Available online at www.sciencedirect.com
ScienceDirect
European Journal of Medicinal Chemistry 44 (2009) 1128–1134
EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY
http://www.elsevier.com/locate/ejmech

Original article

Investigation of the incidence of ''undesirable'' molecular moieties for high-throughput screening compound libraries in marketed drug compounds

Peter Axerio-Cilies [a,b], Ivan P. Castañeda [a,b], Amin Mirza [a], Jóhannes Reynisson [c,*]

[a] *Cancer Research UK Centre for Cancer Therapeutics, Institute of Cancer Research, 15 Cotswold Road, Sutton, Surrey SM2 5NG, UK*
[b] *London Metropolitan University, Department of Health and Human Sciences, Holloway Road, London N7 8DB, UK*
[c] *Nottingham Trent University, School of Science and Technology, Clifton Lane, Nottingham NG11 8NS, UK*

Received 20 March 2008; received in revised form 9 June 2008; accepted 13 June 2008
Available online 26 June 2008

Abstract

A database of 1070 marketed drug compounds was compiled and analyzed in order to assess the occurrence of moieties described in the literature as ''undesirable'' for high-throughput screening compound libraries due to their ability to perturb assay formats. The study revealed a total of 277 compounds, 26% of the database, contained at least one of the moieties. As some of the drug compounds contained more than one ''undesirable'' moiety, the total number was 352. Electrophilic reactive groups, particularly aliphatic esters, were the most abundant type with 55% of the total. Half of the drug compounds incorporating the ''undesirable'' moieties were synthetic organic molecules. These findings suggest that ''undesirable'' moieties do not pose a major hindrance during clinical trials, the most expensive phase of drug development. In addition, their early elimination in the preclinical stage excludes large regions of known drug space due to the reliance on biochemical and cell-based assays. In general, it can be concluded that compounds with ''undesirable'' moieties should not simply be eliminated from compound screening libraries but rather be flagged as potentially problematic. A possible solution is to segregate the compounds containing suspect moieties and screen them when deemed appropriate.

Keywords: Drug chemical space; Toxic and promiscuous chemical moieties; Drug- and lead-like chemical space and drug discovery

1. Introduction

The identification of quality hit compounds is one of the key steps in a drug development programme. As high-throughput screening (HTS) is the main method to find hit and lead compounds, it is clear that well designed screening collections are more likely to yield tractable molecular entities [1–3]. Screening collections are now compiled using the principles of drug-like and lead-like chemical space. This is where compounds are filtered using various molecular descriptors, for example, according to Lipinski's rule of five [4,5] or based on the findings of Oprea [6–8]. This practice can easily be justified by the observation that commercially available compound collections tend to be lipophilic [9] and that the most lipophilic compounds are liable to be discontinued from development in clinical trials [10]. The design of screening collections is discussed in detail in review articles by Lajiness et al. [11], Egan et al. [12] and others.

Another approach used to define drug-like chemical space is to exclude chemical moieties which are known to generate false positives for a range of assay formats. For example, these can aggregate [13] or form covalent bonds with the protein target. Then there are other moieties which have been associated with toxicity issues such as DNA intercalation [14]. Numerous toxic and promiscuous chemical moieties have been identified and are documented in the literature [15–19]. They can be termed ''undesirable'' due to their potential to generate false positives. Medicinal chemists are often able to recognize these toxic molecular moieties and promiscuous inhibitors by inspecting the structural features or chemical

* Corresponding author.
E-mail address: johannes.reynisson@ntu.ac.uk (J. Reynisson).

0223-5234/$ - see front matter © 2008 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2008.06.013

properties of compounds [17]. However, experienced medicinal chemists can have differing opinions on whether these molecular species should be avoided or not, depending on their personal experience and disposition [20].

The use of molecular descriptors to eliminate compounds from the drug development process has lead to the definition of drug-like chemical space. In Fig. 1 it is represented graphically as a sphere in multi-dimensional space. ''Undesirable'' chemical moieties can also be used to further define drug-like chemical space. Whether, they should be shown as cut-offs or embedded as regions within chemical space is debatable. In the interests of clarity they are represented in the same manner as the molecular descriptors.

