Phytochemistry Letters 10 (2014) xlviii–liii

Contents lists available at ScienceDirect

Phytochemistry Letters

journal homepage: www.elsevier.com/locate/phytol

Alkaloids as drug leads – A predictive structural and biodiversity-based analysis☆

Vafa Amirkia, 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.

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 though 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; MWT, molecular weight; pK_a, acid disassociation constant (pK_a).

* Corresponding author. Tel.: +44 20 7753 5844.
E-mail address: m.heinrich@ucl.ac.uk (M. Heinrich).

http://dx.doi.org/10.1016/j.phytol.2014.06.015

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 overestimated 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'-dioctadecanoylethanediamine 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).

[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.

[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.