ARTICLES
PUBLISHED ONLINE: 24 JANUARY 2012 | DOI: 10.1038/NCHEM.1243
nature chemistry

Quantifying the chemical beauty of drugs

G. Richard Bickerton[1], Gaia V. Paolini[2], Jérémy Besnard[1], Sorel Muresan[3] and Andrew L. Hopkins[1]*

Drug-likeness is a key consideration when selecting compounds during the early stages of drug discovery. However, evaluation of drug-likeness in absolute terms does not reflect adequately the whole spectrum of compound quality. More worryingly, widely used rules may inadvertently foster undesirable molecular property inflation as they permit the encroachment of rule-compliant compounds towards their boundaries. We propose a measure of drug-likeness based on the concept of desirability called the quantitative estimate of drug-likeness (QED). The empirical rationale of QED reflects the underlying distribution of molecular properties. QED is intuitive, transparent, straightforward to implement in many practical settings and allows compounds to be ranked by their relative merit. We extended the utility of QED by applying it to the problem of molecular target druggability assessment by prioritizing a large set of published bioactive compounds. The measure may also capture the abstract notion of aesthetics in medicinal chemistry.

The concept of drug-likeness provides useful guidelines for early-stage drug discovery[1,2]. Analysis of the observed distribution of some key physicochemical properties of approved drugs, including molecular mass (M_r), hydrophobicity and polarity, reveals that they occupy preferentially a relatively narrow range of possible values[3]. Compounds that fall within this range are described as 'drug-like'. This definition holds in the absence of any obvious structural similarity to an approved drug. It has been shown that the preferential selection of drug-like compounds increases the likelihood of surviving the well-documented high rates of attrition in drug discovery[4].

Drug-likeness can be rationalized by considering how simple physicochemical properties impact molecular behaviour *in vivo*, with particular respect to solubility, permeability, metabolic stability and transporter effects. Indeed, drug-likeness is often used as a proxy for oral bioavailability. However, drug-likeness provides a broad composite descriptor that implicitly captures several criteria, with bioavailability among the most prominent.

In practical terms, most commonly the assessment of drug-likeness is manifested as rules, the original and most well-known of which is Lipinski's Rule of Five (Ro5) (ref. 5). The rule states that a compound is more likely to exhibit poor absorption or permeation when two or more of the following physicochemical criteria are fulfilled: the M_r is greater than 500, the calculated octanol–water partition coefficient, log *P*, is greater than five, there are more than five hydrogen-bond donors or the number of hydrogen-bond acceptors (nitrogen and oxygen atoms) is greater than ten. The rule does not apply to substrates of biological transporters or natural products. Aside from its predictive power, the widespread adoption of the Ro5 as a guideline for compound evaluation can be attributed also to the fact that it is conceptually simple and straightforward to implement.

Lipinski's insight (that the great majority of orally absorbed drugs occupy a privileged area of molecular property space[5,6]) resulted in greater awareness of the importance of molecular properties in determining oral bioavailability. The rule inspired numerous refinements and investigations into the concept of drug-likeness: Ursu *et al.*[2] provides a comprehensive review of the area. The Ro5 is not without its critics[7], yet in detail the issues tend to be with its qualitative nature, or the focus on oral drug space, as opposed to drug-like thinking *per se*.

Paradoxically, since the publication of the seminal paper by Lipinski *et al.*[5] there appears to be a growing epidemic, which Hann has termed 'molecular obesity'[8], among new pharmacological compounds (Supplementary Fig. S1). Compounds with higher relative M_r and lipophilicity have a higher probability of attrition at each stage of clinical development[4,9–11]. Thus, the inflation of physicochemical properties that increases the risks associated with clinical development may explain, in part, the decline in productivity of small-molecule drug discovery over the past two decades[4]. However, the mean molecular properties of new pharmacological compounds are still considered Lipinski compliant, even though their property distributions are far from historical norms.

Although the Ro5 is predictive of oral bioavailability, 16% of oral drugs violate at least one of the criteria and 6% fail two or more (although this does include natural products and substrates of transporters) (Supplementary Fig. S2a and Supplementary Table S1). High-profile drugs, such as atorvastatin (Lipitor) and montelukast (Singulair), fail more than one of the Lipinski rules (Supplementary Fig. S2b). Despite Lipinski's recommendation that the rule be considered as a guideline, in reality it is used routinely to filter libraries of compounds. The implementation of rules as filters means that no discrimination is achieved beyond a qualitative pass or fail—all compounds that comply with the rules are considered equal, as are all that breach them.

The response to such issues is not to define more refined rules. Instead, methods to quantify drug-likeness are required[12–14]. However, scoring schemes proposed to date, often derived by machine-learning methods, lack the intuitiveness, transparency and ease of implementation of the Ro5. To quantify compound quality we applied the concept of desirability[15] to provide a quantitative metric for assessing drug-likeness, which we call the quantitative estimate of drug-likeness (QED). QED values can range from zero (all properties unfavourable) to one (all properties favourable). The desirability approach can be used to generate functions to describe any set of compounds, depending on requirements. Here we demonstrate the utility of the approach by describing desirability functions derived from a set of orally absorbed approved drugs.

Desirability provides a simple, yet powerful, approach to multi-criteria optimization. It is increasingly utilized in a number of applications in drug discovery, including compound selection[16], library

[1]Division of Biological Chemistry and Drug Discovery, College of Life Sciences, University of Dundee, Dundee DD1 5EH, UK, [2]Gaia Paolini Ltd, 29 High Street, Bridge, Canterbury CT4 5JZ, UK, [3]DECS Computational Compound Sciences, Computational Chemistry, AstraZeneca R&D Mölndal, S-431 83 Mölndal, Sweden. *e-mail: a.hopkins@dundee.ac.uk

