ELSEVIER

The graphical representation of ADME-related molecule properties for medicinal chemists

Timothy J. Ritchie[1], Peter Ertl[2] and Richard Lewis[2]

[1] TJR-Chem, Via Alberto 34C, 21020 Ranco (VA), Italy
[2] Novartis Institutes for BioMedical Research, Novartis Campus, CH-4056 Basel, Switzerland

The importance of striving for and maintaining drug-like physicochemical properties during the hit and lead optimization process is now well documented, and many published studies have suggested optimal ranges and/or limits for key molecule descriptors such as size, lipophilicity, H-bonding characteristics, rotatable bond and aromatic ring counts, particularly with regard to the design of orally administered drugs. The aim of this article is to review various approaches that have been used to represent molecule properties graphically in the context of oral 'drug likeness', with the goal of improving the decision making of medicinal chemists during the drug discovery process.

Importance of ADME-related physicochemical properties in medicinal chemistry

Ever since the late 1990s, when the link between the size, lipophilicity and H-bonding characteristics of drug molecules and their oral bioavailability was demonstrated [1–3], numerous studies have highlighted the importance of physicochemical properties in defining the drug likeness of molecules. From these studies, several 'rules of thumb' have been formulated to help guide medicinal chemists in the design and selection of molecules that should have an increased likelihood of becoming successful oral drugs. These rules were summarized in a recent publication [4].

Over the past few years, there has been some debate as to how strictly such rules are being, or should be, applied and whether in certain cases property limits can be breached [5]. It is apparent, however, that most drug discovery groups have implemented sets of rules or guides that are available to chemists during the hit and lead optimization process and that these rules are often followed strictly.

Visualization of scientific data

Data visualization (i.e. the graphical representation of data to assist viewers in gaining a better understanding of the underlying processes described by the data) is an indispensable part of scientific study. In the context of modern drug discovery, data visualized as an image, chart or plot can provide a simplified view of complex, multidimensional phenomena and ideally reveal correlations between different types of observations. Some visualization tools also offer the possibility to view the data interactively, which enables the researcher to gain better insight into the process under consideration. The visualization of scientific data is also commonly used in presentation graphics to facilitate the communication of key information and scientific results to the reader or viewer.

Many software packages exist, both commercially available and as freeware, that support the analysis and visualization of scientific data. These applications range from simple tools such as Microsoft Excel (http://www.office.microsoft.com/en-us/excel/), and other spreadsheet programs such as OpenOffice.org Calc (http://www.openoffice.org/product/calc.html) and Google Docs Spreadsheet (http://www.docs.google.com/support/bin/topic.py?topic=15115), to sophisticated data visualization tools such as Spotfire (http://www.spotfire.tibco.com/), Vortex (http://www.dotmatics.com/products_vortex.jsp), Miner3D (http://www.miner3d.com/) and JMP (http://www.jmp.com/index.shtml). The scope of this article, however, is not to review these in detail (information on such programs is available from the respective vendors' web sites).

Scope of this article

This overview focuses on how the above-mentioned ADME-related rules of thumb can be rendered and presented to a medicinal chemistry audience in an appropriate graphical representation, using – for example – colour, shape and size to facilitate the selection or deselection of molecules during the drug discovery

Corresponding authors:. Ritchie, T.J. (tim.j.ritchie@gmail.com), Ertl, P. (peter.ertl@novartis.com)

1359-6446/06/$ - see front matter doi:10.1016/j.drudis.2010.11.002

Drug	PSA	Mol Wt	logP	Rotatable bonds
captopril	57.6	217.3	-1.09	3
diprafenone	58.6	369.5	4.62	11
flunitrazepam	78.5	313.3	2.14	2
fluvoxamine	56.9	318.3	2.53	10
moclobemide	41.6	268.7	1.69	4
nafcillin	95.9	414.5	3.27	5
oxprenolol	50.7	265.4	2.07	9
terbutaline	72.7	225.3	1.07	4
tetrahydrocannabinol	29.5	314.5	6.69	4
urapidil	71.7	387.5	1.34	7

Drug Discovery Today

FIGURE 1

Representative drug property data with examples of simple colour and shape highlighting available in Microsoft Excel 2007: polar surface area (PSA) values as horizontal bars; molecular weight (Mol Wt) with graphical pies and grey-scale shading; log *P* with green–yellow–red colouring; and rotatable bond count with colouring if value is >8.

process. In many cases, such visualizations are readily implemented and presented to the chemist via web-based applications.

Simple cell highlighting

Human beings have an innate ability to quickly recognize and assimilate shapes, patterns and colours. This ability has been employed for thousands of years, long before the invention of language, written text and computers. It is known, for example, in engineering, that an analogue speedometer can be read much more easily than a digital version.

A simple example of how shape and colour can be employed to highlight aspects of numerical drug data is shown in Fig. 1. Whereas purely numerical data requires careful inspection to identify particular values, the use of shape (bars, pies), shading or colour can help to draw the viewer's attention to, in this case, high or low values.

