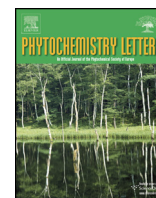


Contents lists available at ScienceDirect

Phytochemistry Letters

journal homepage: www.elsevier.com/locate/phytolAlkaloids as drug leads – A predictive structural and biodiversity-based analysis[☆]Vafa Amirikia, Michael Heinrich^{*}

Centre for Pharmacognosy and Phytotherapy/Research Cluster Biodiversity and Medicines, UCL School of Pharmacy, University of London, 29-39 Brunswick Sq., London WC1N 1AX, UK

ARTICLE INFO

Article history:

Received 2 May 2014

Received in revised form 19 June 2014

Accepted 26 June 2014

Available online 9 July 2014

This paper is dedicated to the memory of Andrew Marston (1953–2013), outstanding phytochemist who is much missed by his friends and colleagues and forms part of a special issue of *Phytochemistry Letters* commemorating his life and work.

Keywords:

Alkaloid development
Natural products
Drug-likeness
Secondary metabolites
Drug discovery
Biodiversity
Modeling

ABSTRACT

The process of drug discovery and development particularly that of natural products, has evolved markedly over the last 30 years into increasingly formulaic approaches. As a major class of natural products initially discovered and used as early as 4000 years ago, alkaloids and the species they are derived from have been used worldwide as a source of remedies to treat a wide variety of illnesses. Yet, a tremendously wide discrepancy between their historical significance and their occurrence in modern drug development exists. Are alkaloids underrepresented in modern medicine?

The physicochemical features of 27,683 alkaloids from the Dictionary of Natural Products were cross-referenced to pharmacologically significant and other metrics from various databases including the European Bioinformatics Institute's ChEMBL and Global Biodiversity Information Facility's GBIF. For the first time we show that market/developmental performance of a class of compounds is linked to its biodiversity distributions, as defined by the GBIF dataset. The potential of such a large-scale data analysis is analyzed against both prevalent rules used to guide drug discovery processes and the larger context of natural product development.

© 2014 The Authors. Phytochemical Society of Europe. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/3.0/>).

1. Introduction

The archeological and historical record shows that people across Asia, Europe, and Africa used alkaloid-containing plants as early as 2000 BCE (Aniszewski, 2007). Applications of such alkaloids included empirical medicines for animals and humans as well as sources of poison for hunting expeditions or executions (Wink, 1998). All throughout the centuries these plants and associated isolated compounds were increasingly and continuously used for, as one scholar encapsulates it, 'Murder, Magic and Medicine' (Mann, 1992). The early 19th century saw breakthroughs in the isolation and

characterization of purified compounds. In the early years of the 19th century, Friedrich Sertürner isolated what we know today as morphine. This led to a cascade of successful isolations and discoveries of isolated compounds by several European scientists including the isolation of xanthine (1817), strychnine (1818), atropine (1819), quinine (1820), and caffeine (1820) (Heinrich et al., 2012). This burst of single compound isolation has been characterized by many, including Sneader, as 'the greatest advance in the process of drug discovery' (Sneader, 2005).

The process of drug discovery as it stands today differs greatly from the ones prominent throughout most of the 20th century decades. Highly popular, yet debated empirical rules aiming to enhance the selectivity of drug candidates have for many years been in the spotlight. Popular terms such as 'lead-like' and 'drug-like' have gained prominence through the work of Lipinski and Congreve (Lipinski, 2000; Rees et al., 2004). As one explores the literature, it is very clear that what exactly druglikeness entails really depends on the intended application of the compound. Properties appropriate for successful metabolism of an orally administered drug differ greatly from, for example, transdermal injections. The applicability

[☆] This paper forms part of a special issue of *Phytochemistry Letters* dedicated to the memory of Andrew Marston (1953–2013), outstanding phytochemist who is much missed by his friends.

Abbreviations: Clog *P*, calculated log *P*; DNP, dictionary of natural products; GBIF, global biodiversity information facility; HBA, hydrogen bond acceptor; HBD, hydrogen bond donor; log *D*, distribution coefficient; log *P*, partition coefficient; MW, molecular weight; p*K*_a, acid dissociation constant (p*K*_a).

^{*} Corresponding author. Tel.: +44 20 7753 5844.

E-mail address: m.heinrich@ucl.ac.uk (M. Heinrich).

and application of such rules to other research areas is an active debate in drug research and development.