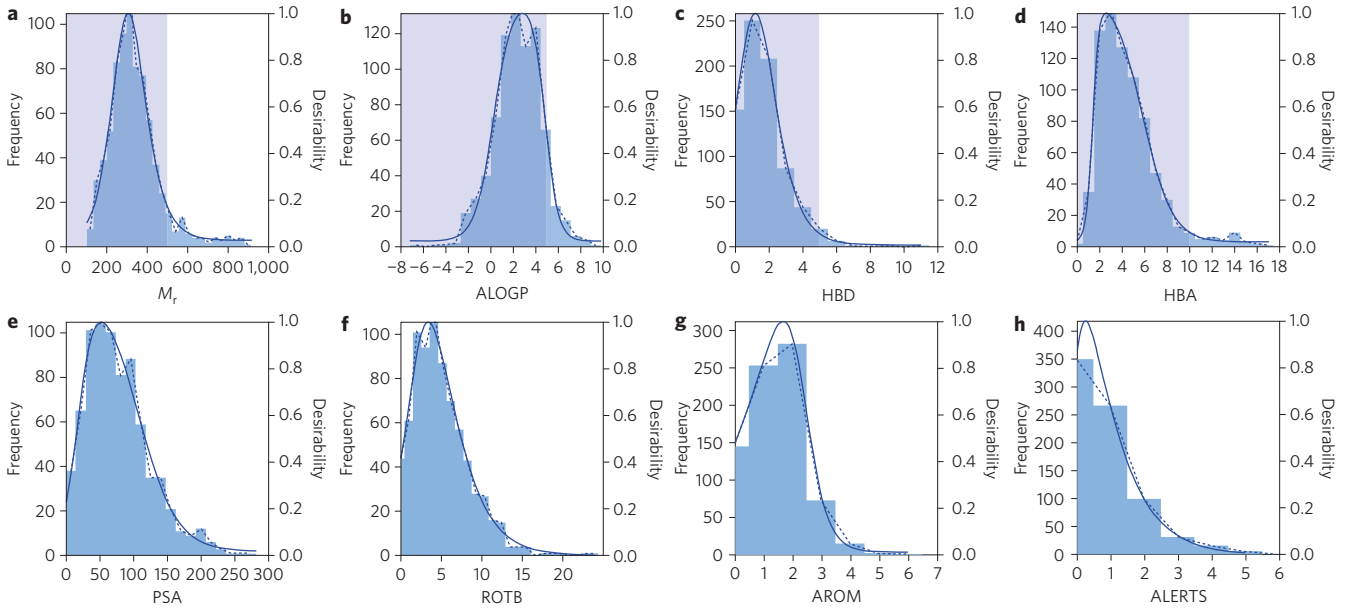

Figure 1 | Histograms of eight selected molecular properties for a set of 771 orally absorbed small molecule drugs. a–h, Molecular properties M_r (**a**), lipophilicity estimated by atom-based prediction of ALOGP (**b**), number of HBDs (**c**), number of HBAs (**d**), PSA (**e**), number of ROTBs (**f**), number of AROMs (**g**) and number of ALERTS (**h**). The Lipinski-compliant areas are shown in pale blue in (**a**), (**b**), (**c**) and (**d**). The solid blue lines describe the ADS functions (equation (2)) used to model the histograms. The parameters for each function are given in Supplementary Table S1.

design[17,18], prioritization of molecular targets, penetration of the central nervous system[19] and estimating the reliability of screening data[20]. The concept was introduced originally by Harrington[15] in the area of process engineering and further refined by Derringer and Suich[21]. Desirability takes multiple numerical or categorical parameters measured on different scales and describes each by an individual desirability function. These are then integrated into a single dimensionless score. In the case of compounds, a series of desirability functions (d) are derived, each of which corresponds to a different molecular descriptor. Combining the individual desirability functions into the QED is achieved by taking the geometric mean of the individual functions, as shown in equation (1):

$$\mathrm{QED} = \exp\left(\frac{1}{n}\sum_{i=1}^{n} \ln d_i\right) \quad (1)$$

Conventionally, desirability functions are defined arbitrarily, usually as monotonic decreasing or increasing functions, or 'hump' functions, at defined parameter ranges and inflection points. Importantly, whereas previous approaches used functions defined by user experiences and expectations[16,19], our approach differs fundamentally in that the functions are derived empirically by describing the underlying property distributions of a set of approved drugs, much as the boundaries defined by Lipinski are. The data used comprise a carefully curated collection of 771 orally dosed approved drugs. Eight widely used molecular properties were selected on the basis of published precedence for their relevance in determining drug-likeness[3,5,22,23]: M_r, octanol–water partition coefficient (ALOGP)[24], number of hydrogen bond donors (HBDs), number of hydrogen bond acceptors (HBAs), molecular polar surface area (PSA), number of rotatable bonds (ROTBs), number of aromatic rings (AROMs)[25,26] and number of structural alerts (ALERTS)[27]. The molecular properties were chosen on the basis that they have all been shown to influence the likelihood of attrition and can all be calculated robustly at high throughput. Histograms of the distribution of the eight molecular properties across the set of oral drugs are shown in Fig. 1. We found that the property distribution data were modelled consistently best as asymmetric double sigmoidal (ADS) functions, which are also shown in Fig. 1 over the same range. The general ADS function is shown in equation (2), where $d(x)$ is the desirability function for molecular descriptor x:

$$d(x) = a + \frac{b}{\left[1+\exp\left(-\frac{x-c+\frac{d}{2}}{e}\right)\right]}\left[1-\frac{1}{\left[1+\exp\left(-\frac{x-c-\frac{d}{2}}{f}\right)\right]}\right] \quad (2)$$

The parameters a, b, c, d, e and f for each of the ADS functions d_{Mr}, d_{ALOGP}, d_{HBD}, d_{HBA}, d_{PSA}, d_{ROTB}, d_{AROM} and d_{ALERTS} are shown in Supplementary Table S2, as are the coefficient of determination (R^2) and the rank among a library of non-linear functions.

Weighted desirability functions. The chosen molecular descriptors may vary in the importance of their contribution to drug-likeness, so each can be weighted by their relative significance. The unweighted QED shown in equation (1) is then replaced with a weighted QED ($\mathrm{QED_w}$), as shown in equation (3):

$$\mathrm{QED_w} = \exp\left(\frac{\sum_{i=1}^{n} w_i \ln d_i}{\sum_{i=1}^{n} w_i}\right) \quad (3)$$

where d is the individual desirability function, w is the weight applied to each function and n is the number of descriptors. Rather than assigning weights subjectively, we rationalized that the optimal set of weights is that which maximizes information content, which can be measured by calculating the Shannon entropy (SE)[28,29] (Supplementary Fig. S3a,b). An exhaustive search of possible weight combinations was performed for the set of approved drugs (see Methods). Three resulting sets of weights (W) were considered (Table 1): (i) the set of weights that gave the maximal information content ($\mathrm{QED_{w,max}}$), (ii) the mean weights

Table 1 | Optimized desirability function weightings by Shannon entropy.

	Shannon entropy	Rank	M_r	ALOGP	HBD	HBA	PSA	ROTB*	AROM*	ALERTS
$QED_{w,max}$	293.42	1	0.50	0.25	0.50	0.00	0.00	0.50	0.25	1.00
$QED_{w,mo}$	293.03	1–1000	0.66	0.46	0.61	0.05	0.06	0.65	0.48	0.95
$QED_{w,u}$	283.08	81,657	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

*The relatively high weightings of the number of AROMs and the number of ROTBs is consistent with a recent analysis from AstraZeneca, which found aromaticity and flatness to be important factors in attrition because of toxicology[52].

of the optimal 1,000 weight combinations that give the highest information content ($QED_{w,mo}$) and (iii) all weights as unity, that is unweighted ($QED_{w,u}$). Interestingly, the $QED_{w,max}$ series gave zero weight to the PSA and HBA parameters, which suggests that the information in these parameters is redundant. To help explain the relative weights we performed a principal component analysis of the unweighted desirability functions (Supplementary Fig. S3c). The results were consistent with the entropy analysis in that the least-correlated descriptors were weighted most highly. The pairwise cross-correlations between each of the properties is shown in Supplementary Fig. S3d and listed in Supplementary Table S4. The complete weighted QED is given in equation (4), where W is the weighting for each respective desirability function:

$$QED_w = \exp\left[\frac{\begin{array}{l}W_{MW}\ln d_{MW} + W_{ALOGP}\ln d_{ALOGP} \\ +W_{HBA}\ln d_{HBA} + W_{HBD}\ln d_{HBD} \\ +W_{PSA}\ln d_{PSA} + W_{ROTB}\ln d_{ROTB} \\ +W_{AROM}\ln d_{AROM} + W_{ALERTS}\ln d_{ALERTS}\end{array}}{\begin{array}{l}W_{MW} + W_{ALOGP} + W_{HBA} \\ +W_{HBD} + W_{PSA} + W_{ROTB} \\ +W_{AROM} + W_{ALERTS}\end{array}}\right] \quad (4)$$

The property descriptors considered, the weights given to these descriptors and the set of data that the functions are derived from can all be varied according to requirements. In this study we considered approved drugs that are dosed orally but, given appropriate data sets, desirability functions could be derived with relative ease to describe the relevant chemical space for parenteral administration, blood–brain barrier penetration[30] or taxonomic species with different permeability barriers.

