

What's in a Name? Drug Nomenclature and Medicinal Chemistry Trends using INN Publications

Marta Serafini,[‡] Sarah Cargnin,[‡] Alberto Massarotti,[‡] Gian Cesare Tron, Tracey Pirali, and Armando A. Genazzani*



Cite This: *J. Med. Chem.* 2021, 64, 4410–4429



Read Online

ACCESS |



Metrics & More

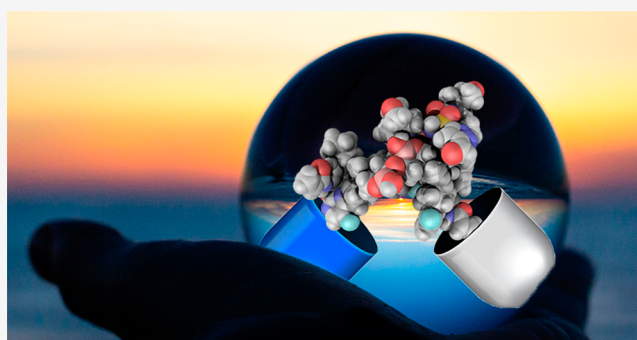


Article Recommendations



Supporting Information

ABSTRACT: The World Health Organization assigns international nonproprietary names (INN), also known as common names, to compounds upon request from drug developers. Structures of INNs are publicly available and represent a source, albeit underused, to understand trends in drug research and development. Here, we explain how a common drug name is composed and analyze chemical entities from 2000 to 2021. In the analysis, we describe some changes that intertwine chemical structure, newer therapeutic targets (e.g., kinases), including a significant increase in the use of fluorine and of heterocycles, and some other evolutionary modifications, such as the progressive increase in molecular weight. Alongside these, small signs of change can be spotted, such as the rise in spirocyclic scaffolds and small rings and the emergence of unconventional structural moieties that might forecast the future to come.



INTRODUCTION

Remember the biblical story of the Tower of Babel, in which the attempt to build a tower that would reach the heavens by the Babylonians was disrupted by the inability of the builders, that spoke different languages, to understand each other? Well, apparently there are today over 7000 spoken languages, and 34 of them are spoken by at least 45 million people.¹ If each pharmacological active principle were to be called differently in each country and language, it would be disastrous for health as well as for scientific progress. Think, for example, of pharmacovigilance, of scientific publications, and of those traveling around the world.

Brand names that exist in most parts of the world well-exemplify this situation. In many countries the same drug is sold with multiple brand names, and, to make matters worse, these brand names change from country to country. While chemists might think that International Union of Pure and Applied Chemistry (IUPAC) names represent unique identifiers, it must be acknowledged that these are complicated, almost impossible to memorize for the lay public, and error prone. A common name, short and easy to pronounce, that identifies the same medicine everywhere in the world, is therefore required.

This was realized soon after World War II, and in 1953 the World Health Assembly, the governing body of the World Health Organization (WHO), passed resolution WHA3.11, which stated that an expert Committee of the WHO “should undertake the selection and approval of non-proprietary names for

drugs” together with the recommendation that national pharmacopoeias should adopt such names.² Such a resolution was the birth of the international nonproprietary names (INN) system for drug identification that we still use today for the effective and safe identification of medicines, for safe prescribing, and for teaching. It has also been a pivotal pillar for the development of the generic market of off-patent drugs, whereby in most countries the INN substitutes the brand name and allows therefore to unlink the manufacturer from the therapeutic effect. The INN has not substituted naming agencies altogether (e.g., in the United States (U.S.), England, Japan, and China, to cite some, national naming agencies are still active), but intense cooperation between these agencies has led, with very few exceptions, to identical names around the world. To imagine a world in which drugs are identified with different names, think if all drugs used were in the situation of paracetamol, salbutamol, and adrenaline, which in the U.S. are known, respectively, as acetaminophen, albuterol, and epinephrine.^{3,4} Fortunately, these are some of the very few exceptions.

Received: February 1, 2021

Published: April 13, 2021



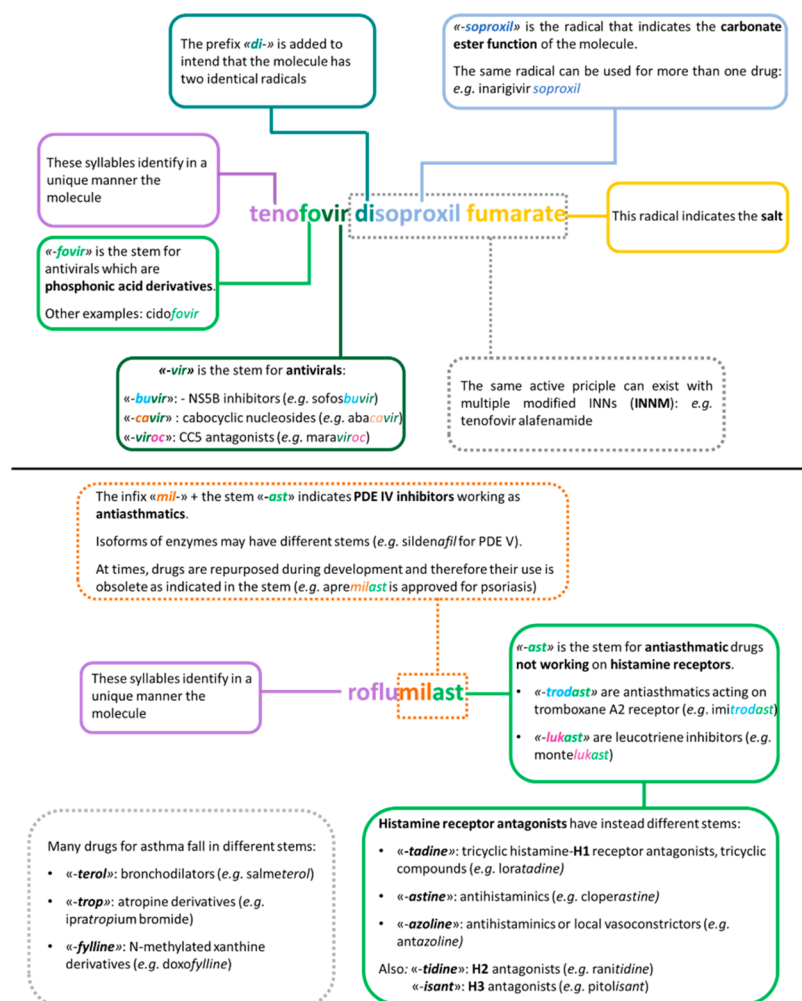


Figure 1. Use of stems, prefixes, infixes, and suffixes in INNs. The definitions of the stems in the figure are the official definitions that can be found in the stem book.⁹

While the present review will concentrate on the INN, it is interesting that contiguous fields have also adopted common names for their products, as a question of safety or consumer protection. Cosmetic ingredients, for example, have common International Nomenclature Cosmetic Ingredient (INCI) names issued by the Personal Care Product Council.⁵ Pesticides and agricultural chemicals also have their standardization for naming, which is the result of several different naming committees. In the U.S., the American National Standards Institute (ANSI) issues common names for chemicals, although other agencies, such as the International Standards Organization (ISO), or national standard organizations (e.g., British) also do so. It may occur that a few molecules receive common names by multiple bodies for different uses, and, while in most instances these names coincide (e.g., ascorbic acid is both an INCI and an INN), this is not always the case (e.g., nicoboxil is an INN, while the corresponding INCI is butoxyethyl nicotinate; the INN cetomacrogol 1000 corresponds to the INCI ceteth-20, both identifying the same tensioactive molecule; the ISO insecticide trichlorfon is equivalent to the INN drug metrifonate, once used for schistosomiasis,⁶ and the INN oxindanac, an anti-inflammatory drug that has never reached the market, is equivalent to the ISO herbicide quinclorac).

How are INNs Defined? INNs are issued upon request to the WHO and, in particular, to the Secretariat of the Expert Advisory Panel of the International Pharmacopoeia and Pharmaceutical Preparations, designated for this purpose, also known as the “INN Expert Group”. The INN Expert Group is composed of experts from all over the world that are led and coordinated by a Secretariat. Experts may include, among others, medicinal chemists, pharmacologists, biochemists, molecular biologists, and clinicians. Given that experts change over time and the scientific community modifies viewpoints on what is important in a drug, the composition of the group may also have an impact on the names chosen for a particular substance, with more or less emphasis given to the chemistry/structure, to the nature of the active principle, or to the mechanism of action. The current composition may be found on the WHO Web site.⁷ To provide consistency across fields, the meetings of the Committee are also attended by a number of other learned scientists, for example, representatives of national naming agencies and pharmacopeias, IUPAC experts, and Anatomical Therapeutic Chemical Classification System (ATC) code experts.

The person or company requesting the INN may propose a name that should abide by the rules and conventions of the INN system, and this may be accepted or modified by the

Committee.⁸ Briefly, the core element of the INN is the stem, which is composed of one or two syllables, and is usually located at the end of the drug name. Such a stem identifies drugs that have a shared feature, usually the mechanism of action, although it can also be a therapeutic use or a chemical/structural characteristic. As an example, think of proton pump inhibitors: *omeprazole*, *esomeprazole*, *pantoprazole*, *rabeprazole* all share the common stem *prazole* that identifies “antiulcer, benzimidazole derivatives” (these brief definitions are given by the Committee once a stem is officially identified).⁹ If drugs are not recognized by the INN Expert Group as being part of a broader category, they will be assigned a unique ending. If other molecules that share the new feature will have an INN requested for them in the future, then these will receive the same ending, and the committee will promote it as an official stem. The WHO publishes a “stem book” and regular updates that can be freely consulted.⁹ Yet, it is not only the stem that qualifies a name, as there are prefixes (a syllable at the beginning of the INN), infixes (a syllable in the middle of the word), or suffixes (a syllable at the end of the INN) that characterize the name and may give information to the learned reader. For example, the syllable *-gli-* characterizes many drugs used in diabetes (e.g., *glibenclamide*, *canagliflozin*, *sitagliptin*, *saroglitazar*, *rosiglitazone*; note that *-glifozin*, *-gliptin*, *-glitazar*, and *-glitazone* are all stems identifying different drug classes and that the prefix *gli-* identifies anti-hyperglycemics, which are sulfonamide derivatives; for examples of the most frequent stems used by the INN Expert Group since 2000, see Table 1 below. Another example of how INNs work is represented by the manner by which chiral switches are represented in INNs, in which an infix illustrates the change, for example, in *esomeprazole* (S isomer of *omeprazole*), *escitalopram* (S isomer of *citalopram*), *esketamine* (S isomer of *ketamine*), *levofloxacin* (levo-rotatory stereoisomer of *ofloxacin*) or *dexketoprofen* (the dextro-rotatory stereoisomer of *ketoprofen*).

In general, INNs are selected for the active moiety of drugs. Pharmaceutical reasons may drive slight modifications to be made during the drug development process or during the lifetime of the product. This is the case of salified forms, ester prodrugs, hydrates or solvate forms, combination products, or complexes. To reduce the number of published INNs, all these cases do not result in the publication of a different INN from that assigned to the active moiety but to the creation of a modified INN (INN_M), which does not necessarily need to be devised by the INN Expert Group and may be created by the manufacturer. Nonetheless, INN_Ms follow strict rules and lead to two- or three-word names, the first of which refers to the active principle, and the subsequent ones are attributable to the inactive moiety. Yet, in some specific cases, the radicals or groups composing the inactive parts are highly complex, and therefore the INN Expert Group selects a common name for them, and WHO publishes periodically a list of “names for radicals and groups” to be used in combination with an already assigned INN.¹⁰

For those drugs that allow it, and that have emerged more recently on the horizon, more rigid schemes have been devised to name substances. This is particularly true for monoclonal antibodies and for advanced gene and cell therapies. The naming of these substances has recently been reviewed elsewhere.¹¹

It has been advocated that INNs, stems, and radicals could be an excellent learning tool for students.

Figure 1 illustrates part of the value of this approach. Starting from a name, links can be made to other drugs of the same class, to drugs that share the same mechanism of action, to drugs that have similar structural features, to drugs that have the same therapeutic use, and to drugs that have been salified or esterified in similar manners, thus creating mnemonic aids. The INN Expert Committee recently also set up a School of INN to educate on how to construct, design, and interpret INNs.¹²

The name chosen by the Committee, which meets in Geneva twice a year, is known as the proposed INN (pINN) and is circulated among all stakeholders (Ministries of Health, industries, learned societies, etc.) that may object to the name for a number of reasons, including trademark infringements, similarity to other substances, or inappropriate meaning in a particular language. If an objection is raised, then the Committee is asked to re-evaluate the name, while if no objections are received by WHO, the name is, approximately a year later, declared a recommended INN (rINN). Each year two lists of the pINNs are published on the WHO Web site¹³ reflecting the choices made by the Committee. At present, list p123 has been published, and the first list dates to 1953. The latest rINN list published in 2020 is r84, and the difference of numeration between the proposed and the recommended list dates back to the initial years, in which several pINN lists were usually incorporated in the same rINN list. Since 2000, the two lists differ minimally, although some substances may see a deferral of their publication because of objections.

Regulatory agencies (e.g., European Medicines Agency (EMA), Food and Drug Administration (FDA), Pharmaceuticals and Medical Devices Agency (PMDA)) require that drugs submitted for their assessment are identified with an INN (or with a name given by the national naming agency), and therefore an INN is in most instances issued before the completion of clinical trials. While the WHO INN Expert Committee has a preference for drugs to be submitted after the beginning of Phase II trials, this is not always the norm, and an INN can be issued before or after, although it is very unlikely that applicants submit an application in the absence of encouraging data from Phase I trials. While receiving an INN for a drug increases the value of the portfolio, as it gives the feeling that it is closer to the market, companies are well-aware that the INN publication will allow competitors to know their intentions on the lead molecule being developed.