One conspicuously lacking class of compounds in this debate has been natural products, which, however, are well known to be of major importance as medicines (e.g. Cragg and Newman, 2005; Newman and Cragg, 2007; Saxton, 1971). It could be argued that the sheer diversity of natural products does not allow for adherence to such rules, yet nevertheless the importance of natural products (and specifically alkaloids) in modern drug discovery cannot be over-estimated as their use has been linked closely the history of human use of such resources (Heinrich, 2013).

Following the initial discoveries and isolations there was a gradual increase in the number of known and medicinally used alkaloids. Currently, the Dictionary of Natural Products (DNP) lists over 27,000 compounds as alkaloids (Hocking, 1997 and updates – *dnp.chemnetbase.com*). Other datasets define and list fewer alkaloids.¹ Much of the uncertainty of how many alkaloids actually exist stems from various issues including: poor chemical identification or structure elucidation, lack of dereplication, chemical ambiguities, and the varying definitions of what exactly constitutes an alkaloid (Rates, 2001). As with natural products as a whole, many have proposed differing classificatory schemes for alkaloids. One popular scheme divides the whole class of compounds into three categories:

- True alkaloids (compounds which derive from amino acid and a heterocyclic ring with nitrogen,
- Protoalkaloids (compounds, in which the N atom derived from an amino acid is not a part of the heterocycle), and
- Pseudoalkaloids (compounds, the basic carbon skeletons of which are not derived from amino acids) (Eagleson, 1994).

The scope of this study encompasses all such variations in definitions by taking the widest categorization of alkaloids as a class of compounds; essentially the 27,000+ found in the DNP (as of April 2014).

In this article we argue that – despite their history of use – alkaloids are considerably underrepresented as new *marketed* or *licensed medicines* ('drugs'). Alkaloids are relatively absent as compared with synthetic, semi-synthetic, and other non-alkaloid natural drug leads which successfully enter the pharmaceutical market today. We argue that barriers to development are strongly correlated to physicochemical properties of compounds. In addition, earlier research suggests that weediness (which in turn is linked to a species abundance) can serve to enhance the search for novel compounds in drug discovery (Stepp, 2004). How does this hold up against often cited challenges associated with access, supply, and production of such alkaloids?

This article examines the similarity of physicochemical and biodiversity characteristics of pharmaceutical and non-pharmaceutical alkaloids in order to pinpoint why alkaloids are underrepresented in the pharmaceutical arena and uses Global Biodiversity Information Facility (GBIF) data to assess this in the context of the species abundance in terms of its geographical distribution. GBIF is undisputedly one of the most comprehensive datasets on the distribution of individual species currently available. GBIF defines an occurrence as documented evidence of a named organism in nature. How does the phytogeographical abundance of a plant species correlate with the 'success' of compounds derived from the taxon to be developed into a marketed drug?

2. Results and discussion

2.1. Alkaloid drugs used as medicines

One would assume that with a 4000+ year history of use, often acting as remedies for a variety of illnesses, alkaloids and alkaloid

containing taxa would play an important and visible role in modern drug development (Bruhn and Bruhn, 1973). Or in the words of G. Cordell (1981) focusing on local and traditional uses: 'For thousands of years, indigenous groups around the world discovered, through self-experimentation with locally available plant extracts, that they could provide materials for hunting prey, culinary enhancement, amelioration from disease, relief of pain, and healing...in this [last] 200-year period, many alkaloids became critical components of the global pharmaceutical armamentarium, and tremendous healing has resulted from their clinical application' (Royal Society of Chemistry, 1971). Our search using the 'Dictionary of Alkaloids (Buckingham, 2010) and other sources identified a total of 53 alkaloids used currently or within the last 50 years for pharmaceutical applications (Table 1). To date less than 0.002% (53/27,000) of alkaloids or alkaloid-based drugs are marketed for such uses internationally (Table 1). It is not surprising that such a diverse set of natural products and their derivatives yield medicines which are used in a variety of applications ranging from cough-suppressants to antimalarial agents. However, in the last 25 years only galanthamine and taxol were newly introduced into biomedicine, and the former in essence through an extension of the therapeutic claims (i.e. from poliomyelitis to Alzheimer's disease, Heinrich and Teoh, 2004). There are only less than 200 others which are commonly used in industrial processes and the manufacturing of commercial goods (for example: N,N'-dioctadecanoyl ethanediamine is an antifoaming agent used in the polymer industry and methylamine hydrochloride is used in the tanning industry).