In this paper we investigated 1070 marketed drug compounds in order to establish how often the ''undesirable'' moieties occur in known drug space. We then established which types of moieties are predominant, the route of administration, and the actual function of the moieties. By working with marketed drug compounds, or known drug space, only molecular entities which have passed clinical trials and successfully marketed were examined. With this approach, we wished to establish whether excluding the ''undesirables'' from screening collections is justified or by doing so researchers might be lessening their chances of finding quality hit and lead compounds, which could be successfully developed into marketable drugs.

2. Methodology

2.1. Compilation of marketed drug library

A collection of marketed drugs was compiled from the DrugBank database [21], filtered for FDA (U.S. Food and Drug Administration) approved drugs. The following information was extracted; generic name, chemical structure, route of administration, therapeutic category, a brief description of the biological target, and mechanism of action when available. Additional databases, Drugs.com [22], RxList.com [23], and PubChem [24], were used to clarify controversial or incomplete information from the DrugBank database. Finally, a small number of salts and solvents were removed from the selection (e.g., calcium carbonate, ethanol and dimethyl sulphoxide) as they were not considered relevant to this study. A total of 1070 molecular entities were entered into Chemdraw for Excel (Cambridgesoft and Microsoft) for analysis.

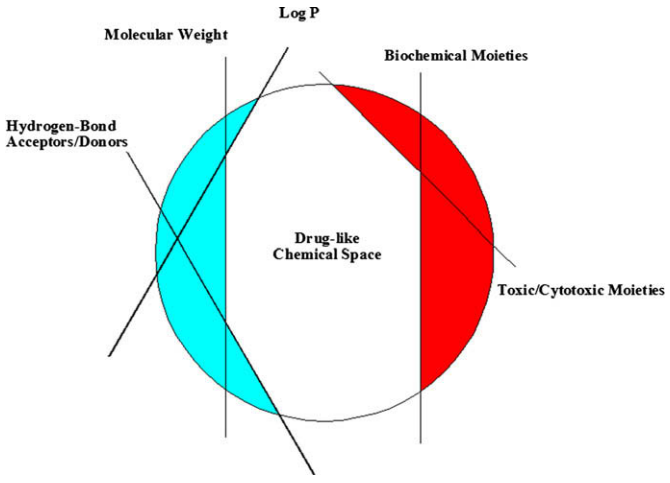

Fig. 1. Graphical representation of drug-like chemical space. It is defined with molecular descriptors (in this case MW, log *P*, and hydrogen bond donors and acceptors) and chemical moieties, which can produce false positives for either biochemical and/or cellular assay formats. This representation is a simplified depiction of the principle, e.g., many more molecular descriptors can be used.

2.2. Definition of the ''undesirables''

According to the literature, known ''undesirable'' moieties are categorised into six types depending on the mechanism by which they distort assay results. The complete list of the ''undesirable'' moieties is given in Supplementary information Table A. Categories 1–3 are known to mainly affect biochemical enzyme based assays whilst categories 4–6 predominantly affect cell-based assays.

2.2.1. Electrophilic reactive moieties

Compounds containing electrophilic functional groups are susceptible to decomposition through solvolysis or hydrolysis mechanisms and are known to react covalently with protein targets and biological nucleophiles [16]. Included in this category are electrophilic ''suicide inhibitors'', which contain moieties known to react with serine, threonine, and cysteine proteases under HTS assay conditions (e.g., α-haloketones, boronic acids, aldehydes, and 1,2-dicarbonyls) [16].

2.2.2. Heteroatom–heteroatom single bonds

Heteroatom–heteroatom single bonds have low bond dissociation energies [25]. These weak bonds are vulnerable to nucleophilic attack by proteins in biochemical assays resulting in false positives.

2.2.3. Common substructures of promiscuous non-specific inhibitors

These moieties act on unrelated targets. They are known to interfere with biochemical screening assays due to their ability to inhibit enzymes by aggregation [26]. They are sensitive to enzyme concentrations, ionic strength, albumin and detergents [13,19].