In our opinion, the pINN and rINN publications¹³ are a unique opportunity to scrutinize drug development as well as the recent trends in drug research and development (R&D) (as they precede marketing by a few years, see below), and, to our knowledge, this has been overlooked by scholars. Moreover, the INN publications have never been scrutinized from a medicinal chemistry viewpoint, unlike other chemical catalogues, such as FDA-approved drugs,^{14–21} including veterinary drugs,²² the Essential Medicines List (EML),²³ or molecules published in medicinal chemistry journals.^{24,25} In the present contribution, we concentrated on the chemical entities published in the INN lists in this millennium. Indeed, the publication of the INN represents at times the first public disclosure of the molecule and, while obviously not all substances that have been assigned an INN turn into approved drugs, these are all compounds that the industry has invested in. It is therefore an intermediate approach between analyzing only successful molecules (approved by the FDA or present on the EML) and analyzing discovery compounds (e.g., molecules

published in the *Journal of Medicinal Chemistry*), as it includes promising molecules, which might turn out to be successes or clinical failures. Our approach of concentrating on molecules in the last 20 years gives us the ability to investigate long-term trends in drug development and in medicinal chemistry, although we acknowledge that historical trends could be also investigated by taking all molecules published since 1953, or, on the contrary, features of newer molecules could be investigated by taking only the past few years.

Bird's-Eye View Analysis of the INNs of the Millennium. From 2000 to September of 2020, 3159 substances have received an INN published as a rINN. Note that r43 (2000) presents 43 substances whose publication had been delayed because of long-lasting objections and most likely date to the 1990s.²⁶ For our analysis, we used drugs listed in the recommended lists from 2000 to 2020 (r43 to r83), and we added the drugs present in the latest two proposed lists (p122 and p123), to include the most recent applications, leading to a total of 3456 molecules.

Not surprisingly, 2021 brought a surge of molecules, for which INNs were sought urgently, to be developed for the coronavirus disease of 2019 (COVID-19), and an extraordinary list (p124-COVID) was created for these drugs. These 25 drugs were not considered in our analysis. p124 is composed of monoclonal antibodies ($N = 10$), followed by RNA-based approaches ($N = 5$), organic compounds ($N = 6$), biologicals ($N = 3$), and advanced therapies ($N = 1$). Figure 2 depicts the

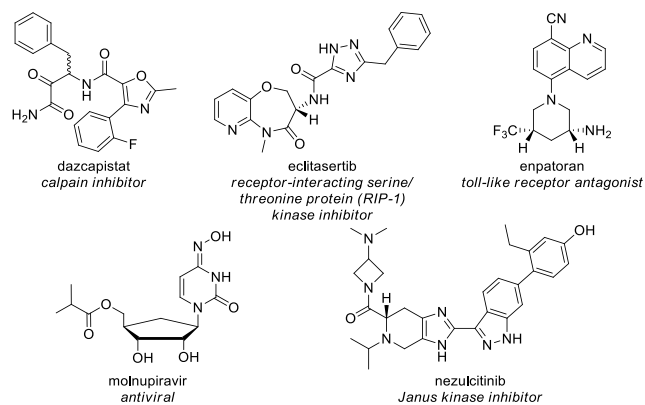


Figure 2. Structures of the five SCEs present in list p124 devoted to drugs potentially of use in the COVID pandemic.

five novel small chemical entities (SCEs) against severe acute respiratory syndrome (SARS) coronavirus 2 (CoV-2) found in this list. It is important to highlight, in this context, that the potential activity is self-declared by the applicant, and it is not up to the Committee to evaluate data that back efficacy or safety claims.

As mentioned above, lists are published on the WHO Web site.¹³ Compounds are described to uniquely identify them (e.g., IUPAC, amino acid sequence), and the lists also include other information. In the proposed lists, organic compounds are also described by their brute formula, the Chemical Abstracts Service (CAS) number and the chemical structure as well as by the broad class they belong to, while recommended lists lack some information, such as the CAS number and drug class. The description of the class is usually broad and heterogeneous: for example, it may refer to its chemistry (e.g., vitamin D analogue), to its pharmacological action (e.g.,

antiviral), or to its mechanism of action (e.g., toll-like receptor antagonist), following a common law on how the first compounds of the class were first described.

At first, we visually inspected all 3456 molecules and classified them in broad categories (see definitions in ref 27), which include “inorganic small chemical entities” ($N = 2$), “organic small chemical entities (SCEs)” ($N = 2018$), “biologicals” ($N = 411$), “monoclonal antibodies” (mAbs) ($N = 578$), “conjugates” ($N = 83$), “DNA/RNA-based therapies” ($N = 91$), “advanced therapies” (ATMPs) ($N = 129$), “polymers” ($N = 25$), “veterinary” ($N = 45$), and “mixtures” ($N = 3$). This classification required a number of compromises, and we do acknowledge that the great variety of compounds included means that some compounds could also have been classified differently. To represent this difficulty, 71 compounds were classified as “other”, as they did not meet our criteria of active principle (diagnostics, excipients, radio-diagnostics, sunscreens) or could not be reconducted to any other category (i.e., belzupacap sarotalocan).²⁸

Figure 3 shows how these categories have changed over time, both from a quantitative and qualitative point of view. Notice that the number of applications over time has dramatically increased, starting from 40 to 60 INNs for each volume in the 2000s and rising to a maximum of 163 in the last published list (p123). The number of SCEs that have received yearly an INN has roughly remained the same (between 70 and 120 depending on the year), although their weight percentage-wise has decreased significantly (from ~80% in volume r43 to ~40% in the last volume), in favor of mAbs, biologicals, ATMPs, conjugates, and DNA/RNA therapies. It is often stated that SCEs are being replaced by biotechnological products, but the analysis made, instead, suggests that biotech compounds are added, and are not a substitute, of traditional medicinal chemistry molecules.

We then decided to evaluate how many of the molecules that are published in the rINN lists eventually become approved drugs in the main global markets, U.S., Europe, or Japan, and how anticipated is the disclosure in the INN publications. The dotted line in Figure 3 shows the percentage of drugs published by the WHO in that particular year eventually approved by the FDA,²⁹ EMA,³⁰ or PMDA.³¹ As it can be observed, between 20% and 30% of drugs from each list are approved. Data from 2017 onward are significantly lower, as most drugs from those years have not been approved yet, and this shows that the INN anticipates by a few years drug approval.

As a whole 639 of 3456 drugs (19%) were approved by at least one agency by Nov 1, 2020. Briefly, 493 (14%) were approved by the FDA, 441 (13%) by the EMA, and 345 (10%) by the PMDA. Notice that the database we used for PMDA includes approval data from 2004, and therefore the number of approved drugs in Japan is underestimated. INNs, by definition, are used globally, and it is possible that some drugs that have been classified as “not approved” in our analysis have been approved elsewhere in the world. To support this statement, we searched different Web sites and databases³² to investigate whether the molecules depicted in the figures of this review and not listed in the FDA, EMA, or PMDA sites had been authorized elsewhere in the world, and we found that a number of these are indeed marketed (mainly in South America or the Far East). This therefore suggests that approved drugs are underestimated, as our analysis does not take into account a number of other markets. When classified

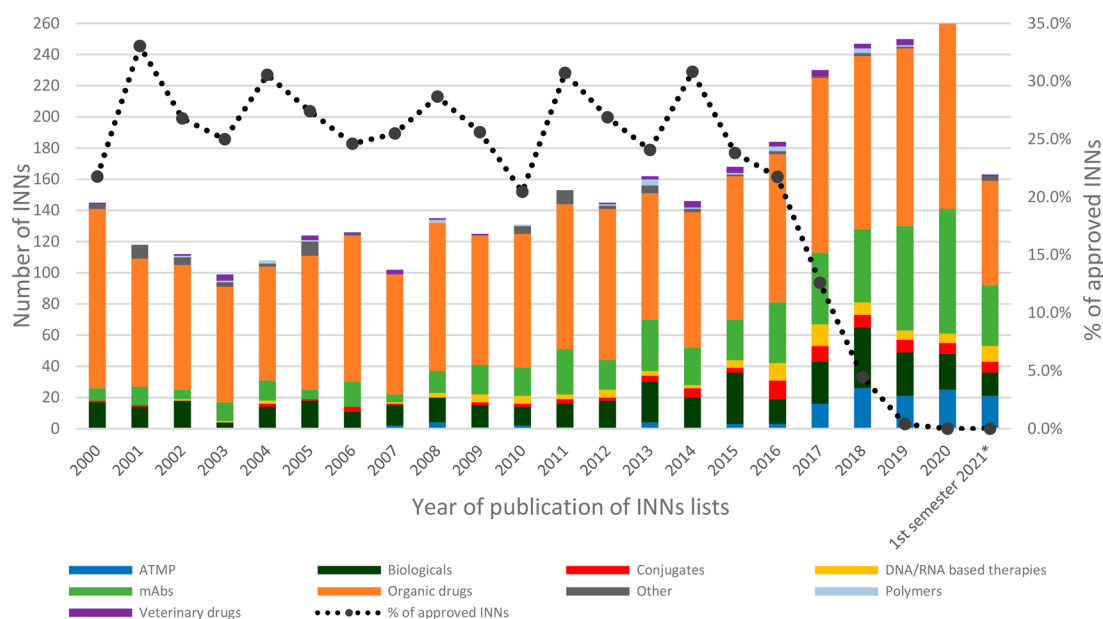


Figure 3. INNs issued between 2000 and 2021. Categories composed of less than 10 substances were excluded from the figure (two inorganics and three mixtures), while the same were included in the calculation of the percentage of approved INNs.

in broad categories, the percentage of approval for each class, considering U.S., Europe, and Japan, was found to be 24% for biologicals, 20% for SCEs, 15% for conjugates, 13% for mAbs, 12% for polymers, 10% for DNA/RNA based therapies, and 5% for ATMP, while no inorganic drug or molecule classified as a mixture was approved.

We then analyzed the time to approval in the three regulatory districts. This is displayed in the Kaplan–Meier plot in Figure 4. This manner of expressing data allows an estimation of events over time and, in this particular instance, the probability that an INN is approved after a given time. As it can be observed in panel A, it is estimated that ~22.5% of the total INNs will be approved by the FDA, and a slightly lower number will be approved by the EMA and PMDA. It is estimated that half of these drugs will be approved within four years in the U.S. and Europe, while it will take slightly longer for Japanese approval (this latter analysis might be nonetheless skewed by the loss of data between 2000 and 2004). Very few drugs are approved after 10 or more years from the INN publication. Notice that our analysis is performed on the rINN list, but a disclosure by the WHO in the pINN lists occurs approximately a year earlier. In brief, therefore, it is expected that, for each new list published, a fifth to a quarter of the drugs will be authorized and that half of these authorizations will occur within the first five years. We also investigated whether the different categories of drugs took different amounts of time to get approved (Figure 4B,C). Briefly, the median approval time for ATMPs was two years; for polymers and biologics it was three years, for small chemical entities it was four years, for mAbs it was five years, and for conjugates it was six years.

■ ANALYSIS OF THE MEDICINAL CHEMISTRY OF THE INNS IN THE LAST 20 YEARS

Having characterized the data set, we then decided to proceed evaluating the medicinal chemistry solely of SCEs listed in the INN publications. While we took inspiration from other reviews on chemical catalogues for a systematic analysis,^{14–22}

we also arbitrarily concentrated on particular aspects or molecules that caught our attention or on details that have recently emerged in the literature. The main goal of the review is to sprout interest in the INN catalogue, and we believe that this data set will be used in the future by others, which might concentrate on other aspects.

The decision to concentrate solely on SCEs meant that we excluded mixtures (i.e., plauracin, guaifylline, and latidectin A3+A4) and polymers ($N = 25$, e.g., paclitaxel poliglumex) from our analyses due to the impossibility of assigning them precise physicochemical characteristics, albeit recognizing that these might be of interest to the medicinal chemist. Last, conjugates were also excluded, as this group contained a number of medicines that do not pertain to the medicinal chemistry arena (e.g., mAbs+toxins/peptides, mAbs+radio-labeled isotopes). Yet, the same group also contained 59 antibody-SCE conjugates (antibody-drug conjugates (ADCs)), and we felt that this deserved to be flagged, given that the synthesis and characterization of chemical payloads are becoming an important field in medicinal chemistry and in anticancer therapy.^{33,34} As it can be observed in Figure 5, there are five different families of payloads that are more often used in conjugates: auristatin analogues (e.g., vedotin), maytansinoids (e.g., emtansine), pyrrolbenzodiazepine derivatives (e.g., talirine), irinotecan analogues (e.g., govitecan), and chelating agents (e.g., tiuxetan; to which a radioactive isotope is usually added). Some of the 21 payloads are conjugated with more than one mAb (e.g., labetuzumab govitecan and sacituzumab govitecan). In a similar manner, the same mAb can be conjugated with distinct payloads (e.g., trastuzumab emtansine and trastuzumab deruxtecan, both of which are currently approved).