Benchmarking. A benchmark study was designed to determine the relative performance of QED, the drug-like classifiers defined by Lipinski *et al.*[5], Veber *et al.*[23] and Ghose *et al.*[22] and the quantitative score of Gleeson *et al.*[31] in distinguishing a set of drugs from a background set of compounds. Assessment of whether a compound is objectively drug-like or otherwise is non-trivial and, as we have already argued, such a binary classification is somewhat misleading. With this consideration in mind, to assess the relative performance of rule-based classifiers and QED we attempted to benchmark their performance qualitatively. The DrugBank database[32] was used as the positive set and the small-molecule ligands of the Protein Data Bank (PDB) were used as the negative set.

The results of the benchmark study are shown in Fig. 2a and Supplementary Fig. S4 in the form of a receiver operating characteristic plot. QED outperforms the Ro5 and Ghose rules and performs marginally better than the Veber rule at a QED of 0.35 (the threshold at which Veber closely approaches that of QED). However, unlike rule-based approaches, this threshold could be modulated to give different levels of sensitivity and specificity according to requirements. The Ghose rule is less sensitive but more specific than the Veber and Lipinski rules. Interestingly, this benchmark suggests that the Veber rule outperforms the Ro5. $QED_{w,mo}$ and $QED_{w,u}$ outperform the quantitative measure of Gleeson regardless of threshold. $QED_{w,max}$ outperforms Gleeson above 0.37, below which it performs comparably. Performances of $QED_{w,max}$, $QED_{w,mo}$ and $QED_{w,u}$ are generally comparable, which suggests that the optimally weighted index ($QED_{w,max}$) can provide similar discrimination despite using fewer molecular descriptors. The best-performing measure alternates between $QED_{w,mo}$ and $QED_{w,u}$, depending on the range being considered, with $QED_{w,max}$ performing marginally worse. Given the somewhat artificial nature of the benchmark, in practical terms $QED_{w,mo}$ and $QED_{w,u}$ could be used interchangeably; only 18 (2.3%) of the DrugStore oral drugs have ΔQED (between $QED_{w,mo}$ and $QED_{w,u}$) of >0.15 and 100 drugs (13.0%) have ΔQED >0.10 (Supplementary Fig. S5a,b). Therefore, $QED_{w,mo}$ is used in all further analyses described here.

Direct comparison of the Ro5 and QED is illustrated in Fig 2b,c for the set of 771 oral drugs. An advantage of QED is its ability to rank compounds whether they fail the Ro5 or not. Interestingly, oral drugs that fail the Ro5 show QED values over a very wide range from nearly 0 to 0.8 (Fig. 2c). Figure 2d shows the differences in the distribution of QED scores for compounds in the ChEMBL database of small-molecule bioactivities[33], small-molecule ligands from the PDB and the set of oral drugs from DrugStore used to derive the functions. Such comparative analyses provide the means to establish the relative drug-likeness of any library of compounds.

Chemical aesthetics. As beauty is in the eye of the beholder, so chemical attractiveness is in the eye of the chemist[34–36]. A study that compared the ability of chemists to assess drug-likeness revealed that, although chemists agreed on the 'attractive' or drug-like structures, subjective human analysis is inconsistent in rejecting undesirable or 'ugly' compounds[35]. In an attempt to use chemists' collective experiences as a means to evaluate drug-likeness, Takaoka *et al.*[34] found the correlation coefficient between drug-like scores assigned by individual chemists to be 0.5–0.6. Lipinski has argued that pattern recognition is the forte of the chemist[37,38]. Wipke and Rogers describe the chemist's knowledge of chemical structures as a *Gestalt* pattern-recognition process[39]. Thus, we suggest that QED is an objective score that may correlate with the tacit knowledge of chemists' subjective assessment of drug-likeness or chemical attractiveness. The advantage of a codified metric on which chemical attractiveness can be judged is its application to ranking very large numbers of compounds. To aid the interpretation, it may be useful to consider QED values in the context of the observed distribution of a large reference set. To illustrate this, QED values that correspond to key percentiles from the ChEMBL database are shown in Supplementary Table S3 and a complete list is provided in the Supplementary Information.

Compared to the binary classification of the Ro5, QED exhibits a continuous scale from the most drug-like drugs (Fig. 3a) to the least drug-like (Fig. 3b). Comparison of the most drug-like drugs that fail the Ro5 (Fig. 3c) and the least drug-like drugs that pass the Ro5 (Fig. 3d) illustrates the potential of QED to rank compounds objectively by the elusive quality of chemical attractiveness.

To assess whether QED reflects chemists' opinions of chemical attractiveness, we compared QED with the manually assigned annotations for 17,117 diverse compounds scored by a survey of 79