Traffic light colouring

The familiar 'traffic light' colouring of red, amber and green was first introduced on street lights around 1920 and is now universally recognized around the globe. Because of its ubiquitous meaning, it is attractive to use the same colours to label molecule properties to indicate 'bad', 'intermediate' and 'good' values. Such traffic light colouring has been used to assist in the selection of HTS screening hits and subsequent leads, profiling of compound libraries and prioritizing compounds for purchasing [6]. As well as highlighting suboptimal properties with colour, the individual traffic light values are used to generate an 'Oral PhysChem Score' (Fig. 2) so large numbers of structures can be easily compared. Traffic light colouring has also been used to show the impact of higher molecular weight and/or *c* log *P* on a range of ADMET parameters [7].

Specialized plots

Optimization of a lead into a drug requires the simultaneous adjustment of several factors. As a result, mechanisms were invented to display bi-, tri- and multi-variate data in meaningful graphical representations. Although one might consider scatter plots and pie charts to be associated with modern presentation graphics and the computer age, it is interesting to note that these were, in fact, invented in the 19th century [8]. Several of these more specialized plots that have been found to be useful when applied to the display of molecule data in the drug discovery context are discussed in the next section.

Craig plots

One of the most important early examples of a graphical display of chemical properties was published by Craig in 1971 [9]. Analyses comparing several substituent constants revealed that the plotting of Hammett sigma and Hansch hydrophobicity pi descriptors produced a non-correlated scatter plot, which could serve as a guide for the synthesis of derivatives designed to cover wide ranges of values for these parameters. Craig plots are still very much in use: interaction with the display, for example, by showing the structures associated with points, enables the user to see into the process and hand select bioisosteric alternatives on a rational basis [10] (Fig. 3).

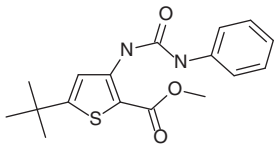
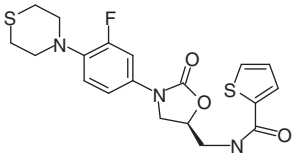
Flower plots

An early example of a graphical display of multiple molecular properties was published by Martin *et al.* [11] in 1995. In this case, 'flower plots' (Fig. 4) were generated to compare the molecular diversity of N-substituted glycine libraries with respect to 16 molecule properties, including lipophilicity, shape and branching, and chemical functionality. In essence, flower plots are bar graphs in which the *x*-axis has been wrapped in a circle. This enables the viewer to make rapid comparisons between objects. A modern interpretation of this radial plot approach has been implemented at Johnson & Johnson, where 'pie bar charts' are used to display selectivity data for compounds screened through a panel of biological target assays [12].

Egg plots

Using literature data on well-absorbed and poorly absorbed drugs, Egan *et al.* [13] created a general computational model for human passive intestinal absorption. This was graphically displayed as a bi-plot using calculated log *P* and polar surface area (PSA) as the *y*- and *x*-axes, respectively, overlaid with an ellipse (or egg) representing the property space occupied by the majority of well-absorbed compounds (Fig. 5). This provides a simple visual cue for profiling new compounds in terms of their potential to be orally absorbed. The direction of movement towards or away from the 'egg' can be used to guide the design of improved analogues. It also reminds the user about the probabilistic nature of the data, with the lighter oval on the periphery capturing 95% of the data.

TL Value	TL Solubility [mg L^{-1}]	TL CLOGP	TL Mwcorr	TL PSA [$Å^2$]	TL Rot Bonds
0	≥50	≤3	≤400	≤120	≤7
1	10-50	3-5	400-500	120-140	8-10
2	<10	>5	>500	>140	≥11

Structure	Oral PhysChem Score	TL Solubility [mg L^{-1}]	TL CLOGP	TL MWcorr	TL PSA [$Å^2$]	TL Rot Bonds
	4	<10	5.1	332.4	67.4	5
	2	10-50	2.6	407.7	61.9	5

Drug Discovery Today

FIGURE 2

In silico ADME traffic light (TL) colouring for several ADME-related properties [6]. The TL values for the five properties are summed to generate an overall 'Oral PhysChem Score'. Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission. *Abbreviations*: CLOGP, logarithm of the calculated partition coefficient; MWcorr, corrected molecular weight; PSA, polar surface area; Rot bonds, number of rotatable bonds.