The different payloads that compose a family (indicated by different INNs with the same stem, such as *-dotin*, *-xetan*, *-tansine*, and *-tecan*) may differ by their linker function or by chemical modifications in the core molecule (although these are not indicated in Figure 5, and a single representative member is depicted). Last, we also found three payloads, namely, ozogamicin, duocarmazine, and clezutoclax, that are

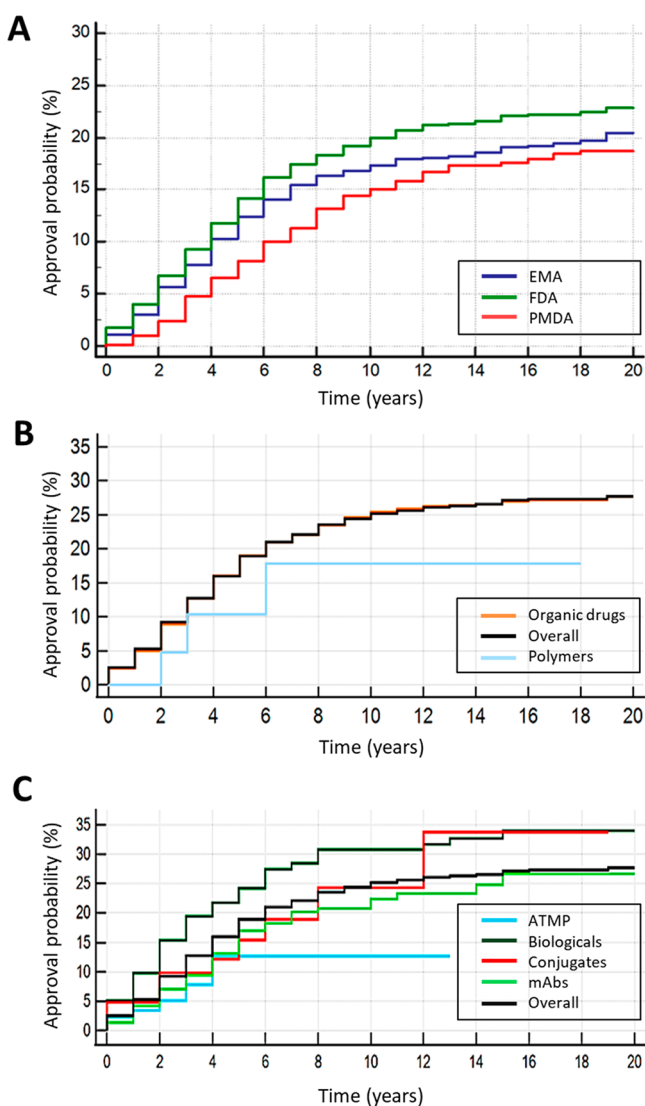


Figure 4. Probability of INN approval from the year of publication on the recommended INN list.

unique representatives, not being part of a family of molecules. Eight distinct payloads have been approved so far in 11 ADCs, with vedotin being approved in three different drugs and ozogamicin in two.

Trends in SCEs Drug Classes Developed. It is probably overambitious to prime the INN catalogue to investigate for diseases that have received the most attention from industry, as the description of drug use present in the database is given early on in development and may not accurately represent the later stages of development (i.e., industries may decide to repurpose a drug for a different disease while in clinical development; see apremilast in Figure 1 as an example). Furthermore, the descriptions are rather vague, and drugs undergo an extension of indications that would not be represented in the initial description.

For this reason, we decided instead to take a “nomenclature” approach and investigated which are the stems (e.g., drug classes) used in at least 10 INNs in the last 20 years. Table 1 lists the most frequent stems found in the 2018 SCEs, their definition as described in the stem book,⁹ and examples of approved drugs for each of them. As it can be observed, a few of these (-ine, -one, -fos-, -abine) are traditional and refer to

chemical moieties, but most stems nowadays refer to pharmacological targets or to drug use.

We then analyzed the frequency of these stems in two different periods, subdividing the SCEs in two subsets of a similar size: period A from 2000 to 2011 composed of 1038 SCEs and period B from 2012 to 2021 composed of 980 SCEs (Figure 6). A qualitative analysis reveals some striking changes. First, tyrosine kinase, cyclin-dependent kinase, serine/threonine kinase, and phosphatidylinositol 3-kinase inhibitors have seen a surge of interest among developers in the last 10 years. Almost 20% of all SCEs issued an INN since 2011 belong to one of these four classes, with tyrosine kinase inhibitors accounting for most of them. It is also of interest that, of the 37 serine/threonine kinase inhibitors, none has so far been approved, possibly owing to the recent investment. Possibly an unexpected finding, remaining in the oncology field, is that some chemotherapeutic classes have not gone out of fashion altogether (e.g., -bulin, -abine), while some others (-tecan) have completely lost popularity. Alongside cancer chemotherapeutic agents, a significant drop in INN assignments is also observed for the diabetes field and for some other drug targets (e.g., neurokinin receptors, vasopressin receptors, and endothelin receptors) that were popular in the first decade of the millennium. Of great interest is that a drug class exclusively developed for a single rare disorder (cystic fibrosis) has sprouted sufficient interest to issue 12 names in the last 10 years (-caftor). Overall, this analysis points, as expected, to a great dynamism of the pharmaceutical industry, with significant shifts in interest and development, which are bound to shape the drugs eventually approved for the market and the patients cared for.

Elemental Composition. We then analyzed the elemental composition of the 2018 SCEs in the INN lists. To speed up and automate the analysis process, a python protocol was prepared using a module designed for a chemoinformatic analysis (RDKit).³⁵ This protocol made it possible to extract the SMILES structures of INN lists from Pubmed and use them both to calculate chemical descriptors (see below) and to count the number of elements present in each molecule.

The elemental composition of the 2018 SCEs was compared to the one described in an analysis made on FDA-approved pharmaceuticals in 2014,²⁰ and the frequency of elements was compared between periods A and B (Figure 7).

Obviously, carbon, hydrogen, oxygen, and nitrogen (CHON) are present in the vast majority of compounds, as was the case for FDA-approved drugs.²⁰ Just after CHON, the INN list presents fluorine in just over 30% of compounds. This shows a change in medicinal chemistry strategies compared to the past, with fluorine taking over from chlorine and sulfur as the most frequent element after CHON.²⁰ While this could be somehow expected,³⁶ the extent of the use of fluorine might possibly surprise some. Indeed, a recent analysis made on FDA-approved drugs from 2015 to 2020 showed that 26% of drugs displayed fluorine,³⁷ already a very high number, but in our analysis we find 40% of drugs in the same time frame (compared to 17% in the first four years of the century) and 55% of fluorinated drugs issued a name in 2020.

No substantial differences were found on the percentage of approved drugs containing chlorine or fluorine, which have remained approximately similar between 2000 and 2020.

Iodine is present in only 19 compounds, a frequency very close to that of boron, which occurs in 14 molecules. The rate of approval of the iodine-containing molecules is lower (11%),

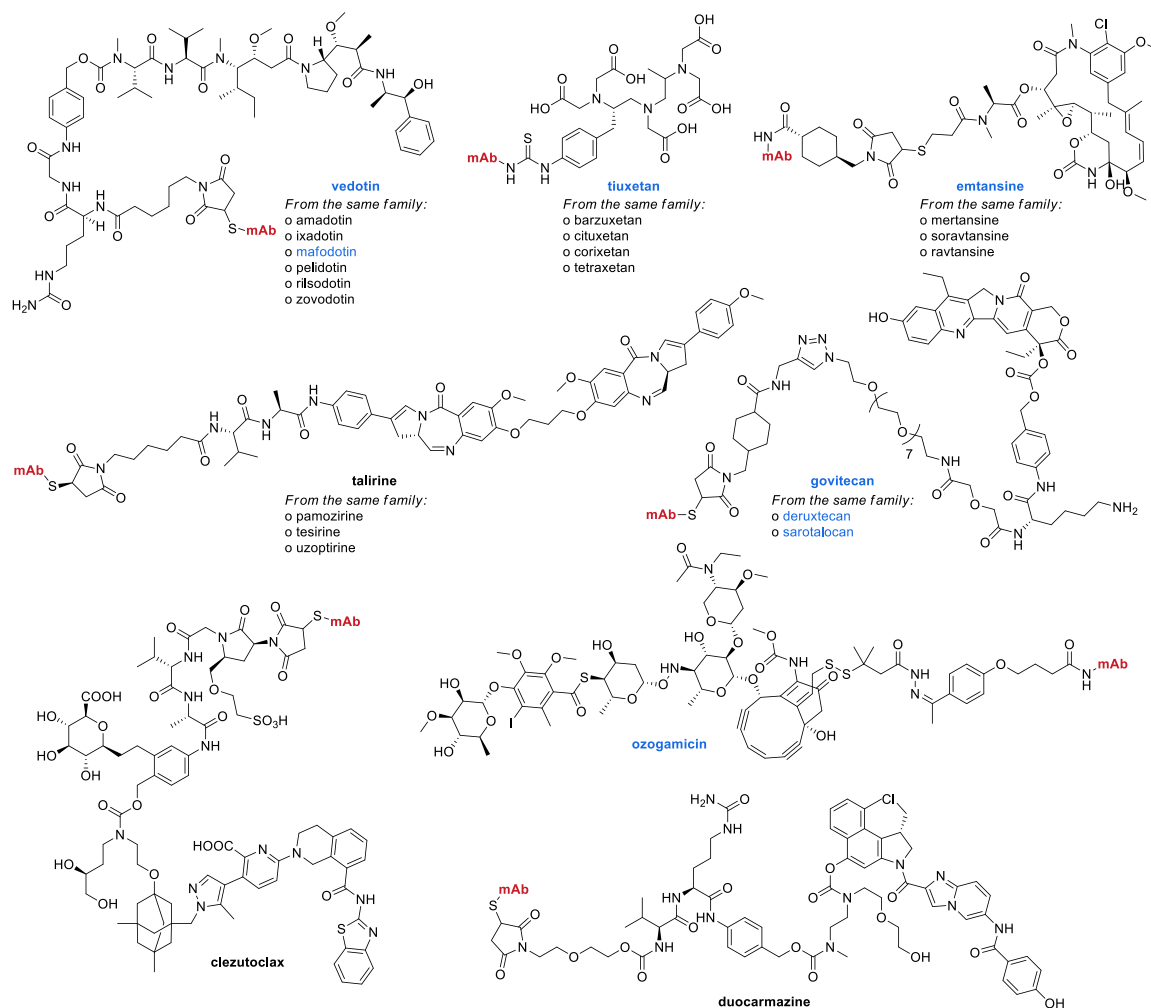


Figure 5. Representative structures of SCEs that are conjugated with mAbs. Payloads that are part of approved ADCs are highlighted in blue.

while the one of boron-containing molecules is significantly higher (43%) compared to the INN SCE benchmark (20%). Of the 14 boron-containing molecules, nine have been published in period B, demonstrating that the use of this element is increasing over time,^{38,39} although numbers remain small. While most of these molecules present boronic acid ($N = 8$), four contain five-membered boron heterocycles, and two contain six-membered boron heterocycles.

The remaining elements are platinum ($N = 5$), silicon ($N = 3$), iron ($N = 3$), palladium ($N = 2$), manganese ($N = 2$), bismuth ($N = 1$), tin ($N = 1$), and arsenic ($N = 1$), and molecules that contain these uncommon elements are represented in Figure 8.

Silicon is a proposed carbon isostere that improves the drug-likeness of bioactive compounds,^{40,41} and, similarly, deuterium has gained academic popularity as an isostere of hydrogen.^{42–45} The INN database lists three silicon-containing (Figure 8) and five deuterium-containing molecules (Figure 9). Some of these SCEs are initially developed with silicon or deuterium (e.g., cositecan or deucravacitinib), although it is interesting to note that most deuterium-containing molecules resulted from deuterium switches and that the cystic fibrosis drug ivacaftor has been the object of both a silicon and a deuterium switch (dirocaftor, Figure 8 and deutivacaftor, Figure 9). While for silicon no prefix has been established yet (but the syllable *-si-* is present in two of the three retrieved

drugs), deuterated molecules can be easily recognized by the prefix *deu-/deut-*, although this is not officially recognized.⁴⁶ The five compounds in Figure 9 represent the only deuterated INNs published so far. However, deuterated molecules are flourishing in the literature. An example of this is provided by the recently published dosimertinib,⁴⁷ a deuterated analogue of osimertinib (r75; 2016). It is likely that this name represents the preferred choice of the drug developer of what the compound should be called, but to our knowledge this is not an official INN and should not be used in the scientific literature, to avoid confusion (indeed, INN rules would suggest that the putative name should be deutosimertinib). Similar confusion is generated by the use of the name donafenib,⁴⁸ which is the deuterated analogue of sorafenib, as this molecule has never been issued an INN (which would putatively be deusorafenib according to the rules). It is important to note that the absence of these two drugs from our analysis is not a drawback of our data set, because these molecules cannot be marketed in Europe or the U.S. unless they reach the WHO for an official INN, which would probably not be the one they have so far used in the scientific literature.