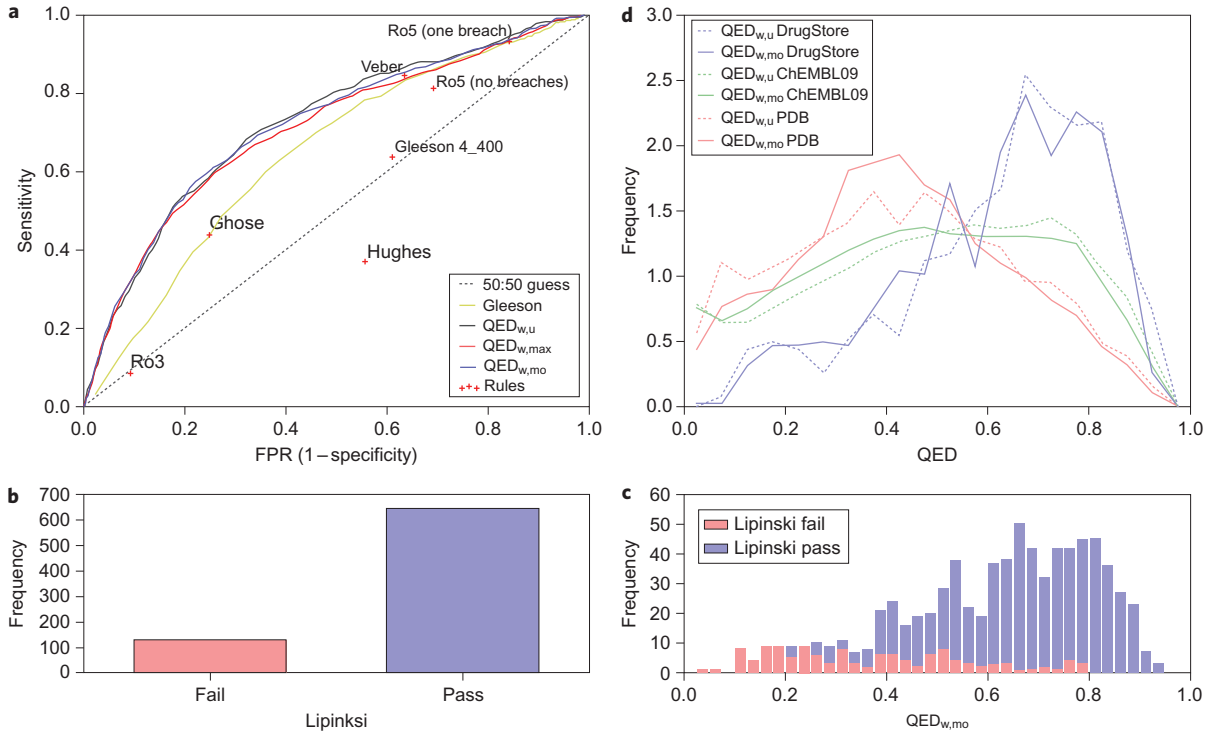

Figure 2 | Benchmarking of QED against other measures of drug-likeness. a, A receiver operating characteristic plot of true positive rate (sensitivity) against false positive rate (FPR (1 – specificity)) describes the difference in performance of different approaches in classifying compounds as drug-like or otherwise. The performances of the rules of Lipinski *et al.*[5], Veber *et al.*[23], Ghose *et al.*[22], Gleeson (4/400) (ref. 50), Congreve *et al.* (Rule of Three (Ro3))[51], Hughes *et al.*[9] and the quantitative method of Gleeson *et al.*[31] are compared to three different QED weighting schemes (maximal entropy ($QED_{w,max}$), mean optimal entropy ($QED_{w,mo}$) and unweighted ($QED_{w,u}$)). Veber *et al.* observed that compounds with fewer than ten ROTBs and PSA $\leq$140 Å^2 (or 12 or fewer hydrogen donors and acceptors) had an increased oral bioavailability in rats[23]. Ghose *et al.* suggested a qualifying range that could be used in the development of drug-like chemical libraries and recommended the following constraints: M_r between 160 and 480, calculated log *P* between −0.4 and 5.6, molar refractivity between 40 and 130 and total number of atoms between 20 and 70 (ref. 22). Gleeson *et al.* have proposed that the most desirable region for absorption, distribution, metabolism and excretion properties lies between $M_r < 400$ and ALOGP < 4 (ref. 50) and recently suggested a quantitative absorption, distribution, metabolism, excretion and toxicity score based on M_r and ALOGP (ref. 31). Hughes *et al.*[9] suggested clinical candidates had a lower likelihood of failing animal *in vivo* toleration toxicity if they had ALOGP < 3 and PSA > 75. For comparison, the Ro3 (ref. 51) for fragment selection is also plotted (where M_r <300, ALOGP $\leq$3, PSA $\leq$60, the number of HBDs $\leq$3 and the number of HBAs $\leq$3). At a threshold that provides an equivalent level of sensitivity as the Ro5, a $QED_{w,mo}$ of 0.40 offers 48% greater specificity than that of the Ro5. Equally, for the same degree of specificity as the Ro5, a $QED_{w,mo}$ of 0.26 offers 12% greater sensitivity. The dashed line represents the line of no discrimination – the level of performance that would be achieved by a random guess. **b**, Direct comparison of the Ro5 and QED shows the drugs failing (red) and passing (blue) Lipinski's Ro5. **c**, Equivalent plot of the QED results of the same set of compounds. The overlapping distributions indicate the greater resolution provided by the quantitative measure – some rather drug-like Lipinski failures are observed, as are some undrug-like passes. **d**, QED distribution for three small-molecule databases: the ChEMBL database of small-molecule bioactivities (green), small-molecule ligands from the PDB (red) and the set of oral drugs used to derive the functions (blue). Both weighted ($QED_{w,mo}$) (solid lines) and unweighted ($QED_{w,u}$) (dashed lines) indices are shown.

chemists from across AstraZeneca's chemistry community (see the Supplementary Information). Each chemist was asked to provide a 'yes' or 'no' answer to the question, 'Would you undertake chemistry on this compound if it were a hit?', for approximately 200 compounds each. Less than one-third of the compounds (5,457, 31.8%) were considered as attractive chemical starting points for hit optimization. Of 11,660 compounds that were considered unattractive, 4,497 (38.6%) were considered to be 'too complex' and 5,243 (45.0%) considered 'too simple' and no reason was assigned to the remainder. The mean QED is 0.67 (s.d. = 0.16) for the attractive compounds, 0.49 (s.d. = 0.23) for the unattractive compounds and 0.34 (s.d. = 0.24) for the unattractive compounds considered 'too complex' (Fig. 3e,f). The difference in QEDs between the attractive and unattractive compounds is statistically significant. The estimated difference in the medians of the attractive and unattractive compounds is 0.164 (Wilcoxon rank-sum test, 95% confidence interval 0.157–0.171). The equivalent value for the difference between the attractive and the 'too complex' set is 0.349 (95% confidence interval 0.340–0.358).

The chemical beauty of drug targets. A logical extension of the concept of compound drug-likeness is to apply it to the problem of target-druggability assessment. Hopkins and Groom[40] postulated that if there are physicochemical limitations to the properties of compounds that are likely to be oral drugs (as Lipinski proposed[5,6]), then drug-binding sites should have complementary properties. An implication of this idea is that not all ligand-binding sites have the appropriate physicochemical and topological properties to bind non-covalently small-molecule drugs with sufficient affinity. Binding sites that do have these characteristics are described as druggable (this definition is independent of any wider biological considerations). A number of algorithms were developed to determine the druggability of proteins based on analysis of the structural and physicochemical properties of an identified binding site[41–44]. A common feature of structure-based druggability analysis methods is the classification of a binding site into the categories of druggable or undruggable based on predefined training data.

Much as we have argued for the benefits of considering drug-likeness in quantitative terms, the druggability of a protein can