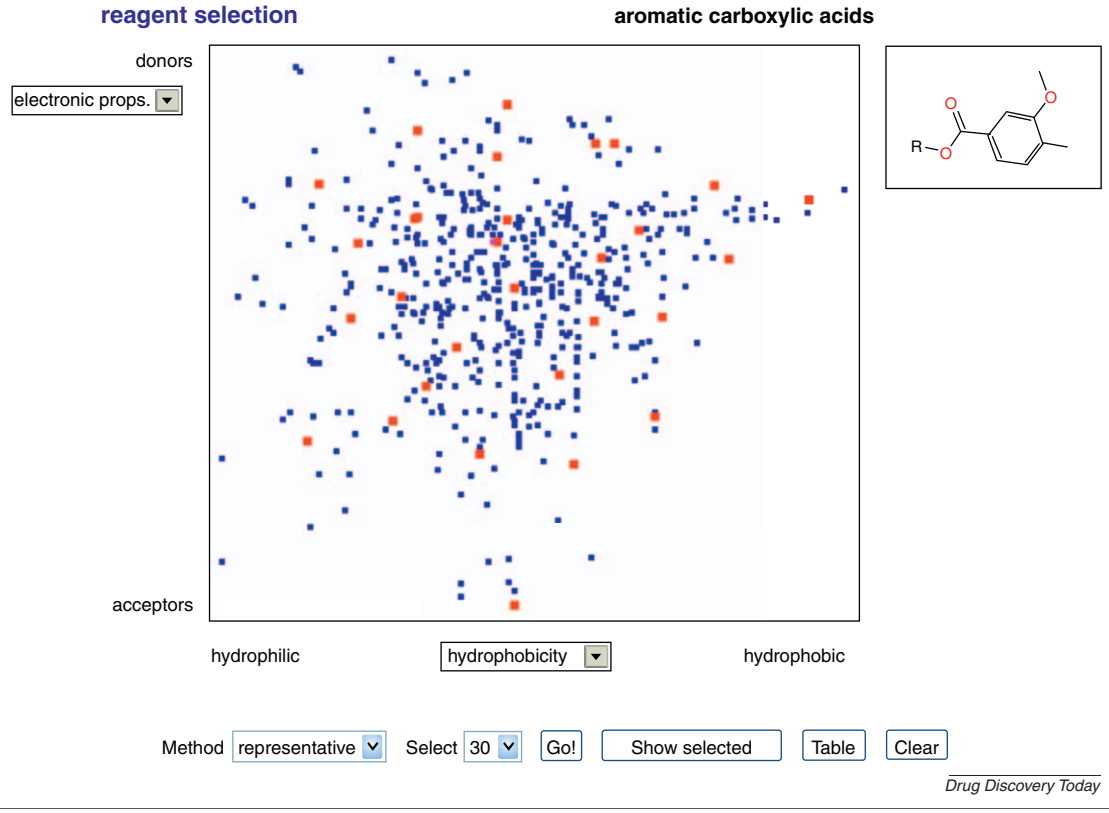


FIGURE 3

A web-based Craig plot implemented at Novartis enables interactive exploration of substituent property space with respect to hydrophobicity and electronic properties.

FIGURE 4

Flower plots showing 16 different side chain properties to display diversity within N-substituted glycine-based combinatorial libraries.Adapted, with permission, from Ref. [11]. Copyright 1995 American Chemical Society.

Oral bioavailability graphs

Another graphical *x*–*y* plot to assist in the estimation of oral bioavailability (%*F*) was published by Mandagere *et al.* [14]. In this case, metabolic stability and Caco-2 cell permeability data were plotted and areas of the graph defined to indicate low, medium or high probability of achieving oral bioavailability (Fig. 6). This model, which was validated with reference drugs, proved useful in the estimation of the bioavailability of compounds in several species based on their position on the map.

Golden Triangle

The Golden Triangle [15] is a visualization tool to help the simultaneous optimization of absorption and clearance of drugs. When plotting molecular weight versus distribution coefficient at pH 7.4 (log *D* 7.4) for a series of molecules, it is apparent that compounds with good permeability and low clearance are concentrated within a triangular shaped area, called the Golden Triangle (Fig. 7). The

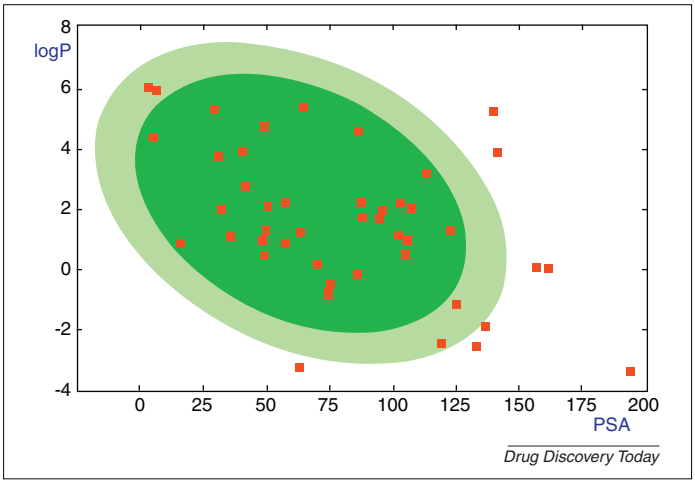


FIGURE 5

The green ellipse in the egg plot denotes a bioavailable region of property space with respect to log *P* and PSA; molecules outside this area might have bioavailability problems. Adapted, with permission, from Ref. [13]. Copyright 2000 American Chemical Society. *Abbreviations*: log *P*, logarithm of the calculated partition coefficient; PSA, polar surface area.

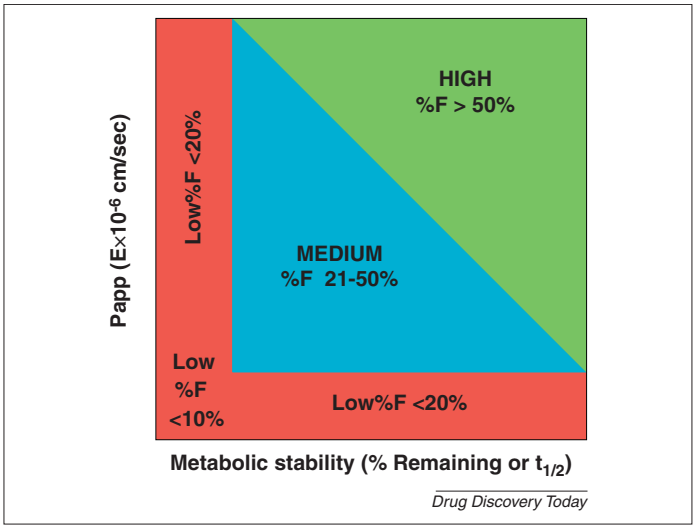


FIGURE 6

Graphical model to estimate oral bioavailability (*F*) using metabolic stability and cell permeability data. Adapted, with permission from, Ref. [14]. Copyright 2002 American Chemical Society. *Abbreviations*: Papp, apparent permeability coefficient; $t_{1/2}$, half-life.