2.2.4. Tight-binding or metal chelating moieties and redox/thiol reactive species

Hydroxamate, oxime, and thiol chelators are known to react with metalloproteinases and other zinc binding mechanisms under HTS assay conditions [16]. This category includes redox/thiol reactive species like quinones and related derivatives. Under certain cellular conditions, quinones can behave as redox active molecules that are susceptible to one or two electron reduction [27]. Quinones can also act as Michael acceptors, triggering cellular damage through alkylation of thiol groups of glutathione, proteins and DNA [27].

2.2.5. Suspected cytotoxic moieties

The moieties that fall in this category are potentially cytotoxic due to their ability to modify proteins and nucleic acids, disrupting cellular function and/or causing mutations. The cytotoxic effect of aldehydes is caused by a reaction with sulfhydryl, imidazole, or amino groups to form adducts via a Michael addition mechanism [28]. Soft electrophiles like aliphatic ketones and cyclohexanones, can alkylate protein nucleophiles and cause cellular damage [29].

2.2.6. Moieties known to cause toxicity issues by DNA intercalation

Acridine and related derivatives bind to DNA by intercalation. These moieties have planar aromatic rings that can fit in between adjacent base pairs of the DNA structure [30]. They are known to be toxic since they can interact with the DNA in many types of tissue [14].

2.3. Substructure analysis

The 1070 drug molecules were visually screened for ''undesirable'' moieties as defined above. If one was present, it was considered a HIT. A full list of drugs containing HITs are given in Table B and without in Table C in Supplementary information. In an attempt to find some correlation between the moieties under study and their presence in drug molecules, the drugs were categorised depending on route of administration and therapeutic target. Seven classifications for the route of administration were used; oral, inhalation, intravenous (IV) injection, intravascular, intramuscular (IM) injection, topical and intraoccular (IO). Six categories were also defined for therapeutic targets; neurotransmission, central nervous system, cardiovascular system, hormonal system, immune system, and chemotherapeutic agents. Additionally, the year of approval for each compound was obtained from the FDA website [31]. This is a useful indication of whether the drugs were discovered prior to the emergence of HTS technology around 1990. Natural products, and their derivatives, were categorised in a special class since these compounds are often very complex and it is difficult to elucidate what purpose certain moieties serve. Finally, the moieties present in each of the drugs were individually assessed by means of structure–activity relationships and categorised according to their main function (moiety function). The analytical process is depicted as a flow chart in Fig. 2.

3. Results

3.1. Distribution according to mechanism

Of the 1070 drug compounds assembled from the public database, 277 (26% of the collection) contained at least one of the moieties discussed. Sixty-one of these drug molecules had more than one ''undesirable'' moiety. In some cases, three ''undesirable'' moieties are incorporated, e.g., sirolimus and tacrolimus. Valrubicin, a natural product derivate contains an exceptional four ''undesirable'' moieties. As a result a total of 352 HITs were recorded [32]. Electrophilic reactive moieties account for more than 50% of the HITs found in the collection. This is depicted in the distribution of ''undesirable'' moiety type shown in Fig. 3. The suspected cytotoxic moieties account for ~24%. Tight-binding or metal chelating moieties and redox/thiol reactive species constitute ~10% of the total HITs. Heteroatom–heteroatom bond-containing moieties and promiscuous inhibitors contribute 5% each, followed by the DNA intercalators with less than 2%. According to Rishton [16], compounds containing electrophilic functional groups