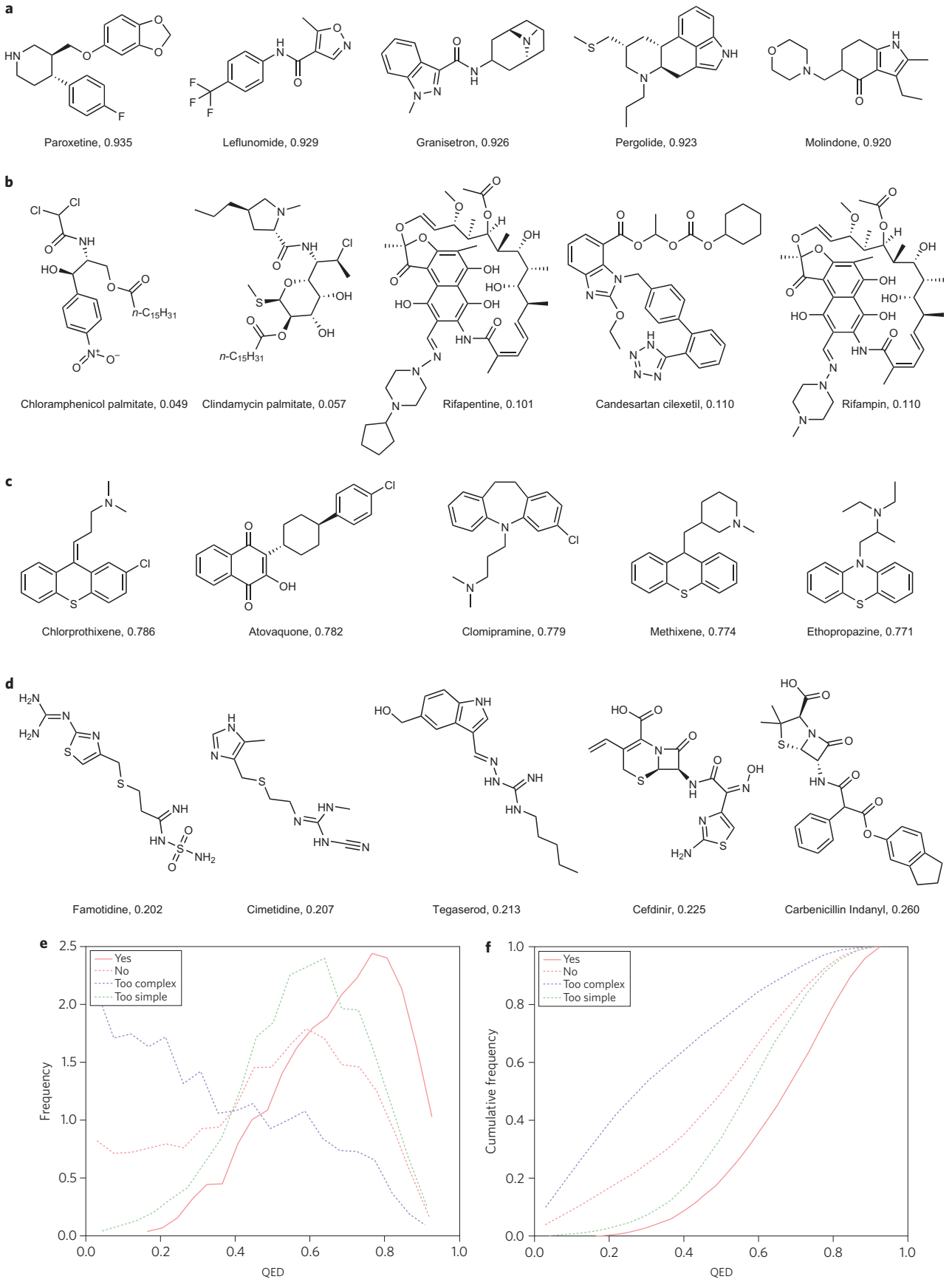

Figure 3 | Chemical aesthetics. a–d, Illustrative subsets of the oral drugs from DrugStore: the five most drug-like drugs (**a**), the five least drug-like drugs (**b**), the five most drug-like Ro5 failures (**c**), the five least drug-like Ro5 passes (**d**) (also see Supplementary Fig. S8). **e**, Results of chemical survey: QED distributions between compounds annotated chemically attractive and unattractive. **f**, Cumulative QED distribution of chemical survey results.

Table 2 | Top human targets* by three different ranking schemes.

	Target (UniProt)	Mean QED	Target (UniProt)	Mean QED best cluster	Target (UniProt)	Proportion clusters with mean QED >0.796
1	Free fatty acid receptor 2 (O15552)	0.861	Neuropeptide Y receptor type 5 (Q15761)	0.935	Vesicular acetylcholine transporter (Q16572)	0.714
2	Sodium channel (Q9NY72, O60939, Q8IWT1, Q07699)	0.849	Serotonin transporter (P31645)	0.932	Phosphodiesterase 7A (Q13946)	0.667
3	Voltage-gated potassium channel (P15382, P51787)	0.835	Serotonin 1a (5-HT1a) receptor (P08908)	0.932	Melatonin receptor 1A (P48039)	0.5
4	Phosphodiesterase 9A (O76083)	0.820	Norepinephrine transporter (P23975)	0.932	Melatonin receptor 1B (P49286)	0.462
5	Aldo-keto-reductase family 1 member C3 (P42330)	0.812	Dopamine transporter (Q01959)	0.931	Norepinephrine transporter (P23975)	0.453
6	Cholinergic receptor, nicotinic, beta 1 (muscle) (Q8IZ46)	0.811	Histamine H1 receptor (P35367)	0.928	Histamine H4 receptor (Q9H3N8)	0.444
7	Sorbitol dehydrogenase (Q00796)	0.809	Dopamine D3 receptor (P35462)	0.927	Dopamine transporter (Q01959)	0.409
8	Sodium channel protein type IV alpha subunit (P35499)	0.809	Dopamine D4 receptor (P21917)	0.925	Serotonin 7 (5-HT7) receptor (P34969)	0.4
9	Endothelial lipase (Q9Y5X9)	0.804	Thromboxane-A synthase (P24557)	0.921	Neuronal acetylcholine receptor protein beta-4 subunit (P30926)	0.4
10	Vesicular acetylcholine transporter (Q16572)	0.798	Serotonin 2c (5-HT2c) receptor (P28335)	0.917	Neuronal acetylcholine receptor protein alpha-7 subunit (P36544)	0.4
11	Lipoxin A4 receptor (P25090)	0.794	Serotonin 2a (5-HT2a) receptor (P28223)	0.916	Hormone-sensitive lipase (Q05469)	0.4
12	Small chondroitin/dermatan sulfate proteoglycan (Q99645)	0.790	Aldose reductase (P15121)	0.912	Serotonin 5a (5-HT5a) receptor (P47898)	0.4
13	Hypoxia-inducible factor prolyl 4-hydroxylase (Q9NXG6)	0.789	Alkaline phosphatase, tissue-nonspecific isozyme (P05186)	0.912	Histamine H1 receptor (P35367)	0.381
14	Huntingtin (P42858)	0.788	Dipeptidyl peptidase IV (P27487)	0.911	Metabotropic glutamate receptor 5 (P41594)	0.375
15	DNA-directed RNA polymerase (Q9Y2Y1, Q9H1D9, P05423)	0.786	Metabotropic glutamate receptor 5 (P41594)	0.911	Carbonic anhydrase VA (P35218)	0.375
16	Amine oxidase, copper containing (Q16853)	0.786	Neuronal acetylcholine receptor protein alpha-7 subunit (P36544)	0.910	Quinone reductase 2 (P16083)	0.375
17	Nuclear factor of activated T-cells cytoplasmic (Q13469, O95644)	0.769	Butyrylcholinesterase (P06276)	0.904	Serotonin transporter (P31645)	0.369
18	Sodium channel protein type VII alpha subunit (Q01118)	0.766	HERG (Q12809)	0.903	Dipeptidyl peptidase II (Q9UHL4)	0.333
19	Alkaline phosphatase, tissue-nonspecific isozyme (P05186)	0.764	Serine/threonine-protein kinase RAF (P04049)	0.902	Serotonin 3a (5-HT3a) receptor (P46098)	0.333
20	Serotonin 1f (5-HT1f) receptor (P30939)	0.762	Adenosine A1 receptor (P30542)	0.901	Steryl-sulfatase precursor (P08842)	0.333

*Only targets with at least ten associated compounds are considered. Full data for the top-ranking targets are provided in Supplementary Tables S5–7.

also be considered as a continuum of chemical tractability[45] rather than as a simple binary categorical assignment, thereby enabling the prioritization of druggable binding sites. QED provides an efficient means to quantify and rank the druggability of targets according to the chemical attractiveness of their associated ligands. QEDs were calculated for each compound in the ChEMBL database[33] of published bioactivity (release ChEMBL09) with an affinity $<10\ \mu M$ for a defined human protein target. The resulting 167,045 compounds are associated with 1,729 human proteins.

Top-ranking targets by three different schemes are shown in Table 2. The first scheme involves ranking targets by the mean QED of their associated ligands (Table 2 and Supplementary Table S5). The mean QED for all targets in the list is 0.478. For the targets of approved drugs the mean QED is 0.492 and for the targets of approved oral drugs the mean QED is 0.539 (with an average standard deviation for a target of 0.231). Drug targets are, indeed, enriched towards the more highly desirable targets, with 70% of the drug targets found in the top 50% of the prioritized target list.