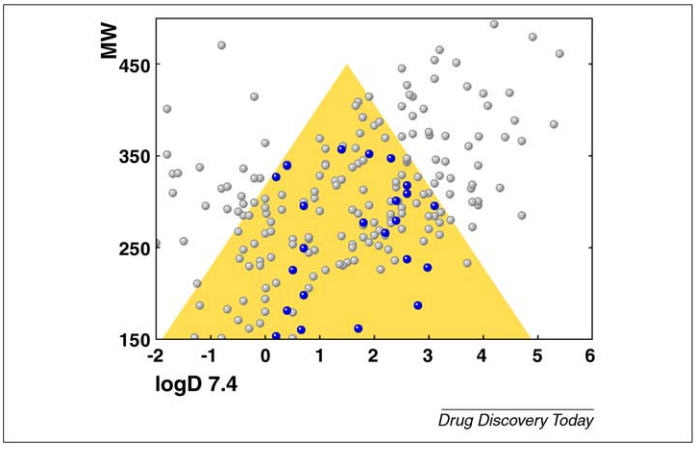


FIGURE 7

Set of project molecules with their log *D* and MW values. Molecules with good permeability (blue circles) are located in the Golden Triangle. Adapted from Ref. [15], with permission from Elsevier. *Abbreviations*: MW, molecular weight; log *D* 7.4, logarithm of the distribution coefficient at pH 7.4.

authors suggest using this simple visualization tool to guide the design of new molecules towards drug-like space.

Face diagrams

An unusual way to visualize multiple molecular properties is to associate numerical property values with various facial features – such as the size of the nose or eyes, eye slant, or mouth curve – and display them as facial diagrams (so-called 'Chernoff faces'). This method is based on the ability of a human brain to readily recognize and compare human faces. A computer program, FACES, has been described [16] to display 11 physicochemical properties for a set of molecules encoded into such face diagrams. With proper encoding, this approach could even offer the possibility of representing molecules with bad ADME properties as sad faces and molecules with good properties as happy faces.

SARANEA

Lounkine *et al.* [17] recently presented an interesting approach for interactive visualization of structure–activity, structure–property and structure–selectivity relationships. The method integrates various SAR and structure–selectivity relationship analysis functions and uses a network-like similarity graph data structure for visualization. The authors named the program SARANEA, which combines 'SAR' and 'ARANEAE', the scientific designation of the order of spiders, because the resulting molecular networks resemble spider webs. The tool enables the systematic detection of activity and selectivity cliffs. The SARANEA Java Program is freely available for download.

Time series plots

In the drug discovery process, one often needs to analyse changes of ADME properties as a function of time (so-called 'time series'). Typical examples are the improvement of bioavailability, potency and so on over the course of a drug development project or the examination of general trends in properties of drugs depending on their year of launch [18]. The classical way to do such an analysis is simply to use one of the axes of an *x*/*y* plot as a time axis. Interactive computer graphics, however, offer more sophisticated ways to visualize time series. A nice example is the free Motion Chart tool from Google (http://www.code.google.com/apis/visualization/documentation/gallery/motionchart.html), which enables the exploration of several indicators over time. Currently, this type of tool is used mostly in economics and finance (e.g. to explore the development of share prices or sales), but interactive time series graphs also offer promising opportunities in supporting drug discovery projects. An example of such a visualization is shown in Fig. 8, a screenshot of an interactive graph that enables exploration of changes in log *P*, PSA and molecular weight for a series of molecules active on different targets depending on the year of publication. A slider at the bottom of the graph allows the