Functional Groups. We next evaluated the occurrence of functional groups, taking inspiration from Ertl et al.,²⁴ who scanned medicinal chemistry journals, both for the most popular groups to search for and for the *in silico* approach to employ (we used the SMARTS listed in the Supporting

Table 1. Stems of SCEs Used at Least 10 Times in the Last 20 Years

ranking of stems in SCEs	stem	definition	example
#1 ^a	-ine	alkaloids and organic bases	atropine (according to WHO, ~17.5% of INNs have used this stem in lists p1-p119) ⁹
#2 ^b	-tinib	tyrosine kinase inhibitors	imatinib
#3 ^b	-stat-/-stat	enzyme inhibitors	atorvastatin, cobicistat
#3 ^b	vir-/-vir-/ -vir	antivirals (undefined group)	maraviroc, remdesivir
#5 ^b	gli-/-gli-	antihyperglycaemics	glibenclamide, dapagliflozin
#5 ^a	-one	ketones	naloxone (according to WHO, ~7.5% of INNs have used this stem in lists p1-p105) ⁹
#7 ^b	-imod	immunomodulators, both stimulant/suppressive and stimulant	fingolimod
#8	-sertib	serine/threonine kinase inhibitors	no drug is yet on the market
#9 ^c	fos-/-fos-	phosphorus derivatives; various pharmacological categories belonging to "fos", other than insecticides, anthelmintics, pesticides, etc.	foscarnet, sofosbuvir
#10	-lisib	phosphatidylinositol 3-kinase inhibitors, antineoplastics	idelalisib
#11 ^b	-ast	antiallergic or anti-inflammatory, not acting as antihistaminics	zafirlukast
#11 ^b	-tant	neurokinin (tachykinin) receptor antagonists	aprepitant
#13	-anib	angiogenesis inhibitors	nintedanib
#13	-bulin	antineoplastics; mitotic inhibitors, tubulin binders	eribulin
#13	-ciclib	cyclin dependent kinase inhibitors	palbociclib
#16 ^b	-abine	arabinofuranosyl derivatives; nucleosides antiviral or antineoplastic agents, cytarabine or azacitidine derivatives	capecitabine
#17	-entan	endothelin receptor antagonists	bosentan
#17	-oxacin	antibacterials, nalidixic acid derivatives	levofloxacin
#19	-conazole	systemic antifungal agents, miconazole derivatives	ketoconazole
#19 ^b	prost-/- prost-/- prost	prostaglandins	prostaglandins (veterinary use), misoprostol, latanoprost
#21	-caftor	cystic fibrosis transmembrane regulator (CFTR) protein modulators, correctors, and amplifiers	ivacaftor
#21	-nicline	nicotinic acetylcholine receptor partial agonists/agonists	varenicline
#23	-tecan	antineoplastics, topoisomerase I inhibitors	irinotecan
#23	-adol-/-adol	analgesics	tramadol
#23	-ciguat	guanylate cyclase activators and stimulators	riociguat
#23	-denoson	adenosine A receptor agonists	regadenoson
#27	-parib	poly-ADP-Ribose polymerase inhibitors	olaparib
#27	-terol	bronchodilators, phenethylamine derivatives	formoterol
#27	-vaptan	vasopressin receptor antagonists	tolvaptan

^aA number of other stems include this one (e.g., -terone for androgens). ^bThese stems possess substems that categorize in more detail. For example, -brutinib, -citinib, -ertinib, and -metinib group together tyrosine kinase inhibitors (-tinib) with the same target (Bruton kinase, Janus Kinase, EGFR, MAPK, respectively). ^cThe stem -fos is, instead, used for insecticides, anthelmintics, pesticides, etc., phosphorus derivatives.

Information using RDKit).³⁵ The analysis was then performed subdividing our data set in period A and period B, in order to make a comparison between the two decades (Figure 10).

As shown in Figure 10, the most abundant group is ether ($N = 962$), with approximately half of the INN molecules displaying this group, followed by amide, either linear or lactam ($N = 840$, 42%), aliphatic amine (primary, secondary, and tertiary; $N = 797$, 39%), and aromatic amine ($N = 633$, 31%). The 45% increase in aromatic amines in period B is not surprising considering that kinase inhibitors are particularly rich in this moiety and have significantly increased over time (e.g., -tinib, -sertib, -lisib, and -ciclib, see Figure 6).

When looking at the functional groups present in at least 60 molecules, no seismic change between the two periods occurs, although a few trends can be observed. Esters, for example, are usually thought to be groups not privileged in drug design, due to their hydrolytic instability. Despite this, ~10% of the INN molecules display this group, including prodrugs, even if a trend toward a reduction in period B is observed. Another little change is represented by sulfur-containing functional groups. While the number of SCEs displaying sulfur is the same in the

two periods (Figure 7), the use of this element has changed, with a decrease in the use of thioether, a little increase of sulfonamide, and a dramatic rise in sulfone,^{17,49} which is spread over different classes of INNs (e.g., in -pirdine, a proposed stem for serotonin receptor antagonists, and in -sertib and -tinib). While α,β -unsaturated carbonyl groups decrease in period B, α,β -unsaturated amides slightly increase in the same period, despite being well-known structural alerts. Interestingly, their marginal rise is consistent with the surge of covalent kinase inhibitors and follows the approval of afatinib,⁵⁰ ibrutinib,⁵¹ and osimertinib,⁵² three tyrosine-kinase inhibitors approved as antineoplastic drugs.

Both carboxylic acids and aliphatic amines have experienced a small decrease in their use over time, and we therefore decided to investigate the fraction of acidic, basic, neutral, and zwitterionic compounds (Figure 11) using an approach similar⁵³ to the one described by Charifson et al.⁵⁴ The distribution of the SCEs in the four categories was similar (neutral 47%; basic: 27%; acidic: 19%; zwitterionic: 7%) compared to that described in Charifson et al. with a slight

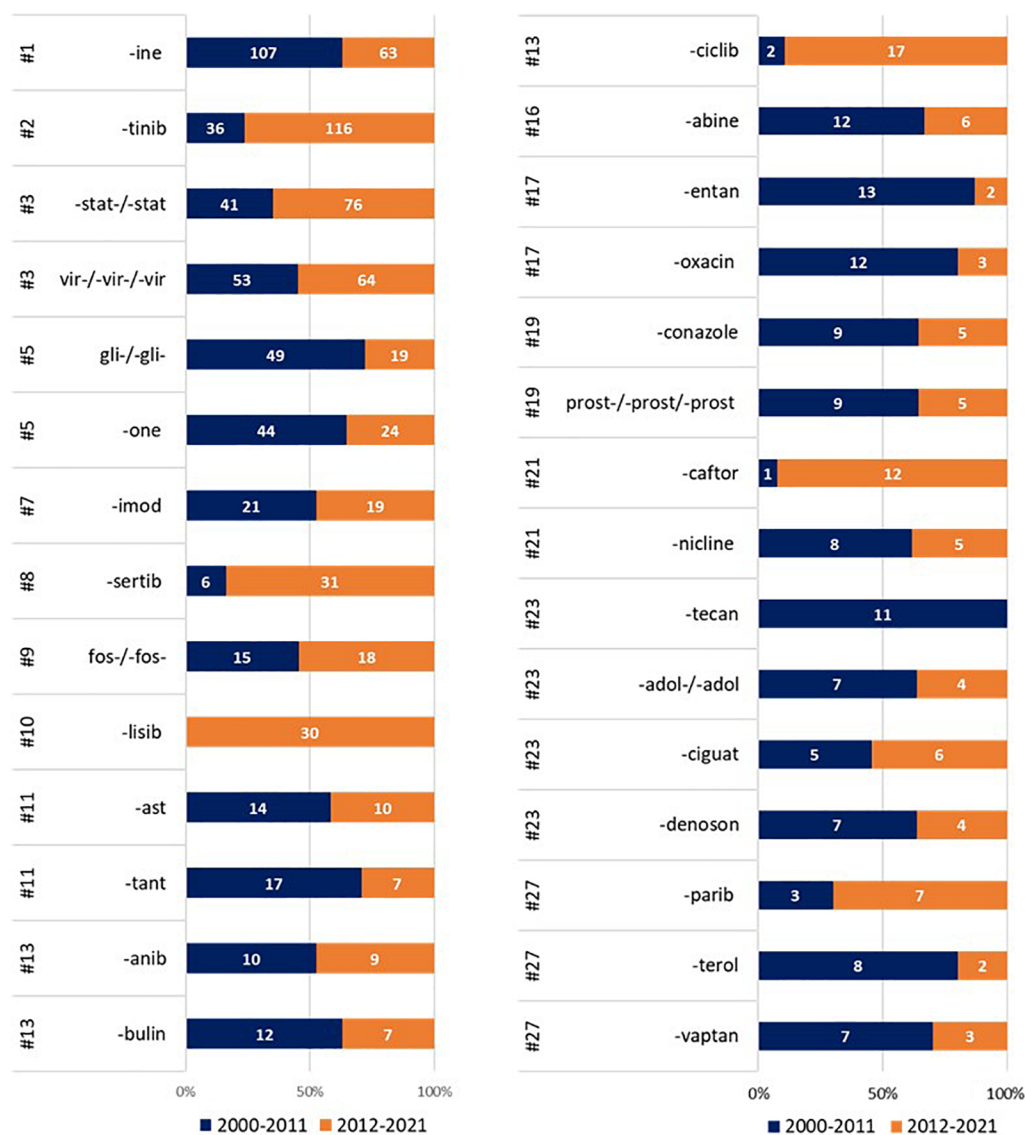


Figure 6. Frequency of the most common stems in period A (blue; 2000–2011; $N = 1038$) and period B (orange; 2012–2021; $N = 980$).

increase of neutral SCEs over time, rising from an average 43% of neutral drugs in period A to 51% in period B.

Interestingly, uncommon functional groups are starting to appear: indeed, we found four aldehydes and three sulfoximines (Figure 12). Despite the fact that aldehydes are considered structural alerts in medicinal chemistry⁵⁵ due to their high reactivity toward a vast array of nucleophiles, in our data set two aldehyde-containing SCEs (i.e., alcaftadine in period A, voxelotor in period B) out of four have been approved, confirming that the safety and utility of this functional group should be assessed on a case by case basis in R&D.⁵⁶ Sulfoximine, a neglected functional group in medicinal chemistry, has made its appearance in the lists since 2014. While none of the three SCEs has been approved so far, a recent article pointed at sulfoximine as an emerging group to further expand the toolbox of medicinal chemists.⁵⁷

The Ring Systems. After having analyzed the functional groups represented in our SCE data set, we investigated the nature of the ring systems by a visual inspection. To do so, we used a slightly modified approach to the original work by Taylor et al.⁵⁸ that considered a ring system as a complete ring

or rings formed by removing all terminal and acyclic linking groups.

From a methodological point of view, all rings and fused rings were retained, together with endocyclic bonds and exocyclic carbonyls, sulfonyls, imines, sulfynyls, and thiocarbonyls. Differently from the previous work,⁵⁸ SCEs displaying steroid core substructures⁵⁹ were grouped, while spiro groups were broken into their corresponding rings, as we dedicated a separate analysis to these substructures (see below). We removed macrocycles (number of atoms greater than 11; $N = 69$) as a potentially special case. Similarly to the approach described above, the resulting data set ($N = 1949$) was divided into two subsets: the first ($N = 994$) from 2000 to 2011 (named period A) and the second ($N = 955$) from 2012 to 2021 (named period B). This allowed us to capture the general occurrence of ring systems in the INN list as well as to compare two decades.

Overall, we found 362 distinct ring systems in period A and 419 in period B, with 224 and 257 ring systems used only in a single SCE. This is a small but significant sign that the chemical novelty is increasing and that the accessible chemical space is expanding and also shows the great chemical diversity

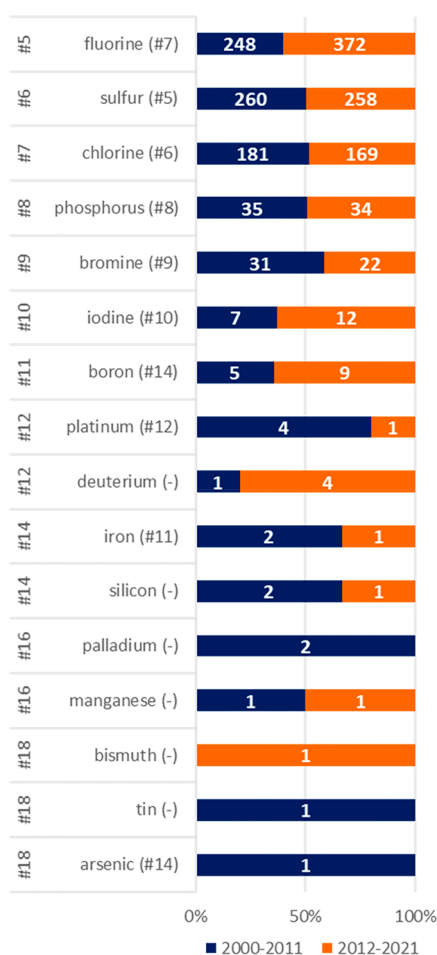


Figure 7. Frequency of elements beyond CHON in SCEs in period A (blue; 2000–2011; $N = 1038$) and period B (orange; 2012–2021; $N = 980$). In the INN lists, four molecules are listed as containing sodium, but a visual inspection revealed that this indicates the salified molecule. These molecules were thereby excluded from the figure. Numbers in parentheses refer to the relative ranking of elements in FDA-approved drugs.²⁰ *The occurrence of each element group refers to the number of SCEs containing at least one.

of INNs. This is further supported by the fact that Taylor et al., when analyzing the 1175 drugs found in the FDA Orange book (which included drugs approved between 1983 and 2012), found 351 distinct ring systems, of which 204 were present only once.

We then compiled the list of the top 30 most frequently used ring systems (Figure 13) and compared it with the one reported by Taylor et al.⁵⁹ The similarity between the two lists confirmed that medicinal chemists still rely on a subset of ring systems that have not changed in the last decades and are in part related to intrinsic properties and synthetic accessibility (e.g., benzene, pyridine, piperidine, cyclohexane). Yet, some substantial changes have occurred: azetidione, oxazole, and indazole are newcomers that were not present in the top 100 ring systems detected by Taylor et al., while pyrazole, pyrazine, and pyrrole have gained prominence. Interestingly, cephalosporins ($N = 4$ in period A; $N = 2$ in period B) and penicillins ($N = 2$ in period A; $N = 0$ in period B) are almost absent in our list, while they are well-represented in the FDA Orange Book, highlighting a progressive decrease of interest in β -lactam antibiotics. Similarly, the phenothiazine core, which is

featured in 11 molecules in the FDA-approved drugs, is almost absent in our data set ($N = 1$ in period A; $N = 1$ in period B).

