graczyk P: **Gini coefficient: a new way to express kinase selectivity against a family of kinases.** *J Med Chem* 2007, **50(23)**:5773-5779.
16. Cheng AC, Eksterowicz J, Geuns-Meyer S, Sun Y: **Analysis of kinase inhibitor selectivity using a thermodynamics-based partition index.** *J Med Chem* 2010, **53(11)**:4502-4510.
17. Shannon CE: **A mathematical theory of communication.** *The Bell Systems Technical J* 1948, **27**:379-423.
18. Atkins P, de Paula J: *Atkins' Physical Chemistry* Oxford University Press: Oxford; 1970.
19. Cheng Y, Prusoff WH: **Relationship between the inhibition constant (K_i) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction.** *Biochem Pharmacol* 1973, **22(23)**:3099-3108.
20. Knight ZA, Shokat KM: **Features of selective kinase inhibitors.** *Chem Biol* 2005, **12(6)**:621-637.
21. Roman DL, Talbot JN, Roof RA, Sunahara RK, Traynor JR, Neubig RR: **Identification of small-molecule inhibitors of RGS4 using a high-troughput flow cytometry protein interaction assay.** *Mol Pharmacol* 2006, **71(1)**:169-175.
22. Inman GJ, Nicolás FJ, Callahan JF, Harling JD, Gaster LM, Reith AD, Laping NJ, Hill CS: **SB-431542 is a potent and specific inhibitor of transforming growth factor-β superfamily type I activin receptor-like kinase (ALK) receptors ALK4, ALK5, and ALK7.** *Mol Pharmacol* 2002, **62(1)**:65-74.
23. Kelly LM, Yu JC, Boulton CL, Apatira M, Li J, Sullivan CM, Williams I, Amaral SM, Curley DP, Duclos N, Neuberg D, Scarborough RM, Pandey A, Hollenbach S, Abe K, Lokker NA, Gilliland DG, Giese NA: **CT53518, a novel selective Flt3 antagonist for the treatment of acute myelogenous leukemia (AML).** *Cancer Cell* 2002, **1(5)**:421-432.
24. Hennequin LF, Stokes ES, Thomas AP, Johnstone C, Plé PA, Ogilvie DJ, Dukes M, Wedge SR, Kendrew J, Curwen JO: **Novel 4-anilinoquinazolines with C-7 basic side chains: design and structure activity relationship of a series of potent, orally active, VEGF receptor tyrosine kinase inhibitors.** *J Med Chem* 2002, **45(6)**:1300-1312.
25. Liu Y, Gray NS: **Rational design of inhibitors that bind to inactive kinase conformations.** *Nature Chem Biol* 2006, **2(7)**:358-364.
26. Simard JR, Klüter S, Grütter C, Getlik M, Rabiller M, Rode HB, Rauh D: **A new screening assay for allosteric inhibitors of Src.** *Nature Chem Biol* 2009, **5(6)**:395-396.
27. Wan PTC, Garnett MJ, Roe SM, Lee S, Niculescu-Duvaz D, Good VM, Cancer Genome Project, Jones CM, Marshall CJ, Springer CJ, Barford D, Marais R: **Mechanism of activation of the RAF-ERK signaling pathway by oncogenic mutations of B-RAF.** *Cell* 2004, **116(6)**:855-867.
28. Vajpai N, Strauss A, Fendrich G, Cowan-Jacob SW, Manley PW, Grzesiek S, Jahnke W: **Solution conformations and dynamics of ABL kinase-inhibitor complexes determined by NMR substantiate the different binding modes of imatinib/nilotinib and dasatinib.** *J Biol Chem* 2008, **283(26)**:18292-18302.
29. Meyers MJ, Pelc M, Kamtekar S, Day J, Poda GI, Hall MK, Michener ML, Reit BA, Mathis KJ, Pierce BS, Parikh MD, Mischke DA, Long SA, Parlow JJ, Anderson DR, Thorarensen A: **Structure-based drug design enables conversion of a DFG-in binding CSF-1R kinase inhibitor to a DFG-out binding mode.** *BioOrg Med Chem Letters* 2010, **20(5)**:1543-1547.
30. Pargellis C, Tong L, Churchill L, Cirillo PF, Gilmore T, Graham AG, Grob PM, Hickey ER, Moss N, Pav S, Regan J: **Inhibition of p38 MAP kinase by utilizing a novel allosteric binding site.** *Nature Struct Biol* 2002, **9(4)**:268-272.
31. Wood ER, Truesdale AT, McDonald OB, Yuan D, Hassell A, Dickerson SH, Ellis B, Pennisi C, Horne E, Lackey K, Alligood KJ, Rusnak DW, Gilmer TM, Shewchuk L: **A unique structure of epidermal growth factor receptor bound to GW572016 (lapatinib): relationships among protein conformation, inhibitor off-rate, and receptor activity in tumor cells.** *Cancer Res* 2004, **64(18)**:6652-6659.

32. Fischmann TO, Smith CK, Mayhood TW, Meyers JE Jr, Reichert P, Mannarino A, Carr D, Zhu H, Wong J, Yang RS, Le HV, Madison VS: **Crystal structures of MEK1 binary and ternary complexes with nucleotides and inhibitors.** *Biochemistry* 2009, **48(12)**:2661-2674.