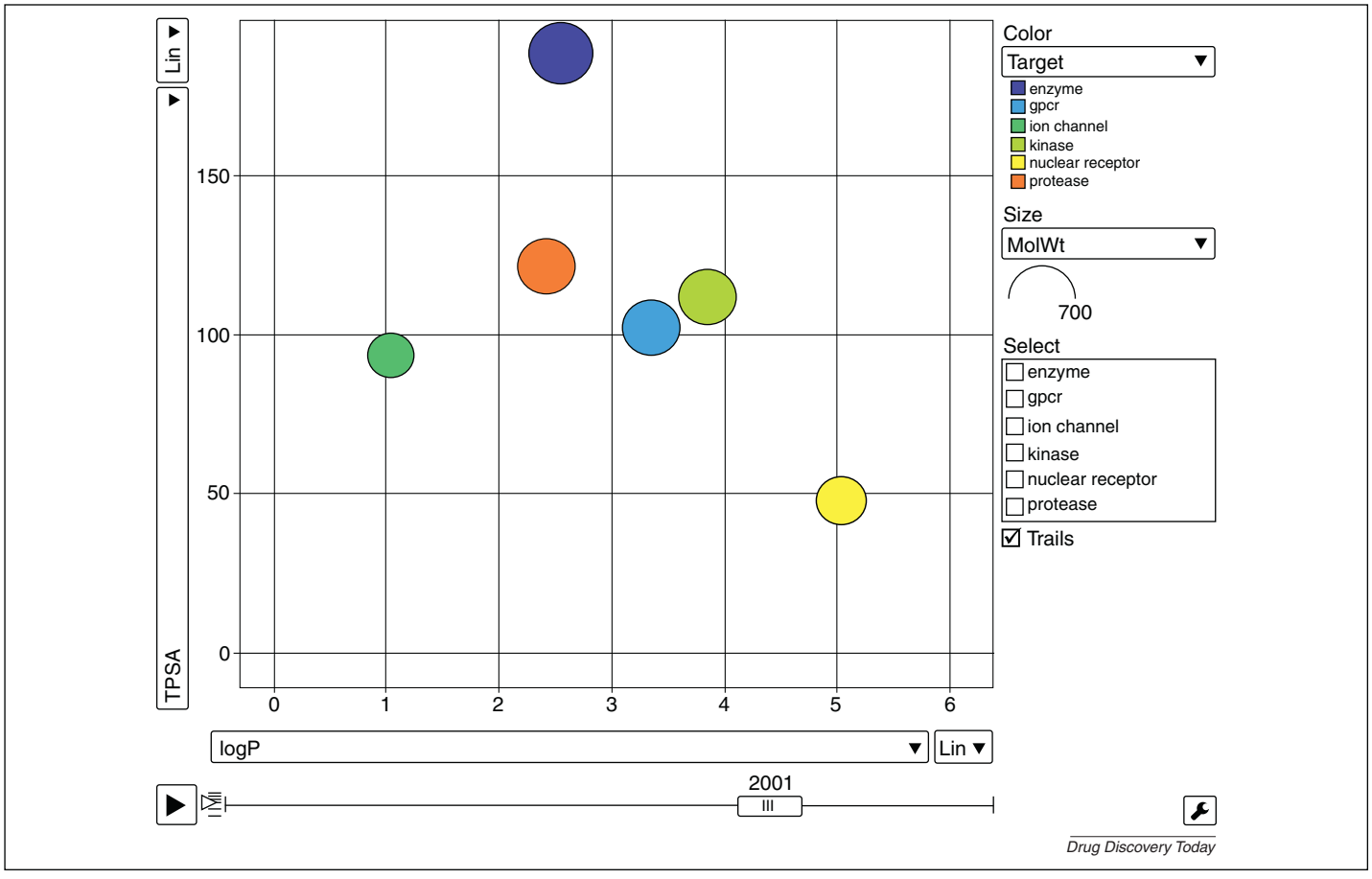


FIGURE 8

A screenshot of an interactive Motion Chart, which allows changes in important molecular properties depending on the target and publication year. *Abbreviations*: TPSA, topological polar surface area; log *P*, logarithm of the partition coefficient; Mol Wt, molecular weight.

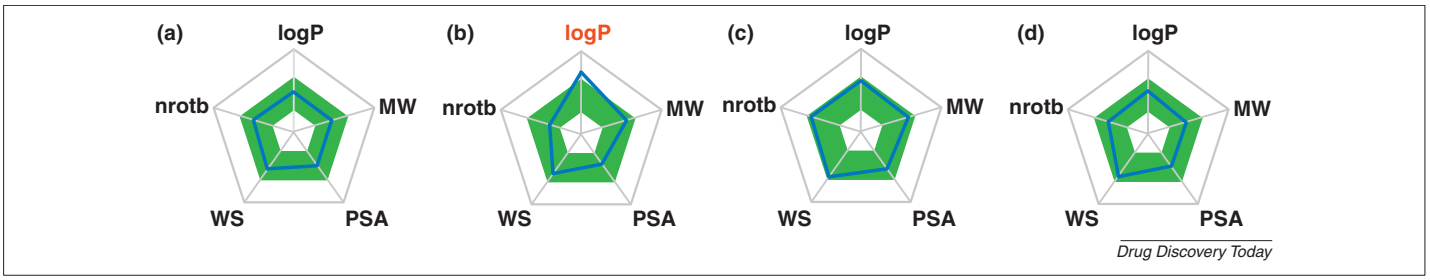


FIGURE 9

Radar plots of molecule properties. The green area represents 'good' property space for oral bioavailability, and the blue pentagon represents values of the five calculated properties for the molecule(s) in question. **(a)** The profile for an average oral drug. **(b)** The profile for a sub-optimal lipophilic molecule. **(c)** The average properties of compounds in the early-stage Novartis development pipeline during the mid-2000s. **(d)** The average properties of compounds in the late-stage Novartis development pipeline during the mid-2000s. *Abbreviations*: log *P*, logarithm of the calculated partition coefficient; Mol Wt, molecular weight; PSA, polar surface area; WS, water solubility score; nrotb, number of rotatable bonds.

view to be changed according to the publication year. Of course, this type of interactive chart should be manipulated in a 'live' environment, either as a web page or as a locally installed stand-alone program. More and more scientific journals – such as the *Journal of Cheminformatics* (http://www.jcheminf.com), the *Journal of Molecular Modeling* (http://www.springerlink.com/content/1610-2940/) and the *Internet Journal of Chemistry* (http://www.hackberry.trinity.edu/IJC/) now offer the possibility of including interactive tools as supporting information, to better illustrate the trends in data.