We then compared the ring systems in the two periods. At first glance, it is impressive that most nitrogen-containing aromatic ring systems are more represented in period B. Pyridine, benzimidazole, and pyrazine have doubled or nearly doubled, pyrimidine has almost tripled, and indazole, despite the small numbers, is experiencing a significant increase over time. While we did not investigate this systematically, our impression is that this revolution is largely attributable to the advent of tyrosine kinase inhibitors (*-tinib*, see Figure 6) and, to a lesser extent, to cyclin-dependent kinase inhibitors (*-ciclib*, see Figure 6) and to the fact that they usually contain adenine mimetic scaffolds in their pharmacophore. An increase also occurs in nitrogen-containing aliphatic rings, including azetidione, pyrrolidine, and morphine, as well as in piperidine and piperazine, albeit to a less extent. This is not surprising, as all these substructures are more and more frequently introduced as solubility enhancers. Cyclopropane has significantly increased in period B, and this is mainly related to the recent trend to escape from flatland, as described below. It is interesting to note that, while 1,2,4-triazole shows a steady trend, 1,2,3-triazole, despite being outside the top 30 list, is experiencing a significant increase, thanks to the advent of the click chemistry approach and its exploitation in drug development ($N = 3$ in period A; $N = 12$ in period B).⁶⁰

Overall, when adding up the top 30 rings, we found an $\sim 25\%$ increase in ring systems (1385 in period A and 1710 in period B). The difference is remarkable, considering that periods A and B display a similar number of molecules ($N = 994$ in period A; $N = 955$ in period B), and we therefore believed that this imbalance deserved in-depth consideration. We wondered what counterbalanced this shortage of ring systems in period A, also in consideration that we observed a small increase in molecular weight (MW) in period B (see below), which per se cannot account for this alone. Once we discarded the possibility that there was a decrease in total phenyl rings in period B (total number of phenyl rings, period A = 897; period B = 880), we then hypothesized that the difference could be related to polycyclic fused ring systems. This was indeed consistent with the fact that, in period A, more steroids ($N = 23$ period A; $N = 14$ period B), irinotecan analogues (*-tecan*; see Figure 6), and taxanes (*-taxel*; see Figure 6) were present, and these, according to the methodology used, were not disconnected. While the polycyclic decrease might partially explain our observation, the numbers are too small to believe that this is the sole explanation. Indeed, it is likely that it is a plethora of small changes that add up to explain a large effect that we observe in ring systems. For example, we also found a difference in the number of exocyclic carbons between period A ($N = 7149$) and period B ($N = 6125$) that, while not being decisive, might partially contribute to fill the observed void.

Symmetric Compounds. While we were visually inspecting the data set, we were impressed by the abundance of symmetric compounds that were present ($N = 29$, referring to C_2 symmetry). It is interesting that there has been a doubling in the last two decades of these molecules, from 10 between 2000 and 2011 to 19 in the last 10 years. Two compounds are actually prodrugs bearing one inactivating portion and two identical molecules of the active principle (lodenafil carbonate, dinalbuphine sebacate). As expected, most of the remaining symmetric compounds are traditional twin drugs designed to

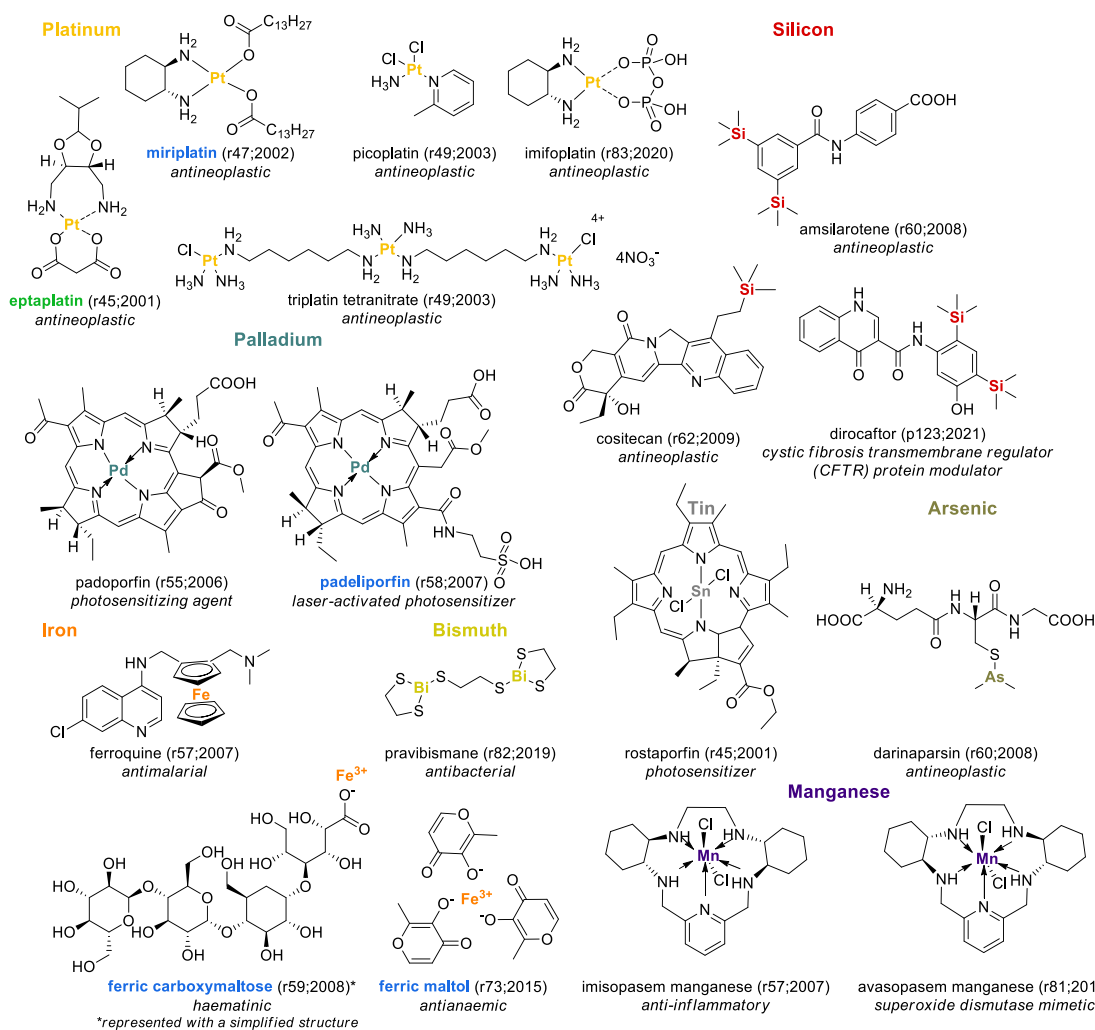


Figure 8. INN molecules containing unconventional elements. EMA-, FDA-, or PMDA-approved SCEs are highlighted in blue, and molecules found to be approved by other agencies are highlighted in green. The numbers in brackets indicate the rINN list and the year in which the molecule was published.

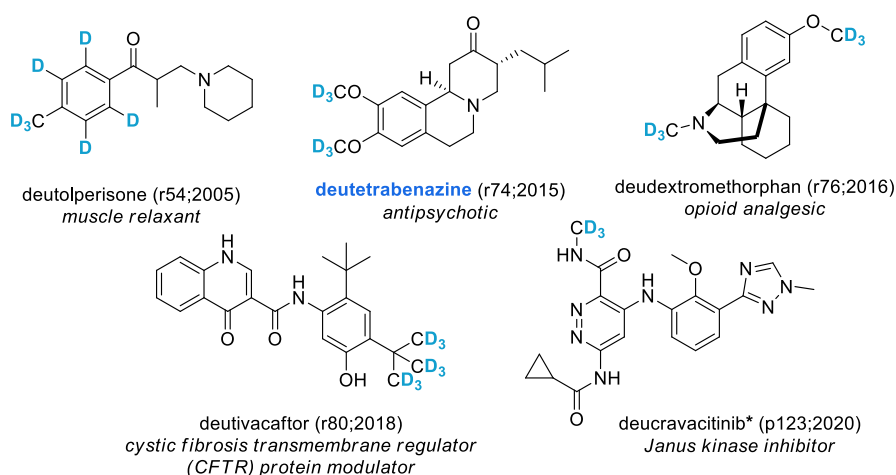


Figure 9. INN deuterated molecules. EMA, FDA, or PMDA approved SCEs are highlighted in blue. The numbers in brackets indicate the rINN list and the year in which the molecule was published. *To avoid a proliferation of stems and definitions, the WHO uses broad categories. Deucravacitinib well exemplifies that this choice is not flawless and might result in grouping molecules with distinct features. Indeed, both the stem (-citinib) and the definition place this drug as a Janus kinase (JAK) inhibitor, although the drug has been designed and developed as a specific inhibitor of TYK2, one of the four protein members of the JAK family (the others being JAK1, JAK2, and JAK3).

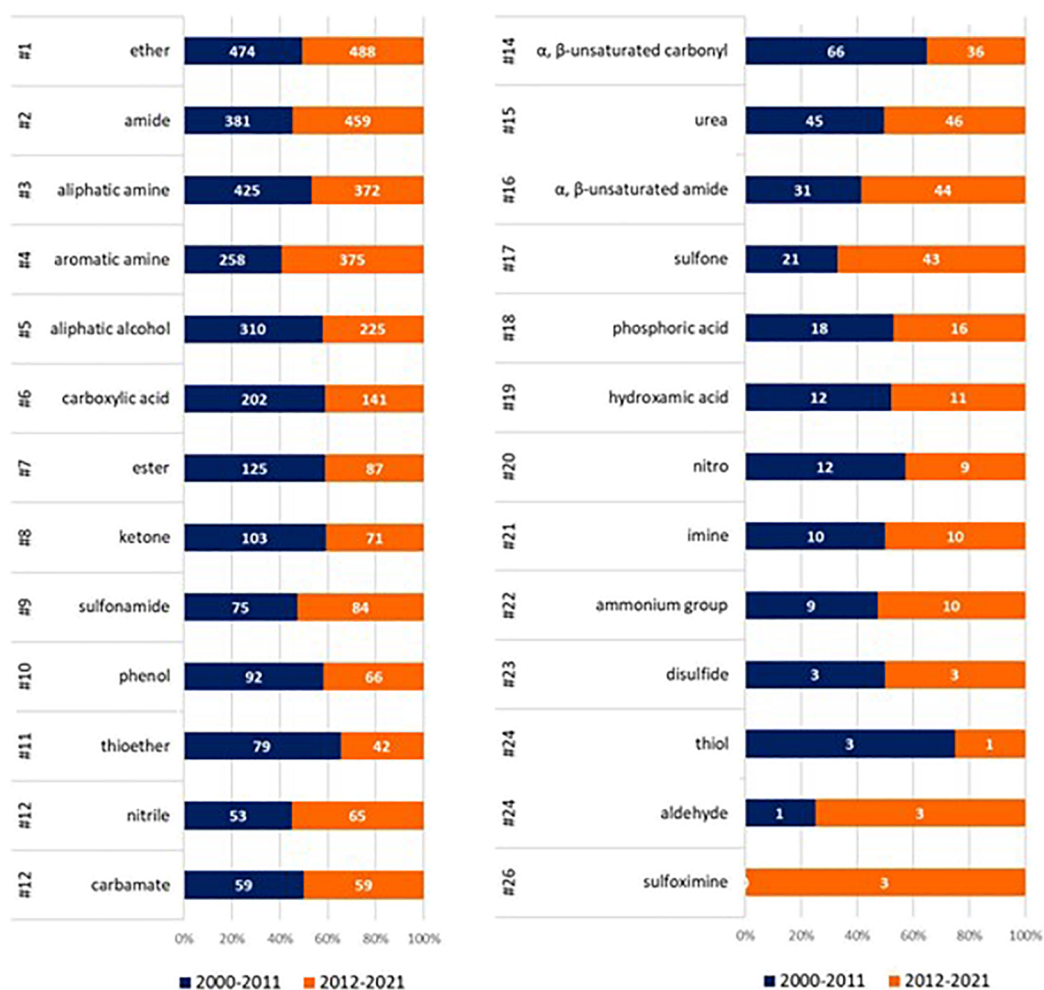


Figure 10. Frequency of selected functional groups in SCEs in period A (blue; 2000–2011: $N = 1038$) and period B (orange; 2012–2021: $N = 980$). *The occurrence of each functional group refers to the number of SCEs containing at least one.

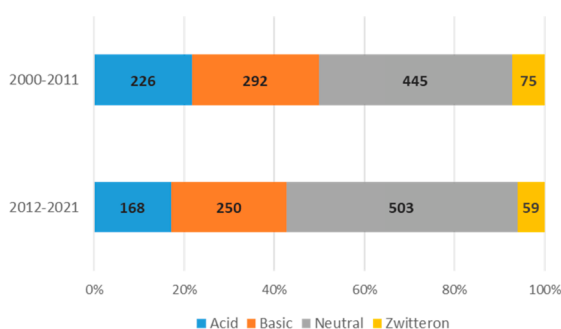


Figure 11. Distribution of neutral, acid, basic, and zwitterionic SCEs in period A (2000–2011: $N = 1038$) and period B (2012–2021: $N = 980$). * To define the acid/base properties of the SCEs, their ionization state at pH 7.4 was considered.

target proteins displaying dimeric structures.⁶¹ Among them, we found the NS5A inhibitors used as a hepatitis C virus (HCV) treatment (i.e., daclatasvir,⁶² ombitasvir,⁶³ pibrentasvir⁶⁴), diquafasol (a P2Y2 antagonist),⁶⁵ firibastat (an aminopeptidase A inhibitor),⁶⁶ and tegavivint (a TBL1 inhibitor).⁶⁷ Similar are those twin drugs intended to bind two identical target proteins (albitiazolium bromide,⁶⁸ miridesap⁶⁹). In this regard, it is interesting to see that the Chemically-Inducible Dimerization (CID) technology⁷⁰ has led to the development of rimiducid,⁷¹ a tacrolimus analogue

that behaves as a protein dimerizer, triggering the homodimerization of Fv-containing drug-binding domains of genetically engineered proteins such as the Caspase 9, Fas intracellular domain, and iCD40 receptor. Molecules where symmetry can be found as a means to complex metals are plerixafor⁷² (zinc) and elesclomol⁷³ (copper). Finally, two porphyrinic compounds are included (i.e., exeporfinium chloride, redaporfin). Figure 14 shows some selected examples of symmetric molecules.