33. Fitzgerald CE, Patel SB, Becker JW, Cameron PM, Zaller D, Pikounis VB, O'Keefe SJ, Scapin G: **Structural basis for p38α MAP kinase quinazolinone and pyridol-pyrimidine inhibitor specificity.** *Nature Struct Biol* 2003, **10(9)**:764-769.

34. Johnson LN: **Protein kinase inhibitors: contributions from structure to clinical compounds.** *Q Rev Biophysics* 2009, **42(1)**:1-40.

35. Hermkens PH, Kamp S, Lusher S, Veeneman GH: **Non-steroidal steroid receptor modulators.** *IDrugs* 2006, **9(7)**:488-494.

36. Egea PF, Klaholz BP, Moras D: **Ligand-protein interaction in nuclear receptors of hormones.** *FEBS Letters* 2000, **476(1-2)**:62-67.

37. Nettles KW, Sun J, Radek JT, Sheng S, Rodriguez L, Katzenellenbogen JA, Katzenellenbogen BS, Greene GL: **Allosteric control of ligand selectivity between estrogen receptors α and β.** *Mol Cell* 2004, **13(3)**:317-327.

38. Raaijmakers HC, Versteegh JE, Uitdehaag JCM: **The X-ray structure of RU486 bound to the progesterone receptor in a destabilized agonistic conformation.** *J Biol Chem* 2009, **284(9)**:19572-19579.

39. Martinez L, Nascimento AS, Nunes FM, Phillips K, Aparicio R, Dias SM, Figueira AC, Lin JH, Nguyen P, Apriletti JW, Neves FAR, Baxter JD, Webb P, Skaf MS, Polikarpov I: **Gaining ligand selectivity in thyroid hormone receptors via entropy.** *Proc Natl Acad Sci USA* 2009, **106(49)**:20717-20722.

40. Togashi M, Borngraeber S, Sandler B, Fletterick RJ, Webb P, Baxter JD: **Conformational adaptation of nuclear receptor ligand binding domains to agonists: potential for novel approaches to ligand design.** *J Steroid Biochem Mol Biol* 2005, **93(2-5)**:127-137.

41. Bourguet W, Vivat V, Wurtz JM, Chambon P, Gronemeyer H, Moras D: **Crystal structure of a heterodimeric complex of RAR and RXR ligand-binding domains.** *Mol Cell* 2000, **5(2)**:289-298.

42. Fradera X, Vu D, Nimz O, Skene R, Hosfield D, Wynands R, Cooke AJ, Haunsø A, King A, Bennett DJ, McGuire R, Uitdehaag JCM: **X-ray structures of the LXRα LBD in its homodimeric form and implication for heterodimer signaling.** *J Mol Biol* 2010, **399(1)**:120-132.

43. Cornell W, Nam K: **Steroid hormone binding receptors: application of homology modeling, induced fit docking, and molecular dynamics to study structure function relationships.** *Curr Top Med Chem* 2009, **9(9)**:844-853.

44. Nabuurs SB, Wagener M, de Vlieg J: **A flexible approach to induced fit docking.** *J Med Chem* 2007, **50(26)**:6507-6518.

45. Knight ZA, Lin H, Shokat KM: **Targeting the cancer kinome through polypharmacology.** *Nature Rev Cancer* 2010, **10(2)**:130-137.

46. Morphy R, Rankovic Z: **Designing multiple ligands - medicinal chemistry strategies and challenges.** *Curr Pharm Design* 2009, **15(6)**:587-600.

47. Hopkins A: **Network pharmacology: the next paradigm in drug discovery.** *Nature Chem Biol* 2008, **4(11)**:682-690.

48. Robinson D, Sherman W, Farid R: **Understanding kinase selectivity through energetic analysis of binding site waters.** *Chem Med Chem* 2010, **5(4)**:618-627.

49. Scapin G: **Protein kinase inhibition: different approaches to selective inhibitor design.** *Curr Drug Targets* 2006, **7(11)**:1443-1454.

50. Raynaud FI, Eccles S, Clarke PA, Hayes A, Nutley B, Alix S, Henley A, Di-Stefano F, Ahmad Z, Guillard S, Bjerke LM, Kelland L, Valenti M, Patterson L, Gowan S, de Haven Brandon A, Hayakawa M, Kaizawa H, Koizumi T, Ohishi T, Patel S, Saghir N, Parker P, Waterfield M, Workman P: **Pharmacologic characterization of a potent inhibitor of class I phosphatidylinositide 3-kinases.** *Cancer Res* 2007, **67(12)**:5840-5850.

51. Yang Y, Chen H, Nilsson I, Muresan S, Engkvist O: **Investigation of the relationship between topology and selectivity for druglike molecules.** *J Med Chem* 2010, **53(21)**:7709-7714.

52. Peters JU, Schnider P, Mattei P, Kansy M: **Pharmacological promiscuity: dependence on compound properties and target specificity in a set of recent Roche compounds.** *Chem Med Chem* 2009, **4(4)**:680-686.

53. Azzaoui K, Hamon J, Faller B, Whitebread S, Jacoby E, Bender A, Jenkins JL, Urban L: **Modelling promiscuity based on in vitro safety pharmacology data.** *Chem Med Chem* 2007, **2(6)**:874-880.