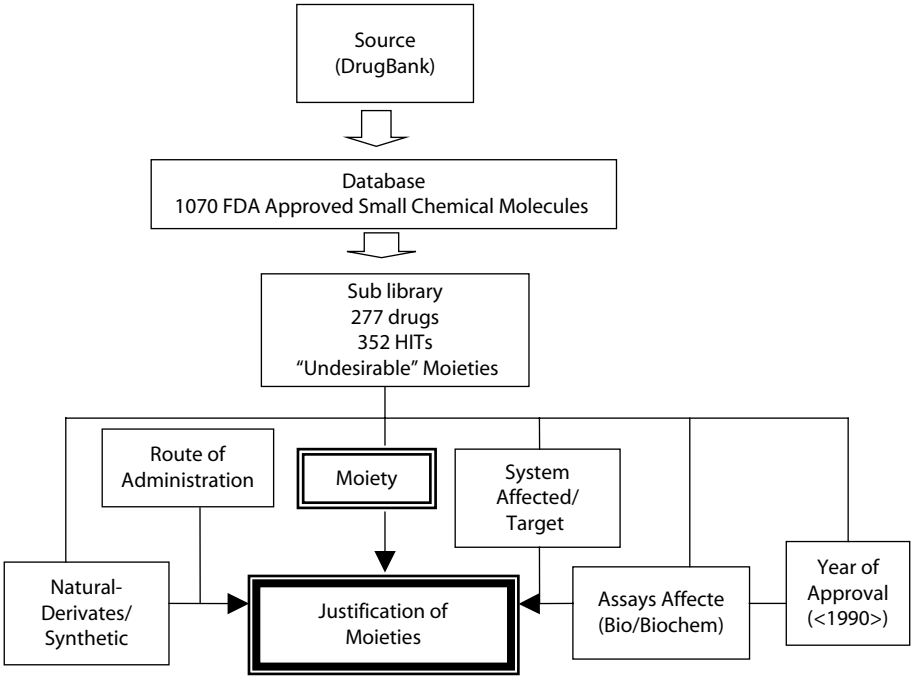

Fig. 2. Flow chart outlining the analytical design of the study.

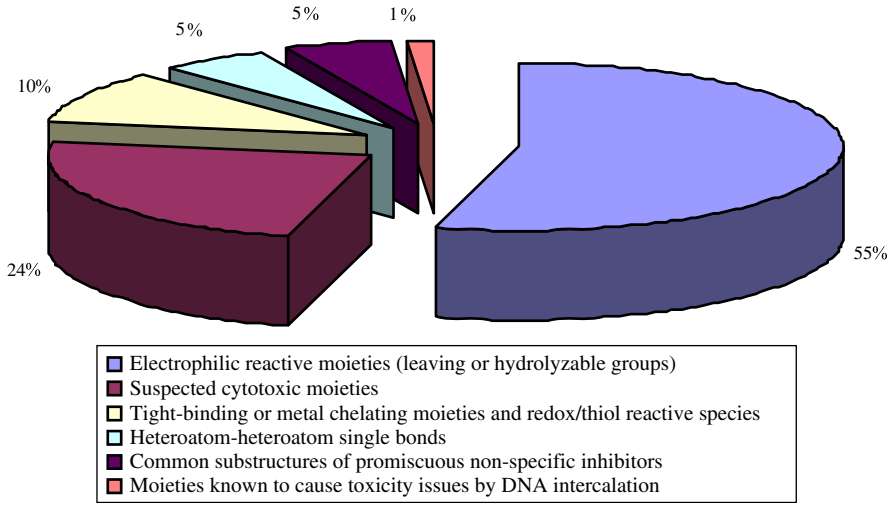

Fig. 3. Percentage of "undesirable" moieties categorised according to mechanism by which they distort assay results. 100% = 352 HITs; in 277 molecules from a library of 1070 marketed drug compounds.

are the most common moieties giving rise to false positives in biochemical assays. Interestingly, they are the most prevalent moieties observed in the sub-library of drugs.

Our designation of moieties that perturb enzyme-based assays includes categories 1–3 which represent 65% of "undesirable" moieties identified. The moieties that perturb cell-based assays are comprised of categories 4–6 and represent ~35% of the chemical entities investigated. However, there is bound to be considerable overlap between these two categories, i.e., moieties that perturb both biochemical enzyme and cell based assays.

3.2. Functional groups

Fig. 4 shows the distribution of HITs according to the type of functional group. Electrophilic aliphatic ester moieties are the most prevalent at 37%. Alkyl halides represent 8% and aliphatic ketones (suspected cytotoxic moieties) account for 14% of the total HITs. Other types of moieties, each of which contributes less than 2% to the total, account for 22% of HITs.