Molecular Complexity. It is often believed that medicinal chemistry is becoming more complex to deal with the increasingly challenging drug targets and in line with increased synthetic accessibility. However, chemical complexity remains an elusive concept.⁷⁴ The ranking of compounds mainly depends on what parameters are used to describe complexity, but an unambiguous definition is yet to be defined. While several complexity descriptors have been reported,⁷⁴ what universal parameters should be used as proxies is still a controversial issue. In the literature, a broad range of possibilities is described, from simple topological or physicochemical descriptors to more complex indexes that combine several features into a single score.^{75,76} Since a universally accepted index does not exist and a systematic evaluation of the proposed alternatives is beyond the scope of this review, we decided to consider only some descriptors that are undoubtedly linked to complexity, well-aware that further work will be required in this field. It is our hope that this

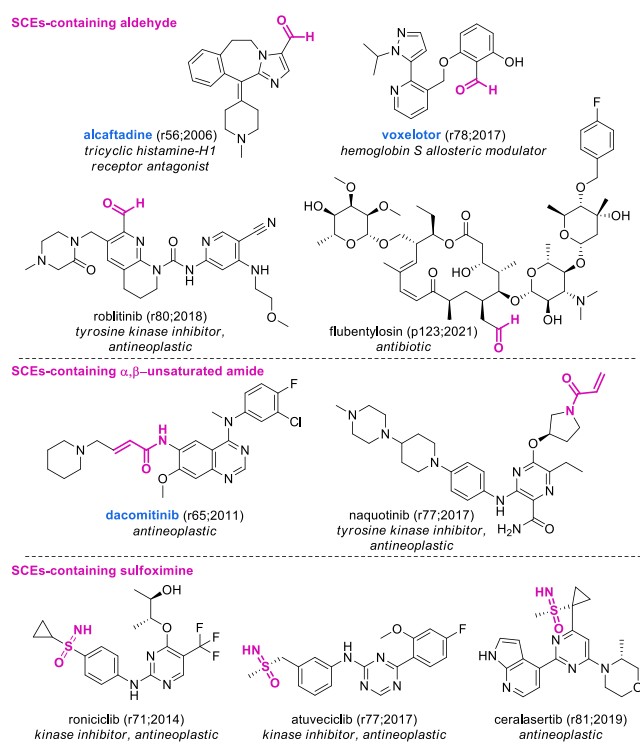


Figure 12. Structures of selected SCEs-containing α,β -unsaturated amide, aldehyde, or sulfoximine. EMA-, FDA-, or PMDA-approved SCEs are highlighted in blue. The numbers in brackets indicate the rINN list and the year in which the molecule was published.

catalogue might be used to prime molecular complexity by others, possibly using a historical approach from 1953.

The first element we investigated was whether molecules are getting larger: bigger and more complex molecules can access greater chemical space, can better complement the three-dimensional target binding site, and can possibly escape from existing Markush formulas. The MW³⁵ of INNs is slowly but steadily increasing over time both in the entire data set and in the restricted data set of approved drugs. This can be seen when evaluating the mean, which may be skewed by particularly large drugs, but also when evaluating the median (Figure 16B), which should exclude artifacts given by outliers. Briefly, there is an upward trend seen for all INN compounds (mean 435 in 2000–2004; mean 467 in 2016–2020). Furthermore, the MW of drugs in period B is \sim 10% higher compared to those in period A (Figure 16A). Approved INN drugs have a slightly higher MW (mean 478) compared to the overall INN compounds (mean 462) when the mean is compared, which we determined to compare our data with that of the literature. Our data (which necessarily include molecules that will fail along the way) are in contrast with the report that MW significantly decreases at each stage of development, from discovery to market,⁷⁶ but are in line with the hypothesis that a correlation exists between high MW and increased selectivity and reduced attrition rate.⁷⁷

The increment in MW in drugs is not a new trend, as noticed by Ivanenkov et al.,⁷⁶ that saw a strong difference between drugs approved in the past decade compared to drugs approved in the first 50 years of the previous century. This finding is also supported by other reports that show that discovery compounds⁷⁸ as well as marketed drugs and oral

drugs¹⁵ have experienced a consistent time-dependent increase, with a dramatic increment over the past decade.⁷⁶

As recently reported by Raymer et al.,⁷⁹ in spite of the fact that MW is increasing, lead-like drugs (drugs below MW 300) still represent a fruitful area of research and a therapeutic opportunity (2011–2016: 17% of drug approvals), and we find 14% of these molecules in our INN data set. To our great surprise, a substantial number of compounds has an MW below 150, of which five are approved worldwide (three in the main regulated areas we focused on), and one is contained in dietary supplements (Figure 15). Similarly, while 500 is considered the threshold for drug-like compounds, 29% of INN compounds are above this limit, and 11.5% of these are represented by macrocycles ($N = 69$).

In 2009, in their seminal paper Lovering et al. suggested that the medicinal chemistry community should “escape from the flatland” by increasing the fraction sp^3 ($Fsp^3 = \text{number of } sp^3\text{-hybridized carbons}/\text{total carbon count}$).⁸⁰ The authors proposed Fsp^3 as an important descriptor of molecular complexity: saturation makes molecules less planar and more structurally complicated, allowing them to access a greater chemical space, without significantly increasing MW. Disruption of molecular planarity is reflected in an increase of aqueous solubility,⁸⁰ target selectivity, and metabolic stability,⁸¹ and this should in principle improve the clinical success. This prediction is consistent with the finding that Fsp^3 increases through the five stages of development, going from 0.36 for discovery compounds to 0.47 for drugs on the market.⁸⁰ It has been suggested that a value of Fsp^3 that is higher than or equal to 0.42 is a suitable benchmark, and 84% of marketed drugs meet this requirement.⁸² In the entire INN data set the mean Fsp^3 is 0.41,³⁵ in accordance with this criterion and with the fact that the INN is usually requested prior to Phase II. Unlike the trend observed by Lovering, Fsp^3 is identical between approved and not approved drugs.

We then evaluated whether medicinal chemistry has escaped from flatland in the last 20 years but found that Fsp^3 is roughly similar between period A and B (Figure 16A), in line with what was reported by Ivanenkov et al. for launched drugs.⁷⁶ This somehow is a surprising finding, and when attempting to find a trend in the last 20 years, we found a small downward trend (Figure 16B), in analogy to what reported when analyzing discovery compounds published in the *Journal of Medicinal Chemistry* in the period of 1995–2009.⁷⁸ Despite this, small signs of change toward an enhanced spatial complexity can be observed at the granular level, and we concentrated on chirality, spirocyclic compounds, and small rings.

The chiral nature of drugs has an impact on molecular complexity and correlates with the chance of approval in the process of R&D.⁸⁰ The number of chiral centers³⁵ increases through the different steps of clinical stages, reaching the maximum in the approved drugs, where 64% of them have at least one stereocenter. We therefore analyzed the counts of stereocenters in our data set, and we found that 60% of our SCEs have at least one stereocenter, a percentage comparable to that of Phase II compounds,⁸⁰ with a constant trend in the calculated period, 2000–2020. On the contrary, in the same period, the average number of stereocenters per SCE is slightly decreased. We did not undertake a classification of the origin of compounds in our data set (e.g., synthetic, natural, etc.), and therefore we are unable to determine whether a change in the origin has an impact on our findings. Eight SCEs have a number of stereocenters equal to or higher than 20, with a

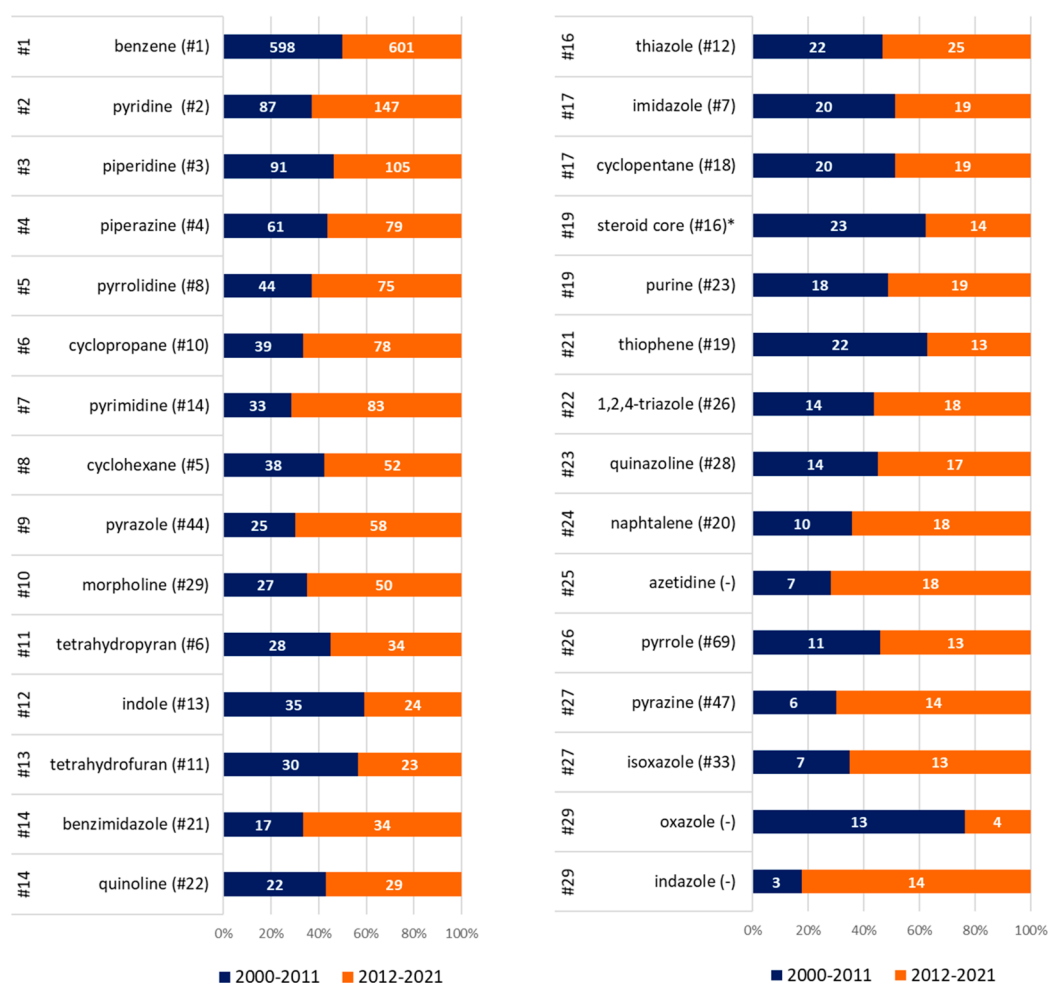


Figure 13. Frequency of selected ring systems in SCEs in period A (blue; 2000–2011; $N = 994$) and period B (orange; 2012–2021; $N = 955$). Numbers in parentheses refer to the relative ranking of ring systems in FDA-approved drugs.⁵⁸ The occurrence of each ring system was determined by counting the number of SCEs containing at least one. *Steroids were grouped⁵⁹, while in Taylor et al. they were subdivided, considering the different ring systems. With our methodology, steroids would have risen to second place in Taylor et al.⁵⁸

maximum of 35 stereocenters in evernimicin (r44; 2000), followed by pixatimod (r79; 2018) with 29 stereocenters. Chirality is also a regulatory issue, and the FDA-guidelines for the development of chiral active substances published in 1992 prompted the development of pure stereoisomers.^{83–85} It is therefore not surprising that only 10% of all the SCEs containing at least one chiral center are named as a mixture of stereoisomers. Finally, the approval percentage of chiral SCEs is 27%, slightly above the mean of 20%, with no differences among the subset of pure stereoisomers (28%) and mixtures of stereoisomers (26%).

We next focused our attention on spirocyclic motifs that, similarly to quaternary carbon stereocenters, provide an opportunity to project substituents in all three dimensions. Their exploitation in medicinal chemistry has been aided by the recent advances in synthetic strategies that allow access to these substructures.⁸⁶ A significant increase in their use is evident when considering the publications with the keyword “spiro” in the medicinal chemistry field, as done by Hiesinger et al.:⁸⁷ a progressive increase can be found starting from 2000. It is therefore not surprising that, when we analyzed the occurrence over time of spirocyclic motifs³⁵ in the INN, more than double spirocycle-containing SCEs have been assigned an INN in period B (Figure 16C) with 62% of them being

assigned a name after 2014 (Figure 16D). The impact of these scaffolds is exemplified by two drugs approved in 2020, oliceridine and risdiplam (r76; 2016 and r80; 2018, respectively).

Besides spirocyclic substructures, aliphatic three- and four-membered rings³⁵ contribute to the overall Fsp³. While cyclopropanes have been exploited for many years in drug discovery, cyclobutanes, azetidines, and oxetanes have become popular only recently, an uptrend that goes hand-in-hand with the advent of synthetic methods that allow their incorporation. Bauer et al. have recently described the occurrence of these cycles in the patent literature (2009–2019) and have found that cyclopropane is the most-used small ring, followed by cyclobutane, azetidine, and oxetane.⁸⁸ Albeit numbers are small, it is evident that, with the exception of oxetane, all rings have increased over time also in the INN lists, with cyclopropane experiencing the most significant increase (Figure 16F). Interestingly, cyclobutane is not represented in the period of 2000–2007 and makes its entrance in 2008, occurring 15 times since then. Overall, the occurrence of these small cycles in period B compared to period A has almost doubled (Figure 16E). Figure 17 shows some representative examples of molecules bearing spirocyclic scaffolds and small rings.