Within a set of ligands for a target, it is useful to consider the QED of distinct chemical series, as even targets perceived as being relatively intractable may have a small proportion of associated chemical matter that is drug-like and of potential interest. To approximate distinct chemical series, all-by-all Tanimoto similarity matrices (International Business Machines Internal Report, 1957) were calculated for each of the 1,729 human protein targets in ChEMBL. Compound similarity is represented as a network with the chemical series identified as distinct subgraphs within the network using a Tanimoto similarity threshold of 0.7. We defined a chemical series as a cluster comprising at least five compounds and an active chemical series as one in which the proportion of actives is at least 0.7 (with an activity threshold of $10\ \mu M$ or ligand efficiency of 0.3). The number of compounds, series and active series for all ChEMBL targets is listed in the Supplementary Information. Chemical similarity networks for four targets are shown in Fig. 4. The chemical network representation in Fig. 4 illustrates the presence of highly desirable chemotypes, even for some targets with low mean QED. The mean QED of the most drug-like active series for each target

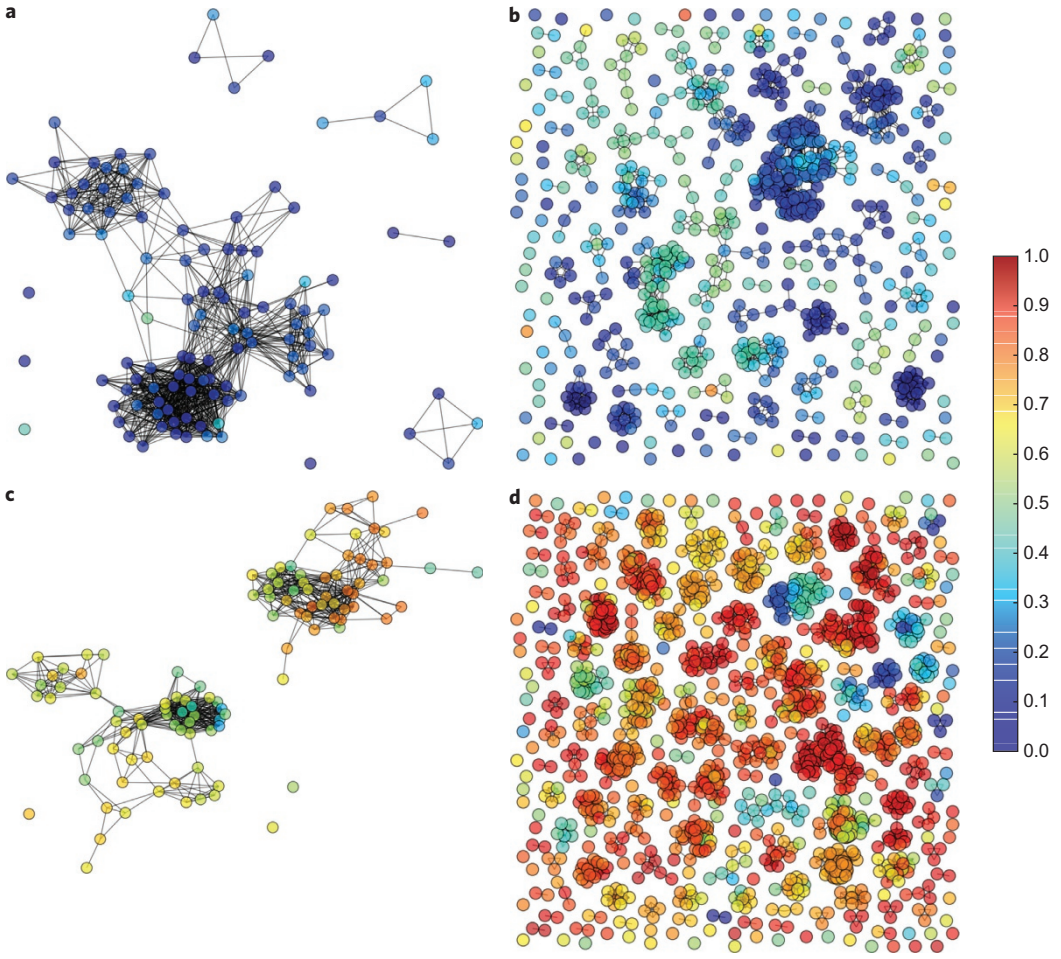

Figure 4 | Structural diversity networks. a, Structural diversity network for matriptase, a target for which the associated bioactive compounds are neither drug-like nor diverse. **b**, Structural diversity network for plasminogen, a target for which the published bioactive compounds are diverse, but not drug-like. **c**, Structural diversity network for 1-acylglycerol-3-phosphate *O*-acyltransferase beta, a target for which the published bioactive compounds are drug-like, but not diverse. **d**, Structural diversity network for norepinephrine transporter, a target for which the published bioactive compounds are both drug-like and diverse. In each of the networks compounds are represented as nodes and are coloured by their respective QED values. An edge connects nodes if they are structurally similar (defined by a Tanimoto threshold ≥ 0.7). The networks provide a useful way to summarize a large amount of data that describe the published bioactivity data for a target in an intuitive and visually digestible form. The four targets chosen each have a considerable number of associated compounds, but illustrate the importance of considering drug-likeness and chemical diversity when prioritizing targets. The network images are generated by the open-source graph visualization software GraphViz.

provides the second ranking scheme, listed in Table 2 and Supplementary Table S6. The mean QED of the best compound cluster of all ChEMBL targets is 0.569 (where the cluster comprises at least five compounds). The mean QED of the best cluster of human drug targets is 0.693. The mean QED when considering only the best cluster of the targets of oral drugs is 0.766.

A third approach to ranking targets is to consider the degree of enrichment of drug-like series. Here, targets are ranked by the proportion of active series that have a mean QED above that of the top 10% of the ChEMBL database (0.796) (Table 2 and Supplementary Table S7).

Conclusion

QED provides the means to rank chemical structures by their merit relative to target functions, which in this case are the properties of oral drugs. Furthermore, by extension of the concept to the set of ligands associated with a drug target, QED provides an efficient means to quantify and rank the druggability of targets. Lipinski's Ro5 gained considerable traction in early-stage drug discovery largely because it is predictive, intuitive and simple to implement. We believe QED compares favourably in each of these regards. Compared to the rule-based approaches, QED offers a richer, more nuanced view of drug-likeness. The QED functions are based on the underlying distribution data of drug properties and, unlike rule-based metrics, can identify cases in which a generally unfavourable property may be tolerated when the other parameters are close to ideal. In so doing the phenomenon of drug-likeness evolves from a binary 'black-and-white' assessment into a more realistic and gradated description of the continuum of compound quality.

Methods

Data set of known drugs. A non-redundant data set comprising 771 approved drugs was derived from the ChEMBL DrugStore database[46]. The selected compounds were all (i) marketed drugs, (ii) classified as therapeutics of small M_r (that is, no nutritional supplements, diagnostic agents or biologics), (iii) of specified molecular structure, (iv) composed of at least six atoms, (v) dependent on a biological macromolecule for their mode of action (that is, exclude chelators and buffers), (vi) orally administered and (vii) systemically absorbed (that is, exclude compounds for which the site of action is in the gastrointestinal tract; for example, orlistat targets gastric lipase and acarbose targets enteric alpha glucosidase).

Molecular properties. Physicochemical properties were calculated using the Pipeline Pilot Chemistry Collection (version 8.0.1.500) from Accelrys (San Diego, California). The properties calculated were M_r, ALOGP (using the atom-based

method by Ghose and Crippen[24]), number of HBDs, number of HBAs, molecular PSA, number of ROTBs and the number of AROMs[25,26]. Finally, a substructure search was performed against each drug using a curated reference set of 94 functional moieties that are potentially mutagenic, reactive or have unfavourable pharmacokinetic properties[27]. The number of matches for each compound was captured (ALERTS). We chose to omit the acid dissociation constant as the available high-throughput computational approaches do not provide sufficient accuracy[47].

Fitting of desirability functions. Histograms were plotted to reflect the distribution of each of the eight molecular properties for the oral drugs. For discrete variables (HBD, HBA, ROTB, AROM and ALERTS) a bin size of one was used. For continuous variables (M_r, ALOGP and PSA) the optimal bin size Δ was estimated by optimizing the cost function $C(\Delta)$ (ref. 48) (equation (5)):

$$C(\Delta) = \frac{2k - v}{\Delta^2} \tag{5}$$

where k and v are the mean and variance, respectively, of the occupancy of bins of size Δ. For PSA a local minimum $C(\Delta)$ was used. A library of functions was fitted to the distributions using TableCurve 2D version 5.01 (Systat Software, California). ADS functions (equation (2)) were found to be the most consistently high-ranking non-linear functions (Supplementary Table S1) and also reflected the important underlying asymmetry. Each function was then normalized by dividing by the maximum function value $d(x)_{max}$ to give a value between 0 and 1.