A quantitative analysis of alkaloids in modern pharmaceutical research and development based on their physicochemical properties

One preliminary step in characterizing the physicochemical makeup of pharmaceutical/medicinal alkaloids is to use metrics used in the commonly used empirical rules to select for druglikeness. At the most basic level, an initial analysis (Table 2 and Fig. 1) of 13 basic physicochemical properties of two sets of alkaloids (those used in marketed pharmaceutical/medicinal products ($n = 53$) and those which are not ($n = 1968$)²) shows averages of each physicochemical property ranging from –56 to +34% ((Pharma Avg./Total Avg.) – 1). The property which exhibits the largest difference between the two sets is the distribution coefficient ($\log D$)³ followed by hydrogen bond donors (HBD), the partition coefficient ($\log P$)⁴, and polar surface area (PSA) respectively. The $\log D$, HBD, $\log P$, and PSA of marketed pharmaceutical products is on average 31–55% lower than that of other alkaloids. These observations do not completely deviate from those general rules of thumb outlined above but rather indicate that adjustments to purely computational screening methods must be made to enhance alkaloid based drug discovery.

Average $\log D$ values for medicinal alkaloids are less than half as compared to other non-medicinal alkaloids. Average $\log P$ values for medicinal alkaloids are less than 40% as compared to other non-medicinal alkaloids. This suggests that ionization, acidity ($\log D$ is decreased as a function of increased pH), and ultimately solubility are potentially the most weighty factors in alkaloid development. These observations are somewhat confirmed by commonly used empirical rules in that they state that $\log P$ values should be <5.0 and <5.6 respectively (cf. Section 2.3).

² Alkaloid naming in both in the DnP and ChEMBL is highly inconsistent and fragmented. This value represents the total number of exact matches between both datasets minus those which have been labeled as 'pharmaceutical alkaloids'.

³ The distribution coefficient is the ratio of the sum of the concentrations of all forms of the compound (ionized plus un-ionized) in each of the two phases.

⁴ The partition coefficient is a ratio of concentrations of un-ionized compound between the two solutions.

¹ In 1988, NAPRALERT contained 16,000 alkaloids. As of 2001, no major additions are made to the dataset. It is estimated that NAPRALERT contains less than 20,000 alkaloids.

Table 1
Record of alkaloids used in marketed medicines, drugs, and clinical environments (excluding additive used in extending the storage life of drawn blood) (Based on The Royal Society of Chemistry, 1971; Cordell, 1981; Schmeller and Wink, 1998; Buckingham, 2010). Derivatives not included.