Bioavailability radar plots

Another useful mechanism to display the calculated phys-chem data of molecules in the context of oral drug-like property space is the 'radar' plot (sometimes called a 'spider' or 'cobweb' plot), available in various statistics software programs, including MS Excel software, and in more recent versions of Spotfire Decision Site.

Figure 9a–d show typical radar plots that can be generated in this way using five molecular descriptors: calculated log *P*, molecular weight, PSA [19], number of rotatable bonds and an aqueous solubility score. The green area of the plot defines the 'oral drug-like' limits for the five properties, which are within the ranges of log *P* −0.7–5.0; molecular weight 150–500; PSA 20–130; number of rotatable bonds 0–9; water solubility score 1–3 (1, highest solubility; 5, lowest solubility). The calculated values for the compound(s) being analysed are displayed as a blue pentagon, which should lie within the green area. Although the initial radar plots were generated as Excel charts, a Java-based web version was quickly implemented to take advantage of the online physicochemical property calculation tools available at Novartis [20,21].

Figure 9a shows a typical radar plot for an oral drug-like molecule (i.e. the blue pentagon lies well within the green area). This type of profile is also obtained by taking the average property values from marketed oral drug data sets (such as reported by Vieth [22]) or from other collections such as the Comprehensive Medicinal Chemistry database from Symyx (http://www.symyx.com/products/knowledge/medicinal_chem/index_print.jsp). By contrast, Fig. 9b is the profile obtained from a higher molecular weight, lipophilic molecule, which has exceeded the log *P* limit and is flagged as such in red.

Figure 9c and d describe the average properties of compounds in the Novartis early- and late-stage development pipeline, respectively, from the mid-2000s. As suggested by Wenlock [23], development candidates with molecular weights and log *P* values higher than those of marketed oral drugs tend not to survive to later clinical stages. The radar plots confirm that compounds that have reached phase II and beyond (Fig. 9d) do seem to have a profile that is closer to that of marketed oral drugs (Fig. 9a), when compared to compounds in phase I and preclinical phases (Fig. 9c).

The radar plot concept can also be used to profile other sets of compounds, such as combinatorial libraries. Figure 10a and b show two Novartis libraries of 3000 compounds each. In this case, the standard deviation from the mean of the properties is indicated by a dotted line, either side of the average values. This gives some measure of diversity in the libraries with respect to the five properties. The 'bad' library (Fig. 10b) was prepared several years ago, when it was very common for pharmaceutical companies to generate high molecular weight, lipophilic combinatorial libraries, before the impact of molecular properties on drug likeness was fully realized. The 'good' plot (Fig. 10a) is an example of a more recent library, which does not exceed the parameter limits and seems more diverse in terms of the standard deviation. Using this approach, virtual libraries or vendor collections of any size can be evaluated.

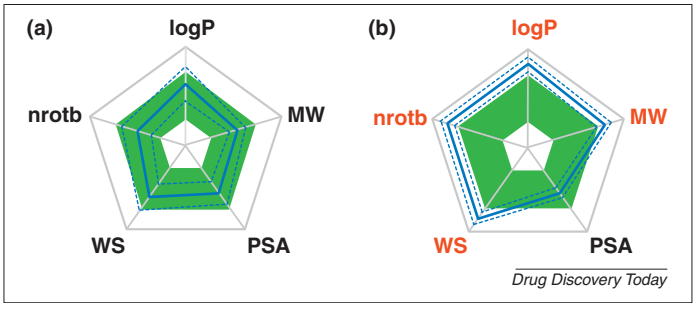


FIGURE 10

Radar plots of two 3000-member Novartis combichem libraries. The blue pentagon represents the mean of the five calculated properties for the data set. The dotted lines define the standard deviation in property values. **(a)** Is a more recent library with better oral drug properties. *Abbreviations*: log *P*, logarithm of the calculated partition coefficient; Mol Wt, molecular weight; PSA, polar surface area; WS, water solubility score; nrotb, number of rotatable bonds.

The examples above describe sets of 'real' compounds, but the radar plot can be equally useful in the design of new drug molecules. Thus radar plots generated from single virtual molecules or sets of molecules have been used by medicinal chemists in Novartis to assist the design process by answering questions such as 'Does this compound possess an oral drug profile, and if not, which properties are sub-optimal?', 'What effect does this new substituent have on the overall radar plot?', 'How does this scaffold change effect the lipophilicity and water solubility?', and so on. Implementation of the bioavailability plots within the Novartis '*In Silico* Profiling' web tool [20] enables easy generation of the plots together with calculation of several important molecular physicochemical properties and drug transport characteristics. In this way, chemists can compare compounds within series and link complex SAR with a graphical view of bioavailability to aid in the design of better compounds.

Although this implementation of the radar plot uses five particular parameters, there is no reason why other combinations of physicochemical properties cannot be used if they are deemed relevant for oral drug likeness; for example, *c* log *D* could replace *c* log *P* or H-bond donor and acceptor counts could be used instead of PSA (see Fig. 11). One must be aware, however, of properties that describe the same thing: a plot using both PSA and H-bond donors and/or acceptors would make no sense because these are strongly correlated.