QED. The individual desirability functions were combined into the QED by taking the geometric mean, which, by logarithmic identities, can be expressed as the exponent of the arithmetic mean of the logarithm transformed identities (equations (1) and (3)).

Assignment of weights. We rationalized that the optimal set of weights is that which maximizes information content, as measured by SE (ref. 28) (equation (6)):

$$\mathrm{SE_w} = -\sum_{i=1}^{n} \mathrm{QED_w} \log_2 \mathrm{QED_w} \tag{6}$$

where $\mathrm{QED_w}$ is the weighted QED calculated with a set of weights w. Each possible combination of weights between 0 and 1 at increments of 0.25 were enumerated exhaustively for all eight molecular descriptors, giving 5^8 (390,625) weight combinations for each of the 771 drugs. The combination of weights that gives the highest entropy provides $\mathrm{QED_{w,max}}$ (Table 1). Inspection of the ranked weight combinations revealed a 'spike' of higher entropy values over the highest-scoring 1,000 combinations (Supplementary Fig. S3a). The mean of each individual molecular property weight over these 1,000 highest ranked entropy scores gives the mean optimal weighted $\mathrm{QED_{w,mo}}$ (Table 1). $\mathrm{QED_{w,mo}}$ may sample more accurately the high entropy combinations and attenuate the quantized nature of the weight increments. The robustness of this procedure was established by assessing the relationship of individual descriptor weights to the ranked entropy scores compared to a randomized series (Supplementary Fig. S3b).

PCA. PCA was performed on the eight unweighted desirability functions calculated on the ChEMBL database (release ChEMBL09) using Pipeline Pilot's R Statistics Component Collection (Supplementary Fig. S3c).

Benchmark study. The benchmarking assessment involved the assignment of positive and negative compound sets. The DrugBank database[32] was used to derive the positive set, and 771 compounds having the word 'oral' in their 'Route of Administration' field were selected. Although we endeavoured to obtain a truly independent positive set for the benchmark, inevitably significant overlap was found between the DrugBank set and the drugs used to derive QED. Of the 771 compounds, 554 were identical structurally and a further 30 had significant structural similarity (Tanimoto score >0.8). Small-molecule ligands from the PDB ligand dictionary[49] were selected as the negative set, as this provides, as well as drugs, a large and diverse source of chemical tools, metabolites, natural products and crystallographic buffers. To prevent ambiguity, 475 compounds that had significant structural similarity to the positive set (Tanimoto score >0.8) were removed, to leave a negative set of 10,250.

Performance measures. The performance measures used are given in equations (7)–(9):

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{7}$$

$$\text{Specificity} = \frac{TN}{FP + TN} \tag{8}$$

$$MCC = \frac{TP \times TN - FP \times FN}{((TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN))^{0.5}} \tag{9}$$

where MCC = the Mathews correlation coefficient, TP = true positives, TN = true negatives, FP = false positives and FN = false negatives.

Target druggability methods. The ChEMBL database includes a highly heterogeneous assortment of published bioactivity data. Bioactivity endpoints were considered only when (i) there was a defined protein molecular target, (ii) the activity type was half-maximum inhibitory concentration, inhibition constant or dissociation content, (iii) the relation was '=', '<' or '≤', (iv) standard units were defined as 'nM' and (v) the activity was greater than 10^{-6} nM (largely to remove misannotations). Typically, a broad range of bioactivity values are reported for a given combination of target and ligand through an amalgamation of biological, technical and annotation errors. Selection of the 'correct' value is non-trivial, particularly when using large-scale automated procedures. Simple calculation of a mean is sensitive to outliers. As such, for each combination of target and ligand we identified the modal log unit of bioactivity and calculated the mean value of activities within that range. Consideration of only human targets resulted in 167,045 unique compounds being associated with 1,729 proteins, to give 310,551 compound target pairs.

For each protein target the Tanimoto structural similarity of each associated compound to every other associated compound was calculated using Pipeline Pilot (FCFP_4 fingerprints) to give an all-against-all similarity matrix. Compound networks were derived from these matrices using the Python package NetworkX.

Received 1 September 2011; accepted 2 December 2011; published online 24 January 2012

References

1. Keller, T. H., Pichota, A. & Yin, Z. A practical view of 'druggability'. *Curr. Opin. Chem. Biol.* **10,** 357–361 (2006).
2. Ursu, O., Rayan, A., Goldblum, A. & Oprea, T. I. Understanding drug-likeness. *Wiley Interdis. Rev.: Comp. Mol. Sci.* **1,** doi: 10.1002/wcms.1052 (2011).
3. Oprea, T. I. Property distribution of drug-related chemical databases. *J. Comput. Aided Mol. Des.* **14,** 251–264 (2000).
4. Leeson, P. D. & Springthorpe, B. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nature Rev. Drug Discov.* **6,** 881–890 (2007).
5. Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Del. Rev.* **23,** 3–25 (1997).
6. Lipinski, C. A. Drug-like properties and the causes of poor solubility and poor permeability. *J. Pharmacol. Toxicol. Methods* **44,** 3–25 (2000).
7. Abad-Zapatero, C. A sorcerer's apprentice and The Rule of Five: from rule-of-thumb to commandment and beyond. *Drug Discov. Today* **12,** 995–997 (2007).
8. Hann, M. M. Molecular obesity, potency and other addictions in drug discovery. *MedChemComm* **2,** 349–355 (2011).
9. Hughes, J. D. *et al.* Physicochemical drug properties associated with *in vivo* toxicological outcomes. *Bioorg. Med. Chem. Lett.* **18,** 4872–4875 (2008).
10. Wenlock, M., Austin, R. P., Barton, P., Davis, A. M. & Leeson, P. D. A comparison of physiochemical property profiles of development and marketed oral drugs. *J. Med. Chem.* **46,** 1250–1256 (2003).
11. Proudfoot, J. R. The evolution of synthetic oral drug properties. *Bioorg. Med. Chem. Lett.* **15,** 1087–1090 (2005).
12. Xu, J. & Stevenson, J. Drug-like index: a new approach to measure drug-like compounds and their diversity. *J. Chem. Inf. Comput. Sci.* **40,** 1177–1187 (2000).
13. Rayan, A., Marcus, D. & Goldblum, A. Predicting oral druglikeness by iterative stochastic elimination. *J. Chem. Info. Model.* **50,** 437–445 (2010).
14. Ohno, K., Nagahara, Y., Tsunoyama, K. & Orita, M. Are there differences between launched drugs, clinical candidates, and commercially available compounds? *J. Chem. Inf. Model.* **50,** 815–821 (2010).
15. Harrington, E. C. Jr The desirability function. *Ind. Qual. Control.* **21,** 494–498 (1965).
16. Cruz-Monteagudo, M. *et al.* Desirability-based methods of multiobjective optimization and ranking for global QSAR studies. Filtering safe and potent drug candidates from combinatorial libraries. *J. Comb. Chem.* **10,** 897–913 (2008).
17. Le Bailly de Tilleghem, C., Beck, B., Boulanger, B. & Govaerts, B. A fast exchange algorithm for designing focused libraries in lead optimization. *J. Chem. Inf. Model.* **45,** 758–767 (2005).
18. Mandal, A., Johnson, K., Wu, C. F. J. & Bornemeier, D. Identifying promising compounds in drug discovery: genetic algorithms and some new statistical techniques *J. Chem. Inf. Model.* **47,** 981–988 (2007).
19. Wager, T. T., Hou, X., Verhoest, P. R. & Villalobos, A. Moving beyond rules: the development of a central nervous system multiparameter optimization (CNS MPO) approach to enable alignment of druglike properties. *ACS Chemical Neurosci.* **1,** 435–449 (2010).
20. Paolini, G. V., Lyons, R. & Laflin, P. How desirable are your IC_{50}s? A method to enhance screening-based decision making. *J. Biomol. Screen.* **15,** 1183–1193 (2010).
21. Derringer, G. & Suich, R. Simultaneous optimization of several response variables. *J. Qualty Technol.* **12,** 214–219 (1980).