| Alkaloid name | Synonyms | Applications | Example product |
|-----------------------------|--|--|--|
| Aconitine | | Rheumatism, neuralgia, sciatica | Aconitysat TM , Bronpax TM , Pectovox TM , Vocadys TM |
| Adenine | | Antiviral agent, pharmaceutical aid used to extend storage life of whole blood | Adenosine, Ansy [®] |
| Ajmaline | Ajmaline, Gilurytmal, Merabitol, Raugalline, Rauwolfine, Rytmalin, Tachmalin | Antiarrhythmic agent | Aritmina TM , Gilurytmal TM , Rauwopur TM , Ritmos TM |
| Atropine | Tropine tropate | Antispasmodic, anti-parkinson, cycloplegic drug | Abdominol TM , Espasmo TM , Protec TM , Tonaton TM |
| Berberine | Berbericine, Umbellatine, | Eye irritations, AIDS, hepatitis | Kollyr TM , Murine TM , Sedacollyre TM |
| Boldine | | Cholelithiasis, vomiting, constipation | Boldoflorine TM , Boldosal TM , Oxyboldine TM , Sambil TM |
| Caffeine | | Neonatal apnea, atopic dermatitis | Agevis TM , Anlagen TM , Thomapyrine TM , Vomex A TM |
| Canescine | Harmony, Raunormine, Recanescine, Reserpidine | Antihypertensive agent | Deserpidine |
| Cathine | Norpseudoephedrine, Norisoeephedrine | Anorectic drug | Amorphan TM , Eetless TM , Recatol TM |
| Cinchonidine | Cinchonan-9-ol | Increases reflexes, epileptiform convulsions | Quinimax TM , Paluject TM |
| Cocaine | | Local anesthetic | Used in highly regulated clinical environments |
| Codeine | Methylmorphine, Codicept, Kodein, Tussipan | Antitussive, analgesic | Antituss TM , Codicaps TM , Tussipax TM |
| Colchicine | | Amyloidosis treatment, acute gout | ColBenemid TM , Colgout TM , Verban TM |
| Diethanolamine | 2,2'-Dihydroxydiethylamine, Diolamine | Base used in pharmaceuticals etc. | Menbutone Diethanolamine |
| Emetine | Ipecine, Methylcephaleine | Intestinal amoebiasis, expectorant drug | Cophylac TM , Ipecac TM , Rectopyrine TM |
| Ephedrine | | Nasal decongestant, bronchodilator | Amidoyna TM , Bronchicum TM , Peripherin TM , Solamin TM |
| Ergometrine | Ergonovine, Ergotrate, Ergobasine, Ergotocin, Ergostetrine | Postpartum/postabortal hemorrhage | Ergometron TM , Ergotrate Maleat TM , Syntometrine TM |
| Ergotamine | | Migrane treatment | Ergostat TM , Lingraine TM , Migral TM , Virdex TM |
| Eserine | Physostigmine | Ophthalmology, antidote/poisoning | Anticholium TM , Antilirium TM , Piloserine TM |
| Galanthamine | Galantamine, Jilkon, Karantonin, Lycoremine | Muscle relaxant, Alzheimer's | Nivalina TM |
| Hydrastine | | Gastrointestinal disorders | Gine Sedans TM , Kollyr TM |
| Hyoscine | Scopolamine | Motion sickness, | Buscopan TM , Hyospasmo TM , Lotanal TM , Transcop TM |
| Hyoscyamine | Daturine, Duboisine | Antispasmodic, anti-parkinson, cycloplegic drug | Bellatard TM , Cystospaz TM , Donnatab TM , Urised TM |
| Lobeline | | Anti-smoking, asthma, cough | Citotal TM , Lobatox TM , Refrane TM , Stopsmoke TM |
| Morphine | | Pain relief, diarrhea | Diastat TM , Duromorph TM , Oramprph TM , Spasmofer TM |
| N,N-Diallylbisnortoxinerine | Alcuronium chloride | Short acting muscular relaxant | Alloferin |
| Narceine | | Cough suppressant | Peneraj TM |
| Nicotine | | Anti-smoking | Nicabate TM , Nicoderm TM , Nicorette TM , Stubit TM |
| Noscapine | Narcotine | Cough suppressant | Bequitusin TM , Degoran TM , Tossamine TM , Tussisedal TM |
| Papaverine | Papaveroline tetramethyl ether | Vasodilator, gastrointestinal disorders | Acticarbine TM , Opdensit TM , Pameion TM , Vasocalm TM |
| Pelletierine | | Tenia infestations | Pelletierine tannate USP |
| Pilocarpine | Ocucarpine, Pilocarpol, Syncarpine | Miotic in treatment of glaucoma, leprosy | Frikton TM , Piladren TM , Salegen, Thiloadren TM , Vistacarpin TM |
| Quinidine | Conquinine, Conchinine, Pitayine | Ventricular and supraventricular arrhythmias, malaria, cramping | Cardioquin TM , Duraquin TM , Quindex TM , Rhythomochin 1 TM |
| Quinine | 6'-Methoxycinchonan-9-ol, Chinin | Malaria, babesiosis, myotonic disorders | Adaquin TM , Biquinate TM , Quinooctal TM , Zynedo-B TM |
| Raubasine | Ajmalicine | Vascular disorders | Circolene TM , Cristanyl TM , Duxil TM , Sarpan TM |
| Rescinnamine | Reserpinine, Anaprel, Apoterin, Cinnaloid, Rescaloid, Moderil, Scinnamina | Hypertension | Detensitral TM , Diuraupur TM , Rauwopur TM |
| Reserpine | | Hypertension, psychoses | Abicol TM , Briserin TM , Sandrii TM , Terbolan TM |
| Rotundine | Argemonine, Bisnorargemonine, | Analgesic, sedative, hypnotic agent | Rotundin-BVP, Transda |
| Sanguinarine | | Antiplatelet agent | Toothpastes and mouthwashes |
| Sparteine | | Uterine contractions, cardiac arrhythmias | Anxoral TM , Diffucord TM , Normotin TM , Tachynerg TM |
| Strychnine | Strychnidin-10-one | Eye disorders | Dysurgal TM , Pasuma TM Retinovix TM , Senirakt TM |
| Synephrine | | Vasoconstrictor, conjunctival decongestant, weight loss | Oxedrine, Sympatol |
| Taxol | Paclitaxel, Taxol A, Anzatax, Yewtaxan | Mamma and ovary carcinoma | Taxol TM |