Furthermore, other calculated ADME parameters such as Caco-2 permeability or metabolic stability could be used if sufficiently robust algorithms are available to do this. One could also consider selecting more stringent limits for the physicochemical properties to generate a 'lead-like' radar plot or a plot directed towards identifying compounds that have the ability to cross the blood–brain barrier.

The decision to use five parameters in the plot (rather than any other number) was a pragmatic one so that a reasonable amount of information could be presented and still be assimilated easily. One could, of course, generate a rule-of-five radar plot, with four axes (*c* log *P*, molecular weight, H-bond donors and H-bond acceptors), using the appropriate limits for each [3].

A modified version of the Novartis radar plot above has been implemented at GlaxoSmithKline and is deployed on the web via the Molecular Operating Environment web application environment (http://www.chemcomp.com/software-med.htm) and as a custom visualization prototype in Tibco Spotfire. In this manifestation of the radar plot, the five descriptors used are molecular weight, aromatic ring count (shown recently to be important for developability [24]), calculated log *P*, and the Lipinski H-bond acceptor and donor counts. There is also an additional 'borderline' range of property values (coloured yellow) to indicate that the properties are approaching 'bad' property space, and the calculated descriptor values are included in the plot labels (Fig. 11).

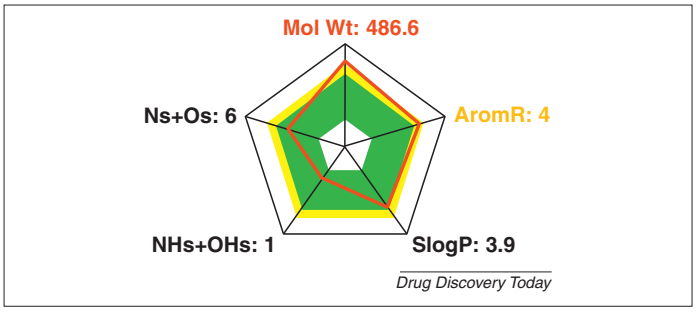


FIGURE 11

A radar plot of calculated physicochemical descriptors as implemented at GlaxoSmithKline. A yellow borderline range between the green acceptable region and unacceptable property space is included in this case. *Abbreviations*: Mol Wt, molecular weight; AromR, aromatic ring count; *S* log *P*, logarithm of the calculated partition coefficient; NHs + OHs, number of hydrogen bond donors; Ns + Os, number of hydrogen bond acceptors.

Molecule 'healthiness' using a traffic light pie chart

Traffic light-based nutritional labelling of foods has been implemented recently by supermarkets and other food suppliers to provide at-a-glance information to enable the selection of healthier food options if desired. An example is shown in Fig. 12 (left chart), indicating whether there are high (red), medium (amber) or low (green) levels of fat, sugar, salt and calories in the food. The adjacent figure shows how this approach can be applied to molecules to indicate their 'health' in terms of important physicochemical properties.

Caveats when using graphical depictions and classification schemes

In essence, the graphical representations described above strive to display multidimensional ADME-related data in a simple, readily assimilated way and help guide decision making in the drug discovery process. The desire for simplicity and clarity inevitably hides some of the shortcomings of the underlying calculations and assumptions, however, which must be borne in mind.

For example, classification schemes tend to deal in hard cut-offs (e.g. molecular weight ≤ 500). A compound with a molecular weight of 501 fails, as does a compound with a weight of 900, but clearly the former is much more likely to be amenable to structural modification to improve the situation. Thus some measure of flexibility around cut-offs (such as borderline ranges) can be more illustrative. In addition, calculations of some physicochemical properties such as log *P*, log *D* and aqueous solubility have inherent errors associated with them. An area for development would be the inclusion of some estimation of error and confidence in the displayed results.

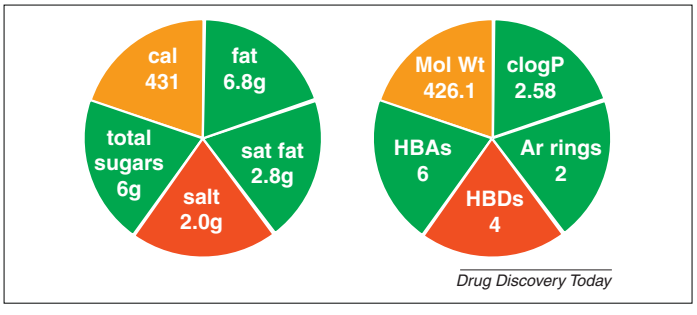


FIGURE 12

Food labelling to indicate nutritional content used on supermarket products (left) together with an analogous chart applied to molecule properties with traffic light colouring (right). *Abbreviations*: Mol Wt, molecular weight; *c* log *P*, logarithm of the calculated partition coefficient; Ar rings, aromatic ring count; HBDs, number of hydrogen bond donors; HBAs, number of hydrogen bond acceptors.

Another important point is that most current approaches produce static images, which must be regenerated if the input (i.e. a molecular structure) is modified. Dynamic displays that give immediate feedback on changes would be more useful for on-the-fly hypothesis generation.

Finally, it should be borne in mind that not all molecules that occupy oral drug-like property space are necessarily drugs, as pointed out by Kubinyi [25]. Compounds might possess molecular properties that should, for example, be compatible with good levels of oral bioavailability, but they might also be highly carcinogenic, have unacceptable off-target biological activity, or contain reactive or metabolically sensitive functionality that precludes their use as drugs.