Natural product derivatives and macrocycles are currently marketed as drugs but are not indicative of the types of compounds that would appear from HTS campaigns. This is because pharmaceutical companies are focusing on identifying small synthetic molecules as lead compounds for drug candidates rather than screening natural products or high molecular weight compounds [7]. Within the 26% of known drug space containing "undesirable" moieties, prodrugs account for 3%. They are designed to be reactive or unstable and so it can be expected that they contain labile moieties. Half of the compounds containing HITs (13%) were of the type likely to emerge from a HTS campaign, i.e., synthetic drug molecules. Finally, macrocycles account for 4% of the HITs identified and natural products 6%.

Golebiowski et al. [33,34] have published two review papers on the success of HTS technology. In them they report on 111 structures, in total, discovered for drug development projects. Eighteen of the structures reported in these papers contain an "undesirable" moiety. The reviews were published in 2001 and 2003, respectively, coinciding with many of the

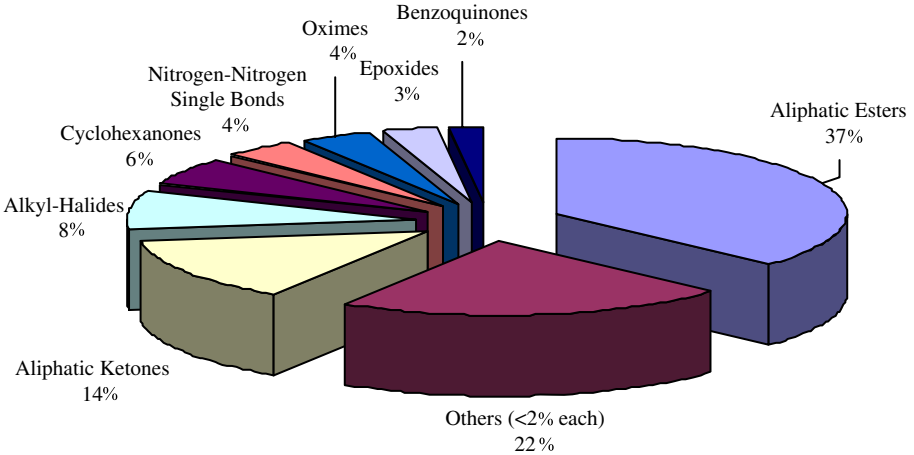

Fig. 4. Percentage incidence of "undesirable" moieties by type of functional group. 100% = 352 HITs; in 277 molecules from a library of 1070 marketed drug compounds.

publications on ''undesirable'' chemical moieties. This suggests the practice of eliminating compounds from libraries, based on ''undesirable'' moieties, is relatively recent. However, although decision makers in pharmaceutical companies are now more aware of these moieties, screening groups still adopt differing approaches to eliminate these compounds and acquire new compounds for their screening collections.

The customary way of distinguishing tractable hits and leads from HTS campaigns is for senior medicinal chemists to review the results and decide which compounds to take forward [3]. Medicinal chemists are often also responsible for the selection of new molecules for screening libraries. Lajiness et al. [20] have reported that the opinions of medicinal chemists vary widely depending on experience and that even the same individual can be quite inconsistent. Medicinal chemists make such judgments based upon solubility, selectivity, ease of synthesis, adsorption, distribution, metabolism, excretion, and toxicity properties but also inevitably upon their disposition [35]. However, the majority of medicinal chemists would agree that the moieties identified by Rishton [16] and Olah et al. [36] (thiols, redox compounds, and Michael acceptors, etc.) are undesirable. Consequently, most would be excluded early on in the drug discovery process.

The results presented in this study show that ''undesirable'' molecular species occur frequently in marketed drug compounds and subsequently the question arises whether they should be removed from screening libraries or not. It has been suggested that a flagging system for these moieties would be more advantageous [17]. Since clinical trials are the most expensive part of drug development it seems that eliminating the chemical structures under investigation is counter-productive. A possible solution to this problem is to keep compounds containing the discussed moieties on separate screen plates and they can be used depending on biological target or assay format used.

3.3. Route of administration and therapeutic target

In an attempt to find some correlation between the moieties under study and their presence in drug molecules, the drugs were classified into groups outlining the route of administration and therapeutic target.