Table 1 (Continued)

| Alkaloid name | Synonyms | Applications | Example product |
|---------------|---|---|---|
| Theobromine | | Asthma, diuretic agent | Atrofed TM , Circovegetalin TM , Dynamol TM , Urodonal TM |
| Theophylline | Austyn, Elan, Elixophyllin, Euphyllin, Nuelin | Asthma, bronchospasms | Adenovasin TM , Aerobin TM , Euphyllin TM , Theochron TM |
| Turbocuranine | Tubarine | Muscle relaxant | Jexin TM , Tubarine TM |
| Vinblastine | | Hodgkin's disease, testicular cancer, blood disorders | Periblastine TM , Velban TM , Velbe TM , Velsar TM |
| Vincamine | | Vasodilator | Aethroma TM , Angiopac TM , Pervin TM , Vincimax TM |
| Vincristine | | Burkitt's lymphoma | Norcristine TM , Oncovin TM , Vincrisul TM |
| Vindesine | | Chemotherapy | DAVA, Eldesine, Eldisine |
| Yohimbine | Aphrodine, Corymbin, Corynine, Yohimex, Hydroergotocin, Quebrachine, Yohimvetol | Aphrodisiac, urinary incontinence | Aphrodine TM , Pasuma TM , Prowess TM , Yohimex TM |

2.2. An analysis based on the empirical rules

Druglikeness rules such as the Rule of Three (Ro3) and Rule of Five (Ro5) were not designed with natural products in mind. Yet, we see that the medicinal alkaloids have 56% less R5 violations when compared with alkaloids at large, thus suggesting that such empirical rules (rules of thumb) are somewhat effective indicators in alkaloid development processes. In looking strictly at MWT values in the DNP for all 27,683 alkaloids we see that 27% pass the Ro3 while 77% pass the Ro5. When added chemical descriptors from ChEMBL are looked at Ro3/Ro5 pass rates decrease to 5% and 60% respectively. It is impressive that pharmaceutical/medicinal alkaloids have 56% less Ro5 violations and a 60% pass rate. If so many alkaloids pass the Ro5, are there other non-physicochemical factors which hinder their development?

Thus, in working toward deepening our understanding of to what extent such rules can be enhanced for research and development, a few modifications to such empirical rules, based on the current dataset, are proposed:

By using the following five parameters, we can predict over 90% of the pharmaceutical/medicinal alkaloids in our dataset. Many of the rules are extensions of the commonly used Ro3/Ro5 metrics which prioritize factors such as compound size and solubility.

1. MWT/PSA ≥ 3
2. HBD ≤ 4
3. BpKa 6–10
4. $\log P$ -1–7
5. Ratio MW/Heavy Atoms 13.2–13.9

When such rules are used to filter the ChEMBL dataset of the 2020 alkaloids included 672 (33%) comply with this rule. Rule 3 (pK_a 6–10) is the most selective in that it filters out 25% of the total alkaloids. The 672 alkaloids represent exactly one third of the total dataset. If this number is 100% accurate and assuming that there are no supply, commercial, and/or identification issues that leaves over 600 alkaloid candidates that have the chemical profile to serve in some commercial pharmaceutical/medicinal capacity. Extrapolating this liberal estimate to the larger DNP dataset suggests, a that there are 6000–7000 alkaloids which carry this 'development potential', as defined as physicochemical druglikeness similar to the Ro3/Ro5 empirical rules.

The goal of this exercise, far from merely introducing another empirical rule into the druglikeness debate, is rather to highlight that several dozen alkaloids used across a variety of pharmaceutical/medicinal applications have corresponding physicochemical properties. Therefore, the empirical rule proposed above selects for >90% of the 53 medicinal as compared to 60% with the Ro5.

2.3. Is this linked to a species abundance?

The more difficult question regarding the quantification of alkaloid biodiversity and the utility and limitations of the GBIF database has already been outlined in previous sections. It is important to note that data has only been extracted for 7435 of the total alkaloid set (dataset is 14.6% complete). Preliminary results are shown in Fig. 2. We see that 93% of all pharmaceutical/medicinal alkaloids have more than 50 occurrences in the GBIF

Table 2

Comparison of average of physicochemical properties between pharmaceutical and/or medicinally significant/insignificant alkaloids.