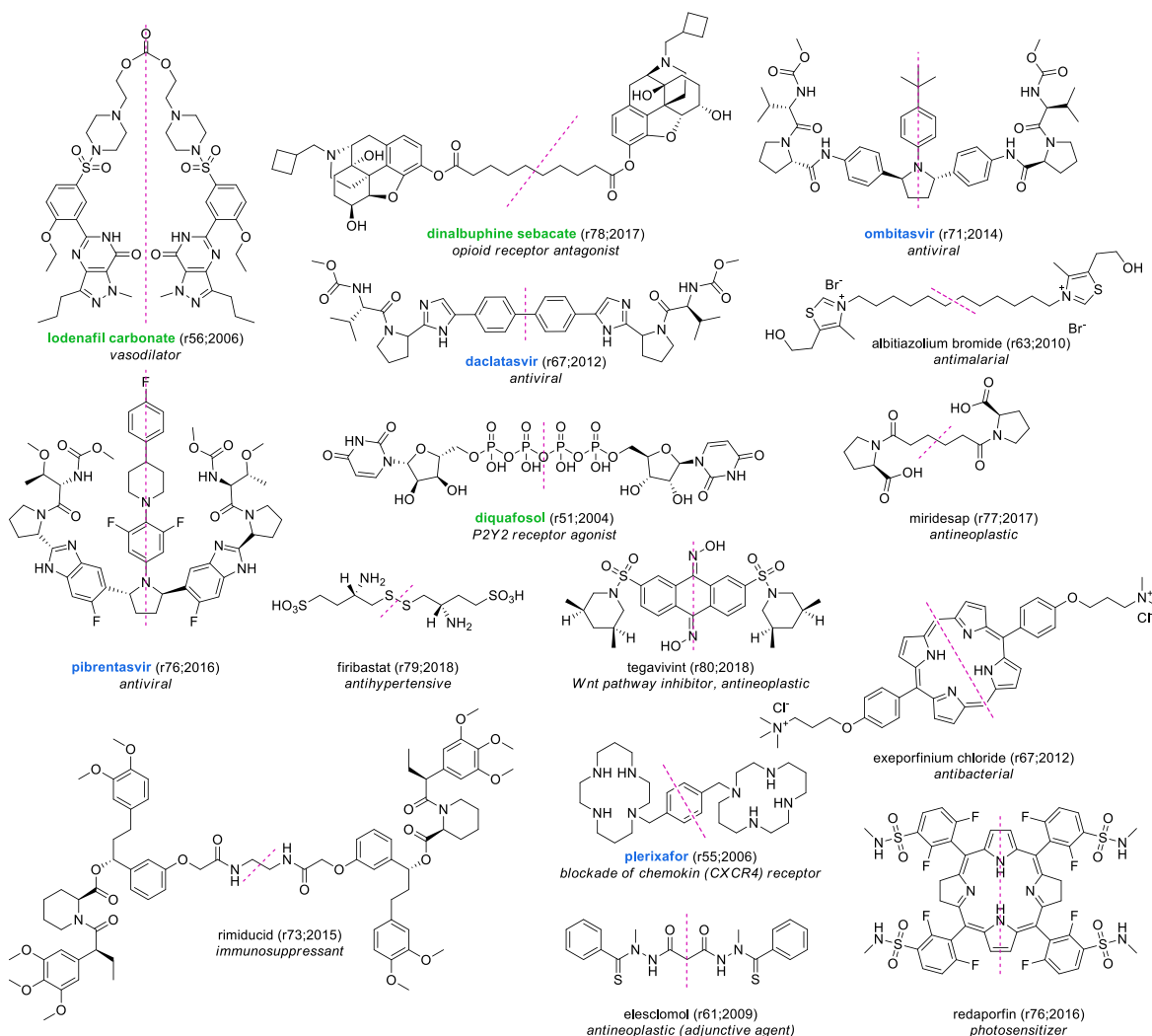


Figure 14. Structures of selected symmetric compounds. EMA-, FDA-, or PMDA-approved SCEs are highlighted in blue, and molecules found to be approved by other agencies are highlighted in green. The numbers in brackets indicate the rINN list and the year in which the molecule was published.

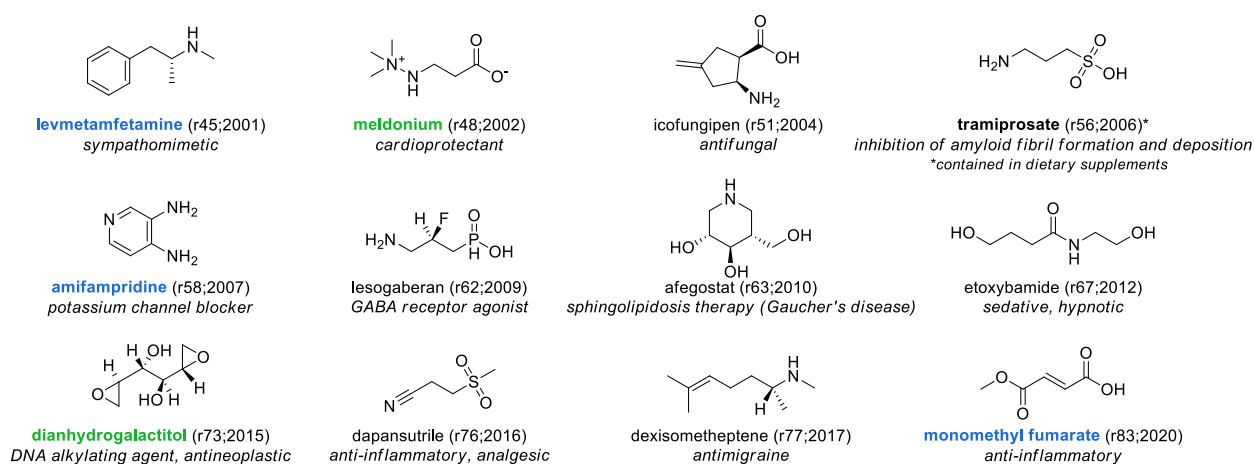


Figure 15. Structures of the 12 SCEs with MW below 150 Da. EMA-, FDA-, or PMDA-approved SCEs are highlighted in blue, and molecules found to be approved by other agencies are highlighted in green. The numbers in brackets indicate the rINN list and the year in which the molecule was published. It is interesting to note that, at times, the INN does not reflect the common name by which a compound is known (which in the case of tramiprosate would be homotaurine).

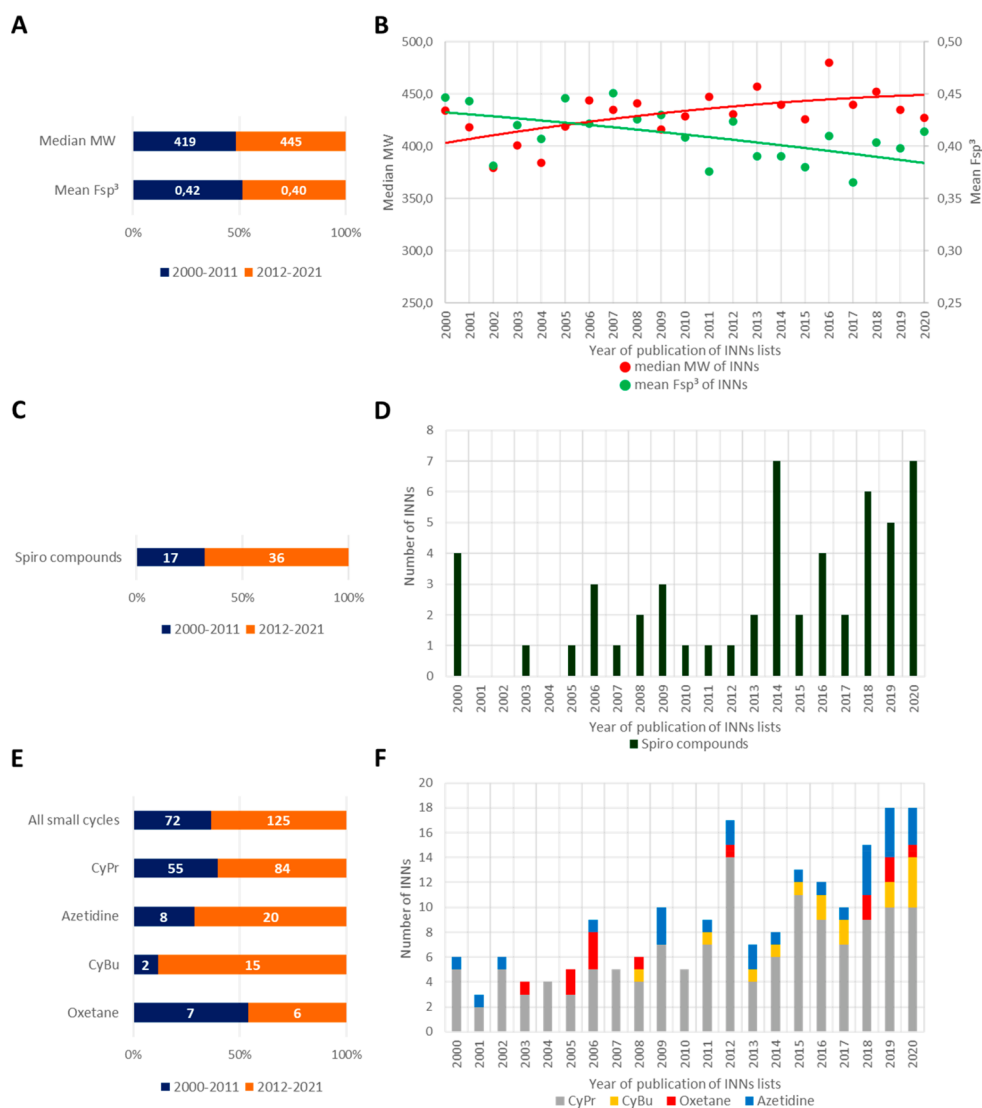


Figure 16. (A) Median MW and mean Fsp³ in period A (2000–2011; $N = 1038$) and period B (2012–2021; $N = 980$) and (B) per year; number of spirocycles (C) in periods A and B and (D) per year and number of SCE-containing small cycles (E) in periods A and B and (F) per year. *The occurrence of spiro and small cycles refers to the number of SCEs containing at least one. Benzofused systems were excluded from the count of spiro compounds. Lactones and lactams were excluded from the count of small cycles.

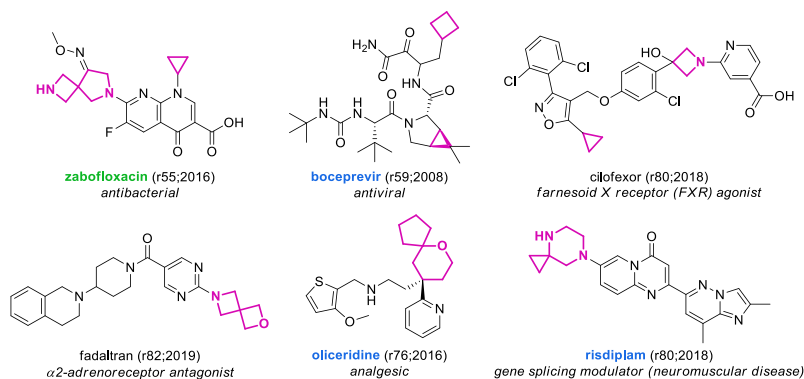


Figure 17. Selected SCEs containing spirocyclic scaffolds and small rings. EMA-, FDA-, or PMDA-approved SCEs are highlighted in blue and molecules found to be approved by other agencies are highlighted in green. The numbers in brackets indicate the rINN list and the year in which the molecule was published.

Overall, therefore, the data from the INN lists show a clear trend in increased MW. Synthetic feasibility has counterbalanced the putative advantages of escaping from flatland, and

this is evident when considering the chemical reactions used in current medicinal chemistry, which is overpopulated by amide

bond formation, Suzuki-Miyaura coupling, and S_NAr reactions.⁸⁹

CONCLUSIONS

A main objective of this review was to disseminate the INN nomenclature schemes and allow readers to recognize the features of a name (stems, infixes, radicals, prefixes, and suffixes) that are important to detect some of the characteristics of the medicine, which may be important for teaching medicinal chemistry and pharmacology, as well as for clinical practice.

A second objective was to evaluate the ~2000 small molecules that received an INN in the last 20 years, (i) paralleling some of the previous analyses done on different chemical catalogues by others and (ii) comparing the two decades of this century.

Many before us, possibly more skilled, have scrutinized chemical catalogues to describe the essence of pharmaceutical R&D. This has been done on FDA-approved drugs,^{14–21} patents by the pharmaceutical industry,^{88,90} publications by academics,²⁴ and drugs under clinical investigation.⁹¹

Now we propose to use the publications related to the INN, freely accessible on the WHO Web site,¹³ as a new chemical catalogue, that presents different features, possibly advantageous, over other databases. First, compared to patents and publications, these molecules should represent lead compounds that industry has decided to invest in at the clinical level. Second, these publications anticipate the market by approximately four to five years and therefore can foresee trends and changes before the databases on approved medicines. These advantages are obviously counterbalanced by the fact that ~75% of the molecules in this catalogue will never be approved.

A peculiar finding of our analysis is that medicinal chemistry is largely unchanged, despite the significant modification in drug targets that we describe. Therefore, at first glance, one might be led to believe that medicinal chemistry is largely conservative. Yet, a closer look under the lens shows subtle evolutionary changes that make the baseline constantly drift. Molecular weight is a good example of this, and, while the constant drift might be related to the change in drug targets, it should be noted that this trend follows similar transformations observed since 1910.⁷⁶ Small signs that might pave the path to more significant changes in the future, or might just represent a historical coincidence, also creep up in our analyses. In particular, we identified a cluster of approved boron-containing drugs, a few molecules that incorporate deuterium, and an increased exploitation of small rings (e.g., cyclopropane, cyclobutane, azetidine) and of spirocyclic scaffolds. These small changes are going hand-in-hand with the advent of new synthetic strategies^{87–89} that ease access to such structural features.