22. Ghose, A. K., Viswanadhan, V. N. & Wendoloski, J. J. A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. *J. Comb. Chem.* **1,** 55–68 (1999).
23. Veber, D. F. *et al.* Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* **45,** 2615–2623 (2002).
24. Ghose, A. K. & Crippen, G. M. J. Atomic physicochemical parameters for three-dimensional structure-directed quantitative structure–activity relationships I. partition coefficients as a measure of hydrophobicity. *J. Comput. Chem.* **7,** 565–577 (1986).
25. Lovering, F., Bikker, J. & Humblet, C. Escape from flatland: increasing saturation as an approach to improving clinical success. *J. Med. Chem.* **52,** 6752–6756 (2009).
26. Ritchie, T. J. & Macdonald, S. J. The impact of aromatic ring count on compound developability – are too many aromatic rings a liability in drug design? *Drug Discov. Today* **14,** 1011–1120 (2009).
27. Brenk, R. *et al.* Lessons learnt from assembling screening libraries for drug discovery for neglected diseases. *ChemMedChem* **3,** 435–444 (2008).
28. Shannon, C. E. A mathematical theory of communication. *Bell System Technical J.* **27,** 379–423, 623–656 (1948).
29. Hosseinzadeh Lotfi, F. & Fallahnejad, R. Imprecise Shannon's entropy and multi attribute decision making. *Entropy* **12,** 53–62 (2010).
30. Wager, T. T. *et al.* Defining desirable central nervous system drug space through the alignment of molecular properties, *in vitro* ADME, and safety attributes. *ACS Chemical Neurosci.* **1,** 420–434 (2010).
31. Gleeson, M. P., Hersey, A., Montanari, D. & Overington, J. Probing the links between *in vitro* potency, ADMET and physicochemical parameters. *Nature Rev. Drug Discov.* **10,** 197–208 (2011).
32. Knox, C. *et al.* DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res.* **39,** D1035–D1041 (2011).
33. ChEMBL https://www.ebi.ac.uk/chembldb/
34. Takaoka, Y. *et al.* Development of a method for evaluating drug-likeness and ease of synthesis using a data set in which compounds are assigned scores based on chemists' intuition. *J. Chem. Inf. Comput. Sci.* **43,** 1269–1275 (2003).
35. Lajiness, M. S., Maggiora, G. M. & Shanmugasundaram, V. Assessment of the consistency of medicinal chemists in reviewing sets of compounds. *J. Med. Chem.* **47,** 4891–4896 (2004).
36. Muresan, S. & Sadowski, J. in *Molecular Drug Properties – Measurement and Prediction* (ed. Mannhold, R.) 441–457 (Wiley-VCH, 2008).
37. Lipinski, C. A. in *Molecular Informatics: Confronting Complexity* (eds Hicks, M. G. & Kettner, C.) (Beilstein-Institut, 2002).
38. Lipinski, C. A. Overview of hit to lead: the medicinal chemist's role from HTS retest to lead optimisation hand off. *Top. Med. Chem.* **5,** 1–24 (2009).
39. Wipke, W. T. & Rogers, D. Artificial intelligence in organic synthesis. SST: starting material selection strategies. An application of superstructure search. *J. Chem. Inf. Comput. Sci.* **24,** 71–81 (1984).
40. Hopkins, A. L. & Groom, C. R. The druggable genome. *Nature Rev. Drug Discov.* **1,** 727–730 (2002).
41. An, J., Totrov, M. & Abagyan, R. Comprehensive identification of 'druggable' protein ligand binding sites. *Genome Inform.* **15,** 31–41 (2004).
42. Cheng, A. C. *et al.* Structure-based maximal affinity model predicts small-molecule druggability. *Nature Biotechnol.* **25,** 71–75 (2007).
43. Halgren, T. A. Identifying and characterizing binding sites and assessing druggability. *J. Chem. Inf. Model.* **49,** 377–389 (2009).
44. Schmidtke, P. & Barril, X. Understanding and predicting druggability. A high-throughput method for detection of drug binding sites. *J. Med. Chem.* **53,** 5858–5867 (2010).
45. Southan, C., Boppana, K., Jagarlapudi, S. A. & Muresan, S. Analysis of *in vitro* bioactivity data extracted from drug discovery literature and patents: ranking 1,654 human protein targets by assayed compounds and molecular scaffolds. *J. Cheminform.* **3,** 14 (2011).
46. Overington, J. P., Al-Lazikani, B. & Hopkins, A. L. How many drug targets are there? *Nature Rev. Drug Discov.* **5,** 993–996 (2006).
47. Manchester, J., Walkup, G., Rivin, O. & You, Z. Evaluation of pK_a estimation methods on 211 druglike compounds. *J. Chem. Inf. Model.* **50,** 565–571 (2010).
48. Shimazaki, H. & Shinomoto, S. in *Advances in Neural Information Processing Systems* Vol. 19 (eds Schölkopf, B., Platt, J. & Hoffman, T.) 1289–1296 (MIT Press, 2007).
49. Dimitropoulos, D., Ionides, J. and Henrick, K. in *Current Protocols in Bioinformatics* (eds Baxevanis, A. D., Page, R. D. M., Petsko, G. A., Stein, L. D. & Stormo, G. D.) 14.13.11–14.13.13 (Wiley, 2006).
50. Gleeson, M. P. Generation of a set of simple, interpretable ADMET rules of thumb. *J. Med. Chem.* **51,** 817–834 (2008).
51. Congreve, M., Carr, R., Murray, C. & Jhoti, H. A 'rule of three' for fragment-based lead discovery? *Drug Discov. Today* **8,** 876–877 (2003).
52. Luker, T. *et al.* Strategies to improve *in vivo* toxicology outcomes for basic candidate drug molecules. *Bioorg. Med. Chem. Lett.* **21,** 5673–5679 (2011).

Acknowledgements

This research received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement N° 223461 and the Scottish Universities Life Sciences Alliance. We thank J. Overington for providing the DrugStore data, R. Brenk for the provision of SMARTS for structural alerts and I. Carruthers for assistance with DrugStore database queries. We also thank the chemistry community of AstraZeneca for participating in the chemistry survey.

Author contributions

A.L.H. conceived the approach and designed the algorithm. G.R.B. implemented the algorithm, performed the calculations, identified the functions and performed the analysis. G.V.P. developed the use of Shannon entropy as the weighting scheme, J.B. wrote the Pipeline Pilot implementation, S.M. coordinated the survey of AstraZeneca chemists, A.L.H and G.R.B. co-wrote the manuscript and G.R.B. produced the figures. All authors commented on the manuscript.

Additional information

The authors declare no competing financial interests. Supplementary information accompanies this paper at www.nature.com/naturechemistry. Reprints and permission information is available online at http://www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.L.H.