| Variable | Pharmaceuticals/drugs in ChEMBL (n = 53) | | Other alkaloids in ChEMBL (n = 1968) | | % Difference in Avg. (Pharma Avg./ Total Avg.) - 1) |
|------------------|--|------------------|--------------------------------------|------------------|---|
| | Range 90% of values fall within (SD) | Category average | Range 90% of values fall within (SD) | Category average | |
| MWT | 162.23–809.41 (164.78) | 375.88 | 219.23–840.70 (178.70) | 446.26 | -15.77 |
| AlogP | -0.02 to 4.89 (1.46) | 1.46 | -0.92 to 7.2 (2.42) | 2.97 | -14.38 |
| PSA | 32.78–153.45 (42.24) | 65.91 | 28.23–243.51 (61.11) | 96.29 | -31.55 |
| HBA | 2–12 (2.93) | 5.17 | 2–13 (3.40) | 5.76 | -10.24 |
| HBD | 0–3 (1.02) | 1.19 | 0–6 (1.90) | 2.27 | -47.32 |
| #Rotatable Bonds | 0–10 (3.33) | 4 | 0–16 (5.10) | 4.87 | -17.92 |
| ApKa | 8.60–13.93 (2.44) | 11.44 | 3.41–13.71 (3.54) | 10.12 | 13.02 |
| BpKa | 5.90–9.98 (1.54) | 7.91 | 1.01–10.56 (3.31) | 6.41 | 23.39 |
| ACDlogP | -0.65 to 5.75 (2.03) | 1.84 | -1.21 to 7.72 (2.85) | 3.11 | -40.66 |
| logD | -1.09 to 5.2 (2.06) | 0.98 | -2.52 to 7.33 (2.89) | 2.22 | -55.75 |
| Aromatic rings | 1–4 (0.99) | 1.89 | 0–4 (1.32) | 1.61 | 17.42 |
| Heavy atoms | 13–46 (11.92) | 27.60 | 16–59 (12.54) | 31.62 | -12.68 |
| QED weighted | 0.25–0.88 (0.22) | 0.66 | 0.11–0.85 (0.23) | 0.49 | 35.26 |
| #Ro5 violations | 0–2 (0.65) | 0.28 | 0–2 (0.90) | 0.64 | -56.01 |

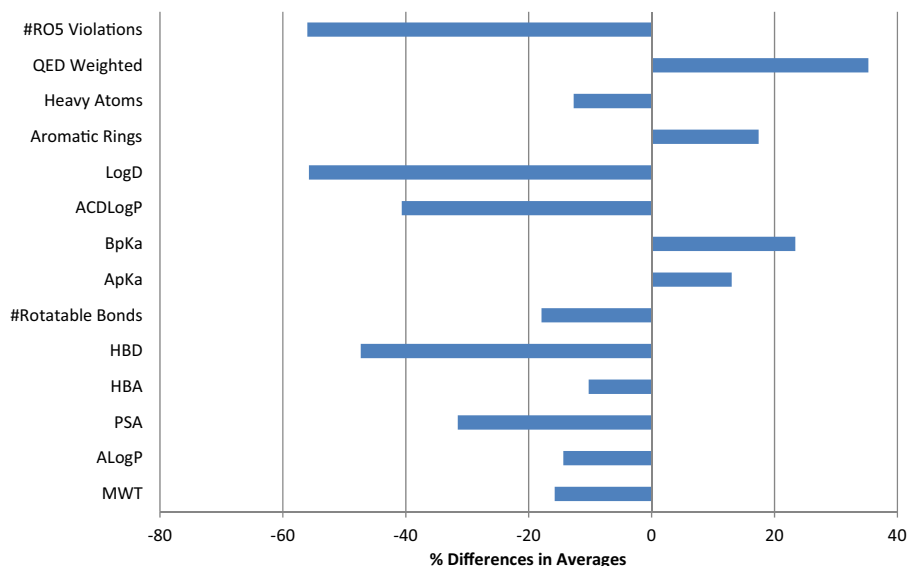


Fig. 1. % Difference in averages of physicochemical properties between pharmaceutical/medicinally significant and insignificant alkaloids.