Concluding remarks

As we have seen in this overview, the old adage that a picture is worth a thousand words is more than true for scientific visualization in drug discovery. As the importance of physicochemical and ADME-related properties in drug design has been increasingly recognized, both simple and more sophisticated graphing and visualization techniques have been applied to these data to assist medicinal chemists in designing and selecting molecules with optimal properties. Although many of the visualization methods described here are proprietary, there are some freely available resources to calculate ADME-related physicochemical parameters that have been summarized recently [26]. In addition, more interactive data analysis tools are being developed to help chemists better understand the relationships within the often complex multi-dimensional data. As the pharmaceutical industry increasingly focuses on finding ways to reduce attrition and increase the probability that drug candidates will become successful marketed drugs, it would seem that approaches to visualize ADME-related properties and summarize data in meaningful ways will continue to be developed.

References

1 Navia, M. and Chaturvedi, P. (1996) Design principles for orally bioavailable drugs. *Drug Discov. Today* 1, 179–189

2 Palm, K. *et al.* (1997) Polar molecular surface properties predict the intestinal absorption of drugs in humans. *Pharm. Res.* 14, 568–571

3 Lipinski, C.A. *et al.* (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 46, 3–26

4 Muchmore, S.W. *et al.* (2010) Cheminformatic tools for medicinal chemists. *J. Med. Chem.* 53, 4830–4841

5 Vieth, M. and Sutherland, J.J. (2006) Dependence of molecular properties on proteomic family for marketed oral drugs. *J. Med. Chem.* 49, 3451–3453

6 Lobell, M. *et al.* (2006) *In silico* ADMET traffic lights as a tool for the prioritization of HTS hits. *ChemMedChem* 1, 1229–1236

7 Gleeson, M.P. (2008) Generation of a set of simple, interpretable ADMET rules of thumb. *J. Med. Chem.* 51, 817–834

8 Friendly, M. (2009) A brief history of data visualization. In *Handbook of Computational Stastistics: Data Visualization III*. Springer-Verlag pp. 1–34

9 Craig, P.N. (1971) Interdependence between physical parameters and selection of substituent groups for correlation studies. *J. Med. Chem.* 14, 680–684

10 Langdon, S.R. *et al.* (2010) Bioisosteric replacement and scaffold hopping in lead generation and optimization. *Mol. Inf.* 29, 366–385

11 Martin, E.J. *et al.* (1995) Measuring diversity: experimental design of combinatorial libraries for drug discovery. *J. Med. Chem.* 38, 1431–1436

12 Kolpak, J. *et al.* (2009) Enhanced SAR maps: expanding the data rendering capabilities of a popular medicinal chemistry tool. *J. Chem. Inf. Model.* 49, 2221–2230

13 Egan, W.J. *et al.* (2000) Prediction of drug absorption using multivariate statistics. *J. Med. Chem.* 43, 3867–3877

14 Mandagere, A.K. *et al.* (2002) Graphical model for estimating oral bioavailability of drugs in humans and other species from their caco-2 permeability and *in vitro* liver enzyme metabolic stability rates. *J. Med. Chem.* 45, 304–311

15 Johnson, T.W. *et al.* (2009) Using the Golden Triangle to optimize clearance and oral absorption. *Bioorg. Med. Chem. Lett.* 19, 5560–5564

16 Larsen, R.D. (1986) FACES (features associated with chemical entities). *J. Chem. Educ.* 63, 1067–1068

17 Lounkine, E. *et al.* (2010) SARANEA: a freely available program to mine structure–activity and structure–selectivity relationship information in compound data sets. *J. Chem. Inf. Model.* 50, 68–78

18 Leeson, P.D. and Davis, A.M. (2004) Time-related differences in the physical property profiles of oral drugs. *J. Med. Chem.* 47, 6338–6348

19 Ertl, P. *et al.* (2006) Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its applications to the prediction of drug transport properties. *J. Med. Chem.* 43, 3714–3717

20 Ertl, P. *et al.* (2003) Web-based cheminformatics and molecular property prediction tools supporting drug design and development at Novartis. *SAR QSAR Environ. Res.* 14, 321–328

21 Ertl, P. *et al.* (2004) Web-based cheminformatics tools deployed via corporate intranets. *Drug Disc. Today: BIOSILICO* 2, 201–207

22 Vieth, M. *et al.* (2004) Characteristic physical properties and structural fragments of marketed oral drugs. *J. Med. Chem.* 47, 224–232

23 Wenlock, M.C. *et al.* (2003) A comparison of physicochemical property profiles of development and marketed oral drugs. *J. Med. Chem.* 46, 1250–1256

24 Ritchie, T.J. and Macdonald, S.J.F. (2009) The impact of aromatic ring count on compound developability – are too many aromatic rings a liability in drug design? *Drug Discov. Today* 14, 1011–1020

25 Kubinyi, H. (2003) Drug research: myths, hype and reality. *Nat. Rev. Drug Discov.* 2, 665–668

26 Ertl, P. and Jelfs, S. (2007) Designing drugs on the internet? Free web tools and services supporting medicinal chemistry. *Curr. Top. Med. Chem.* 7, 1491–1501