For route of administration of the synthetic drug HITs the oral administration was found to be the most prevalent with a 45% share of the compounds investigated. This can be understood since most drugs are intended to be formulated for this route. IV and topical administration are also significant at ~20% each. Drugs administered by inhalation, intramuscular (IM) injection and intraocular (IO) comprise less than 10% each.

For the distribution of primary therapeutic effect it is apparent that largest percentage of the compounds containing an ''undesirable'' moiety fall in the category of chemotherapeutic compounds (37%) followed by drugs intended to affect the hormonal system (19%). Fifteen percent of the drugs are utilized for cardiovascular treatment. Ten percent are used for the central nervous system (CNS) and for neurotransmission, respectively. Finally, 6% are for immune system treatments and for 3% of the drugs investigated no therapeutic effect was found. Even though chemotherapeutic drug compounds are the most abundant, there is relatively even distribution with all categories represented. It is difficult to deduce whether this distribution is simply due to the observation that more drugs are commercially available for some therapies than others.

3.4. Justification – What is the function of the ''undesirable'' moiety?

It is important to understand the structure–activity relationships (SAR) for each drug molecule and moieties under study in the context of its molecular target. The drugs were individually assessed based on their SAR and

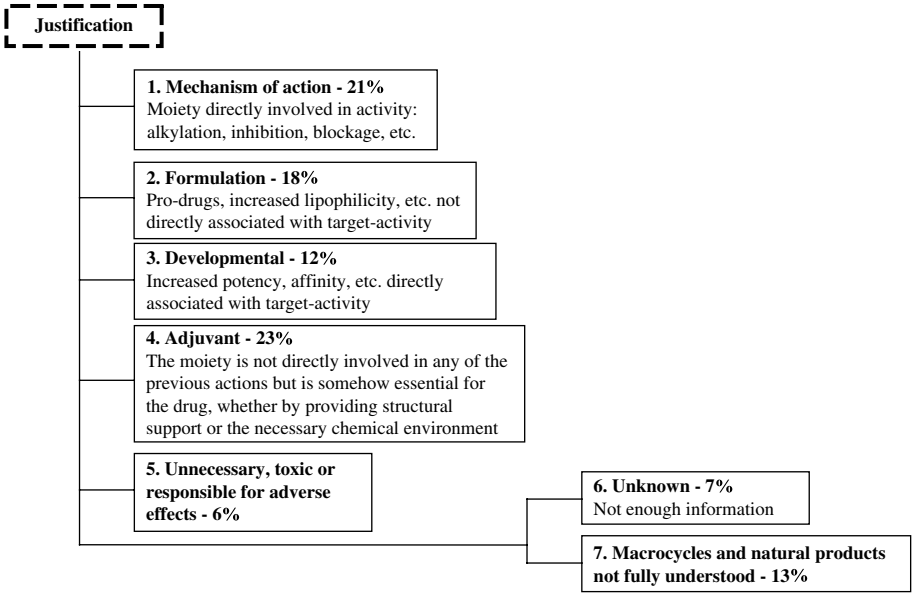

Fig. 5. Results of classification of function for the ''undesirable'' moieties. 100% = 352 HITs; in 277 molecules from a library of 1070 marketed drug compounds.

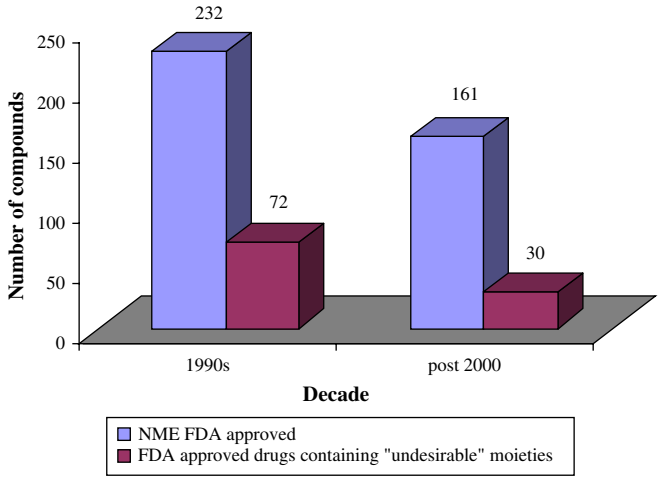

Fig. 6. Correlation of NMEs (New Molecular Entities) and the occurrence of ''undesirable'' moieties.

categorised according to the main function of the ''undesirable'' moiety. Five primary roles were defined; mechanism of action, formulation, developmental, adjuvant, and unnecessary/toxic species. Moieties of unknown function and natural products/macrocycles were segregated. The results of the classification are shown in Fig. 5.