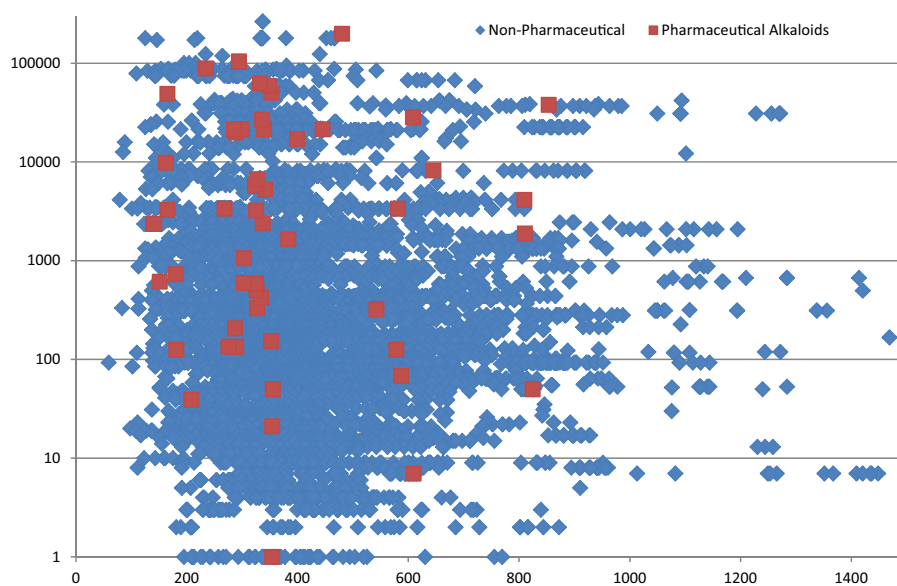


Fig. 2. Molecular weight (g/mol) and GBIF occurrences (limited data set ($n = 7435$) note logarithmic scale). Stripes of horizontal data points refer to families of alkaloids all derived from the same natural sources. Squares represent alkaloids in the pharmaceutical arena. Diamonds represent non-pharmaceutical alkaloids.

database. Only two alkaloids (chondocurine and vincamine) used have less than 10 occurrences in the GBIF database.

When averaging the two data sets the average occurrence of the pharmaceutical alkaloids set is 17,952 (s.d. = 35,595) occurrences while the non-pharmaceutical set averages at 4,165 occurrences (s.d. = 15,072). The standard deviation of the non-pharmaceutical set is significantly higher when calculated as a percent of the category average. This is logical considering the wide variation of abundances of alkaloid producing plants around the globe. These results lend support to those who argue that supply issues are the dominating indicator of successful research, development, and commercialization of natural products.

Many such as Principe (1991) have cited supply constraints as a key obstacle in the development of natural products. For example, Harvey states that natural products are unattractive to many pharmaceutical companies because of perceived difficulties relating to the complexities of natural product chemistry and to the access and supply of natural products resulting in technical

difficulties relating to the (larger scale) isolation of bioactive natural products (Harvey, 2008). One effort which shows much promise was put forth by the Global Biodiversity Information Facility (GBIF), which describes itself as operating 'through a network of nodes, coordinating the biodiversity information facilities of participant countries and organizations, collaborating with each other and the Secretariat to share skills, experiences and technical capacity.'⁵ Biodiversity data is served through four 'portals'; occurrences, datasets (smaller datasets endorsed and subsequently published by GBIF).

3. Conclusions

Overall, these data demonstrate that alkaloids are under represented in the context of newly introduced medicines. Although commonly employed empirical rules used to hone in

⁵ <http://www.gbif.org/whatisgbif>. Accessed December 14th 2013.

on drug-like compounds or lead-like fragments do filter out more than 50% of all alkaloids, other factors must be analyzed to more accurately pinpoint how best to tap into the vast remainder of undeveloped alkaloids.

For the first time we show that biodiversity distributions, as defined by the GBIF dataset, help in understanding to what extent a taxon's distribution relates to market/developmental performance. 93% of the pharmaceutical/medicinal alkaloids have >50 GBIF occurrences (Principe, 1991), indicating that a taxon's abundance considerably affects the development of an alkaloid into a medical product. This supports the view that supply constraints are a considerable concern given that most plants of interest today are usually indigenous only to biodiversity-rich countries especially of the tropics and subtropics.

Thus a larger sample size, both of calculated/observed chemical properties and host plant species distribution data will increase the accuracy of such an analysis. The specificity of pharmacological action of alkaloids and potential toxicological concerns have not been addressed in this context and in subsequent analyses these may need to be taken into consideration. It is likely that as empirical rules in drug development have been developed over the last 15 years to better hone in on drug-like compounds, rules and additional insights regarding natural products and specific natural product classes such as alkaloids will begin to emerge to drive development in this under-tapped area drug development. Presumably, future discovery of drug-like alkaloids will begin with abundant, easily accessible and scalable plants rather than a set of specific empirical rules which narrow down compounds of interest.

4. Materials/methods

The initial data set of 27,683 alkaloids was imported from the Dictionary of Natural Products web portal (dnp.chemnetbase.com/) into Microsoft Excel 2010. A maximum of 33 data types, both qualitative and quantitative, were extracted for each of the 27,683 alkaloids. Highly incomplete (ex. solubility) and irrelevant (ex. DnP classification codes) data types were omitted. Modifications were made to the format of some data to ensure consistency.

Data from ChEMBL (<https://www.ebi.ac.uk/chembl/>) was manually queried for each of the compounds listed in the initial DNP extract (synonyms from each of the two datasets also included). Due to the wide variance between keywords and formats between the two datasets, automating this process would not yield many 'hits.' Therefore, this initial 'bridging' of datasets was performed manually in the form of each query and subsequent data import being performed manually. This initial effort yielded 2020 'hits' (7%) and it is estimated that there are <500 potential remaining compounds that exist both in DnP and ChEMBL datasets. Similar to the DNP, not all data types in ChEMBL were deemed relevant and analyzed (ex. Molregno, Max Phase, and Med Chem Friendly).

GBIF (www.gbif.org/occurrence) data was manually queried and exported from the web portal into Microsoft Excel. Currently it contains 424,254,844 occurrences of organisms in nature including 117,909,945 (27.8%) records from the kingdom Plantae. Occurrences include collected and documented specimens, citations, and records in nature. For example: the DNP reports that the alkaloid monocrotaline can be found under the heading of five taxa: *Crotalaria retusa* L., *Crotalaria spectabilis* Roth, *Crotalaria aegyptiaca* Benth., *Crotalaria burhia* Benth. (all Fabaceae) and *Lindelfia spectabilis* (Boraginaceae). Occurrences in GBIF for these five plant species total to 3222 (2575, 440, 144, 27, and 36 respectively). A preliminary calculation of this nature was possible for 27% of all the alkaloids listed in the DNP (7435/27,783).

References

- Aniszewski, T., 2007. *Alkaloids – secrets of life*, 1st ed. Elsevier, Amsterdam.
- Buckingham, J., 2010. *Dictionary of Alkaloids*, 1st ed. CRC Press, Boca Raton, FL.
- Bruhn, J.G., Bruhn, C., 1973. Alkaloids and ethnobotany of Mexican peyote cacti and related species. *Econo. Bot.* 27 (2) 241–251.
- Cordell, G., 1981. *Introduction to Alkaloids*, 1st ed. Wiley, New York.
- Cragg, G., Newman, D., 2005. Plants as a source of anti-cancer agents. *J. Ethnopharmacol.* 100 (1) 72–79.
- Eagleson, M., 1994. *Concise Encyclopedia Chemistry*, 1st ed. Walter de Gruyter, Berlin.
- Harvey, A., 2008. Natural products in drug discovery. *Drug Disc. Today* 13 (19) 894–901.
- Heinrich, M., 2013. Ethnopharmacology and Drug Discovery. In: Reedijk, J. (Ed.), Elsevier Reference Module in Chemistry, Molecular Sciences and Chemical Engineering. Elsevier, Waltham, MA, 11 September, <http://dx.doi.org/10.1016/B978-0-12-409547-2.02773-6>.
- Heinrich, M., Teoh, H.L., 2004. Galanthamine from snowdrop – the development of a modern drug against Alzheimer's disease from local Caucasian knowledge. *J. Ethnopharmacol.* 92, 147–162.
- Heinrich, M., Barnes, J., Gibbons, S., Williamson, E.M., 2012. *Fundamentals of Pharmacognosy and phytotherapy*, 2nd ed. Churchill Livingstone, Edinburgh.
- Hocking, G., 1997. *A Dictionary of Natural Products*, 1st ed. Plexus Pub, Medford, NJ.
- Lipinski, C.A., 2000. Drug-like properties and the causes of poor solubility and poor permeability. *J. Pharm. Tox Meth.* 44 (1) 235–249.
- Mann, J., 1992. *Murder, Magic, and Medicine*, 1st ed. Oxford University Press, Oxford.
- Newman, D., Cragg, G., 2007. Natural products as sources of new drugs over the last 25 years. *J. Nat. Prod.* 70 (3) 461–477.
- Principe, P.P., 1991. Valuing the biodiversity of medicinal plants. In: Akerele O, Heywood, V., Syngae, H. (Eds.), *The Conservation of Medicinal Plants*. Cambridge University Press, New York, pp. 79–124.
- Rates, S., 2001. Plants as source of drugs. *Toxicol.* 39 (5) 603–613.
- Rees, D., Congreve, M., Murray, C., Carr, R., 2004. Fragment-based lead discovery. *Nat. Rev. Drug Disc.* 3 (8) 660–672.
- The Royal Society of Chemistry, *The Alkaloids: Vol. 1, Specialist Periodical Reports*, The Royal Society of Chemistry, London, 1971.
- Saxton, J., 1971. *The Alkaloids. Vol. 1. A review of the literature published between January 1969 and June 1970*, 1st ed. Chemical Society, London.
- Schmeller, T., Wink, M., 1998. Utilization of Alkaloids in Modern Medicine. Springer, Heidelberg, pp. 435–459.
- Sneider, W., 2005. *Drug Discovery*, 1st ed. Wiley, Hoboken, NJ. (USA).
- Stepp, J., 2004. The role of weeds as sources of pharmaceuticals. *J. Ethnopharmacol.* 92 (2) 163–166.
- Wink, M., 1998. *A Short History of Alkaloids*. Springer, Heidelberg, pp. 11–44.