Cursory analysis suggests similar distribution between the four principal moiety roles, i.e., between 12 and 23%. These functions are not rigid as significant overlap exists between the roles that the moieties play. However, the analysis does suggest that 74% of the identified ''undesirable'' moieties are integral parts of the drug molecules and cannot be eliminated from their structures without sacrificing efficacy.

3.5. Time of marketing of the drug compounds

The policy of excluding ''undesirable'' moieties can be assessed by correlating their incidence with when the drugs were FDA approved. In Fig. 6 the number of new molecular entities (NMEs) and the number of the drug compounds containing ''undesirable'' moieties are plotted for the 1990s and the current decade [37]. It can be assumed that in the 1990s the vast majority of NMEs was not initially discovered utilising HTS technology as it was still in its infancy. Approximately 30% of NMEs approved by the FDA during the 1990s contain ''undesirable'' moieties. Less than 20% of NMEs approved this decade contain an ''undesirable'' moiety. The decline in NMEs approved and marketed each year is a well-established trend [38]. It is tempting to draw the conclusion that the hasty elimination of these ''undesirable'' regions of chemical space could actually contribute to the decline in the number of NMEs. Undoubtedly, this decline is caused by many different phenomena, however, the parallels presented in Fig. 6 are difficult to ignore.

4. Conclusions

Of the original library of 1070 drug compounds 26% contain at least one of the moieties deemed ''undesirable'' for assay formats used in HTS campaigns. Half of the compounds containing these moieties were of the type likely to emerge from a HTS campaign, i.e., synthetic drug molecules. Electrophilic reactive species, particularly aliphatic esters, represent more than half the ''undesirable'' moieties found in the collection. As the study has been performed on marketed drugs (known drug space), the strong presence of these moieties in the database suggests that there might not be a clear reason for eliminating them in earlier stages of preclinical work. Overall, our observations support the approach that criteria for the exclusion of these moieties in screens should be flexible and applied in context considering both the biological target/system affected and the assay format used.

Appendix. Supporting information

Supporting information associated with this article can also be found in the online version, at doi: 10.1016/j.ejmech.2008.06.013.

References

[1] R.P. Hertzberg, A.J. Pope, Curr. Opin. Chem. Biol. 4 (2000) 445–451.
[2] T. Mander, Drug Discov Today 5 (2000) 223–225.
[3] A.M. Davis, D.J. Keeling, J. Steele, N.P. Tomkinson, A.C. Tinker, Curr. Top. Med. Chem. 5 (2005) 421–439.
[4] C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, Adv. Drug Deliv. Rev. 23 (1997) 3–25.
[5] C.A. Lipinski, J. Pharmacol. Toxicol. Methods 44 (2000) 235–249.
[6] S.J. Teague, A.M. Davis, P.D. Leeson, T.I. Oprea, Angew. Chem. Int. Ed. 38 (1999) 3743–3748.
[7] T.I. Oprea, A.M. Davis, S.J. Teague, P.D. Leeson, J. Chem. Inf. Comput. Sci. 41 (2001) 1308–1315.
[8] T.I. Oprea, Mol. Divers. 5 (2000) 199–208.
[9] N. Baurin, B.C. Richardson, I. Chen, N. Foloppe, A. Potter, A. Jordan, S. Roughley, M. Parratt, P. Greany, D. Morley, R.E. Hubbard, J. Chem. Inf. Comput. Sci. 44 (2004) 643–651.
[10] M.C. Wenlock, R.P. Austin, P. Barton, A.M. Davis, P.D. Leeson, J. Med. Chem. 46 (2003) 1250–1256.
[11] M.S. Lajiness, M. Vieth, J. Erickson, Curr. Opin. Drug Discov. Devel. 7 (2004) 470–477.
[12] W.J. Egan, W.P. Walters, M.A. Murcko, Curr. Opin. Drug Discov. Develop. 5 (2002) 540–549.
[13] S.L. McGovern, in: J. Alvarez, B.K. Shoichet (Eds.), Virtual Screening in Drug Discovery, Taylor & Francis, Boca Raton, 2005, pp. 107–123.
[14] X.L. Yang, A.J. Wang, Pharmacol. Ther. 83 (1999) 181–215.
[15] G.M. Rishton, Drug Discov Today 2 (1997) 382–384.
[16] G.M. Rishton, Drug Discov Today 8 (2003) 86–96.
[17] O. Roche, P. Schneider, J. Zuegge, W. Guba, M. Kansy, A. Alanine, K. Bleicher, F. Danel, E.M. Gutknecht, M. Rogers-Evans, W. Neidhart, H. Stalder, M. Dillon, E. Sjögren, N. Fotouhi, P. Gillespie, R. Goodnow, W. Harris, P. Jones, M. Taniguchi, S. Tsujii, W. von der Saal, G. Zimmermann, G. Schneider, J. Med. Chem. 45 (2002) 137–142.
[18] S.L. McGovern, E. Caselli, N. Grigorieff, B.K. Schoichet, J. Med. Chem. 45 (2002) 1712–1722.
[19] S.L. McGovern, B.T. Helfand, B. Feng, B.K. Schoichet, J. Med. Chem. 46 (2003) 4265–4272.
[20] M.S. Lajiness, G.M. Maggiora, V. Shanmugasundaram, J. Med. Chem. 47 (2004) 4891–4896.
[21] D.S. Wishart, C. Knox, A.C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, Z. Chang, J. Woolsey, Nucleic Acids Res. 34 (2006) D668–D672.
[22] <www.drugs.com>.

[23] <www.RXList.com>.
[24] <www.pubmed.gov>.
[25] J.A. Kerr, Chem. Rev. 66 (1966) 465–500.
[26] Y.B. Feng, A. Shelat, N.T. Doman, R.K. Guy, K.B. Shoichet, Nat. Chem. Biol. 1 (2005) 146–148.
[27] E.P. Blower, C. Yang, A.M. Fligner, S.J. Verducci, L. Yu, S. Richman, N.J. Weinstein, J. Pharmacol. 2 (2002) 259–271.
[28] K. Ichihashi, T. Osawa, S. Toyokuni, K. Uchida, J. Biol. Chem. 276 (2001) 23903–23913.
[29] L.A. Arnold, E. Estébanez-Perpiñá, M. Togashi, N. Jouravel, A. Shelat, A.C. McReynolds, E. Mar, P. Nguyen, J.D. Baxter, R.J. Fletterick, P. Webb, R.K. Guy, J. Biol. Chem. 280 (2005) 43048–43055.
[30] J. Reynisson, G.B. Schuster, S.B. Howerton, L.D. Williams, R.N. Barnett, C.L. Cleveland, U. Landman, N. Harrit, J.B. Chaires, J. Am. Chem. Soc. 125 (2003) 2072–2083.
[31] <www.fda.gov>.
[32] From this point forward, the results will be based on the number of total HITs that were identified in the database (352 HITs) and not the number of compounds that were identified as having an undesirable moiety.
[33] A. Golebiowski, R.S. Klopfenstein, E.D. Portlock, Curr. Opin. Chem. Biol. 5 (2001) 273–284.
[34] A. Golebiowski, R.S. Klopfenstein, E.D. Portlock, Curr. Opin. Chem. Biol. 7 (2003) 308–325.
[35] P.D. Leeson, A.M. Davis, J. Med. Chem. 47 (2004) 6338–6348.
[36] M.M. Olah, C.G. Bologa, T.I. Oprea, Curr. Drug Discov. Technol. 1 (2004) 211–220.
[37] The FDA does not supply information about year of approval for 32 compounds containing HITs so these are not used.
[38] B. Hughes, Nat. Rev. Drug Discov. 7 (2008) 107–109.