Two other results caught our attention: first, among the elements beyond CHON, fluorine is being used above expectations and, in the latest INN publications, is represented in ~40% of the molecules (with a peak of 55% in the publications from 2020); second, in the last 10 years there has been an increase in the use of N-containing ring systems, with some heterocycles significantly contributing to this (e.g., pyridine, pyrimidine, and pyrazole). Fluorine use had been predicted,³⁵ and it is highly likely that the increase in N-containing ring systems is partly linked to the fact that the kinome has become a popular target. We also observed a

strong decrease in beta-lactam containing drugs, in steroids, and in phenothiazines, most likely representing a change in therapeutic areas.

While this database will be of great use to scrutinize drugs under development and to inform on trends in medicinal chemistry, it is unlikely to ever allow a determination of the features that confer success to a molecule, as too many variables, alongside chemistry, influence this aspect. In this respect, an acknowledgment of a referenced paper⁷⁸ quotes an anonymous referee that provides an enlightening truth: “*Drugs have to survive multiple hurdles followed by attritional factors including toxicity, clinical safety, efficacy in humans, differentiation, market viability, organizational strategy, regulatory approval and acceptance by payers. It is not a surprise that drug-likeness resists accurate description.*”

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jmedchem.1c00181>.

Data set of the INN molecules and list of SMARTS used for searching functional groups (XLSX)

AUTHOR INFORMATION

Corresponding Author

Armando A. Genazzani – Department of Pharmaceutical Sciences, Università del Piemonte Orientale, 28100 Novara, Italy; orcid.org/0000-0003-1923-7430; Phone: 0039-0321-375827; Email: armando.genazzani@uniupo.it

Authors

Marta Serafini – Department of Pharmaceutical Sciences, Università del Piemonte Orientale, 28100 Novara, Italy; orcid.org/0000-0002-5305-8359

Sarah Cargnin – Department of Pharmaceutical Sciences, Università del Piemonte Orientale, 28100 Novara, Italy

Alberto Massarotti – Department of Pharmaceutical Sciences, Università del Piemonte Orientale, 28100 Novara, Italy; orcid.org/0000-0001-9306-8845

Gian Cesare Tron – Department of Pharmaceutical Sciences, Università del Piemonte Orientale, 28100 Novara, Italy

Tracey Piralì – Department of Pharmaceutical Sciences, Università del Piemonte Orientale, 28100 Novara, Italy; orcid.org/0000-0003-3936-4787

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jmedchem.1c00181>

Author Contributions

[‡](M.S., S.C., A.M.) These authors contributed equally. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

Biographies

Marta Serafini received her degree in Pharmaceutical Chemistry and Technology in 2015 and her Ph.D. in 2021 from the Università del Piemonte Orientale (Novara, Italy). At present, she is a post-Doc in Medicinal Chemistry funded by Fondazione AIRC (Associazione Italiana per la Ricerca sul Cancro). Her fields of interest focus on the discovery of bioactive compounds towards unconventional targets.

Sarah Cargnin received her degree in Pharmacy in 2011 and her Ph.D. in Pharmaceutical and Food Biotechnologies in 2016 at the Università del Piemonte Orientale (Novara, Italy). At present, she is a post-Doc in Pharmacology at the Dept. of Pharmaceutical Sciences in Novara. Her main research interest focuses on personalized medicine.

Alberto Massarotti received his M.Sc. in medical and pharmaceutical biotechnology in 2006 and his Ph.D. in 2009 from the Università del Piemonte Orientale. He is currently an Associate Professor of medicinal chemistry at the same university. In 2014 he cofounded the Start Up IXTAL. His field of interest covers the development and application of novel computational methodologies to discover new lead compounds.

Gian Cesare Tron is Full Professor of medicinal chemistry at the Università del Piemonte Orientale (Novara, Italy). He received his degree in Pharmaceutical Chemistry and Technology in 1994 and his Ph.D. in Organic Chemistry in 2001 from the Università di Torino (Italy) under the supervision of Prof. G. Appendino. He has spent sabbatical leaves in the laboratories of Prof. J. Zhu (Institut de Chimie des Substances Naturelles, Gif-sur-Yvette, France), Prof. V. Aggarwal (School of Chemistry, Bristol, UK), and Prof. V. Fokin (The Scripps Research Institute, La Jolla, United States). His research interests concern the discovery of new multicomponent reactions and their application in the field of medicinal chemistry.

Tracey Pirali received her Ph.D. in Science of Bioactive Compounds from the Università del Piemonte Orientale (Novara, Italy) in 2007. In 2016 she cofounded the Start Up ChemiCare. Currently, she is Associate Professor of Medicinal Chemistry at the Università del Piemonte Orientale, and her research interests include the discovery of small molecules targeting IDO and SOCE.

Armando A. Genazzani received his medical degree from the Università di Catania (Italy) and obtained his D.Phil. in Pharmacology from the University of Oxford (UK). At present, he is Head of Department and Full Professor of Pharmacology at the Università del Piemonte Orientale (Novara, Italy). He has participated in the 22nd Expert Committee for Essential Medicines in 2019, has been previously a member of the WHO INN Expert Committee, is part of the Technical and Scientific Committee of the Italian Medicines Agency (AIFA), and is a member of the Committee for Medicinal Products for Human Use (CHMP) of the European Medicines Agency (EMA).

ACKNOWLEDGMENTS

M.S. is supported by Fondazione AIRC (Associazione Italiana per la Ricerca sul Cancro) fellowship for abroad (Rif. 25278). S.C. holds a grant from Ministero della Salute (Bando Finalizzata 2016-GR-2016-02363848). T.P. and A.A.G. hold grants from Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR) (PRIN 2017 Nos. 2017WJZ9W9 and 2017CBNCYT), from Telethon Foundation (GGP19110), and from Fondazione AIRC (IG201821842). The authors thank O. Mapelli for initial contributions in data mining. A.G. thanks Dr. R. Balocco, the entire INN Secretariat, and all his colleagues at the WHO for the rewarding experience serving the INN Expert Group between 2014 and 2019.

ABBREVIATIONS USED

ADC, antibody-drug conjugate; ANSI, American National Standards Institute; ATC, Anatomical Therapeutic Chemical Classification System; ATMP, advanced therapies; CAS, Chemical Abstract Service; CHON, carbon, hydrogen, oxygen and nitrogen; CID, Chemically-Inducible Dimerization; EGFR,

epidermal growth factor receptor; EML, Essential Medicines List; EMA, European Medicines Agency; FDA, Food and Drug Administration; Fsp³, fraction sp³; HCV, hepatitis C virus; INCI, International Nomenclature Cosmetic Ingredient; INN, international nonproprietary name; INN, international nonproprietary name modified; ISO, International Standards Organization; IUPAC, International Union of Pure and Applied Chemistry; mAb, monoclonal antibody; MAPK, mitogen-activated protein kinase; MW, molecular weight; NSSB, nonstructural protein 5B; pINN, proposed INN; PMDA, Pharmaceuticals and Medical Devices Agency; rINN, recommended INN; SCE, small chemical entities; U.S., United States; WHO, World Health Organization

REFERENCES

- (1) (a) Wikipedia. https://en.wikipedia.org/wiki/List_of_languages_by_total_number_of_speakers (accessed 2020-12-21). (b) *Ethnologue: Languages of the World*, 23rd ed.; Eberhard, D. M., Simons, G. F., Fennig, C. D., Eds; SIL International: Dallas, TX, 2020. Online version: <http://www.ethnologue.com> (accessed 2020-12-21).
- (2) World Health Organization. https://apps.who.int/iris/bitstream/handle/10665/86211/WHA3.11_eng.pdf (accessed 2020-12-21).
- (3) Aronson, J. K. "Where name and image meet"—the argument for "adrenaline". *BMJ*. **2000**, *320*, 506–509.
- (4) Aronson, J. K. Medication errors resulting from the confusion of drug names. *Expert Opin. Drug Saf.* **2004**, *3*, 167–172.
- (5) Personal Care Products Council. <https://www.personalcarecouncil.org/resources/inci/> (accessed 2020-12-21).
- (6) Kramer, C. V.; Zhang, F.; Sinclair, D.; Olliaro, P. L. Drugs for treating urinary schistosomiasis. *Cochrane Database Syst. Rev.* **2014**, *8*, CD000053.
- (7) World Health Organization. https://www.who.int/medicines/services/inn/inn_expert_group/en/ (accessed 2020-12-21).
- (8) World Health Organization. <https://www.who.int/medicines/services/inn/publication/en/> (accessed 2020-12-21).
- (9) Stem Book 2018. <https://www.who.int/publications/i/item/who-emp-rht-tsn-2018-1> (accessed 2021-01-27).
- (10) Radical Book 2015. <https://www.who.int/medicines/services/inn/RadicalBook2015.pdf?ua=1> (accessed 2021-01-27).
- (11) Robertson, J. S.; Chui, W. K.; Genazzani, A. A.; Malan, S. F.; López de la Rica Manjavacas, A.; Mignot, G.; Thorpe, R.; Balocco, R.; Rizzi, M. The INN global nomenclature of biological medicines: A continuous challenge. *Biologicals* **2019**, *60*, 15–23.
- (12) School of International Non-proprietary Names. <https://extranet.who.int/soinn/> (accessed 2020-12-21).
- (13) World Health Organization. <https://www.who.int/medicines/publications/druginformation/innlists/en/> (accessed 2020-12-21).
- (14) Das, P.; Delost, M. D.; Qureshi, M. H.; Smith, D. T.; Njardarson, J. T. A survey of the structures of US FDA approved combination drugs. *J. Med. Chem.* **2019**, *62*, 4265–4311.
- (15) Shultz, M. D. Two decades under the influence of the rule of five and the changing properties of approved oral drugs. *J. Med. Chem.* **2019**, *62*, 1701–1714.
- (16) Delost, M. D.; Smith, D. T.; Anderson, B. J.; Njardarson, J. T. From oxiranes to oligomers: Architectures of U.S. FDA approved pharmaceuticals containing oxygen heterocycles. *J. Med. Chem.* **2018**, *61*, 10996–11020.
- (17) Scott, K. A.; Njardarson, J. T. Analysis of US FDA-approved drugs containing sulfur atoms. *Top Curr. Chem.* **2018**, *376*, 5.
- (18) DeGoey, D. A.; Chen, H. J.; Cox, P. B.; Wendt, M. D. Beyond the rule of 5: Lessons learned from AbbVie's drugs and compound collection. *J. Med. Chem.* **2018**, *61*, 2636–2651.
- (19) Vitaku, E.; Smith, D. T.; Njardarson, J. T. Analysis of the structural diversity, substitution patterns, and frequency of nitrogen heterocycles among U.S. FDA approved pharmaceuticals. *J. Med. Chem.* **2014**, *57*, 10257–10274.

(20) Smith, B. R.; Eastman, C. M.; Njardarson, J. T. Beyond C, H, O, and N! Analysis of the elemental composition of U.S. FDA approved drug architectures. *J. Med. Chem.* **2014**, *57*, 9764–9773.

(21) Ilardi, E. A.; Vitaku, E.; Njardarson, J. T. Data-mining for sulfur and fluorine: an evaluation of pharmaceuticals to reveal opportunities for drug design and discovery. *J. Med. Chem.* **2014**, *57*, 2832–2842.

(22) Scott, K. A.; Qureshi, M. H.; Cox, P. B.; Marshall, C. M.; Bellaire, B. C.; Wilcox, M.; Stuart, B. A. R.; Njardarson, J. T. A Structural analysis of the FDA Green Book-approved veterinary drugs and roles in human medicine. *J. Med. Chem.* **2020**, *63*, 15449–15482.

(23) Serafini, M.; Cargnin, S.; Massarotti, A.; Pirali, T.; Genazzani, A. A. Essential medicinal chemistry of essential medicines. *J. Med. Chem.* **2020**, *63*, 10170–10187.

(24) Ertl, P.; Altmann, E.; McKenna, J. The most common functional groups in bioactive molecules and how their popularity has evolved over time. *J. Med. Chem.* **2020**, *63*, 8408–8418.

(25) Brown, D. G.; Boström, J. Where do recent small molecule clinical development candidates come from? *J. Med. Chem.* **2018**, *61*, 9442–9468.

(26) These substances have been labeled with an asterisk in the Supporting Information database, since their proposed INNs were submitted by WHO several years before their publication in the rINN list.

(27) On the one hand, the category of biologicals includes proteins, including those in which some amino acid residues have been modified, low molecular weight heparins, purified hormones, and small peptides. On the other hand, when a short peptide (up to four-five amino acids) was present in an otherwise SCE (e.g., vintafolide, r69;2013), this substance was inserted in the SCE category. Note that the number of mAbs is slightly overestimated, as often, in the same rINN list, the applicant applied for the mAb itself together with a second application for the relative conjugate (30 examples in the analyzed lists). Pegylation, indicated as either the prefix “peg” or the suffix “pegol”, has not been considered in the classification of the substances. RNA/DNA-based therapies broadly comprise antisense, siRNA, and mRNA-based therapies. Advanced therapies comprise gene and cell therapies, oncolytic viruses, or bacteria. The category of veterinary drugs has been compiled following the indication reported in the proposed INN volumes, and not on chemical or biological structures, but it must be recognized that also drugs not specifically categorized as veterinary during the INN application process may be then developed for veterinary use. Surprisingly, the lists also include a few sunscreens, possibly because these are not considered cosmetic ingredients for the U.S. legislation.

(28) Belzupacap sarotalocan is described in list p122 as “A modified human papillomavirus (HPV) type 16-derived empty nanoparticle, 55 nm in diameter conjugated to approximately 200 molecules of a phthalocyanine-based photosensitizer (sarotalocan group). Each nanoparticle is comprised of 72 capsomeres, made of 5 molecules of modified viral capsid protein L1 [P78 > R, T176 > N, D273 > T, N285 > T, S288 > N, T353 > P, T389 > S] and one molecule of viral capsid protein L2”.

(29) FDA-approved drugs. <https://www.accessdata.fda.gov/scripts/cder/daf/> (accessed 2020-09-02).

(30) EMA EPARs for human and veterinary medicines. <https://www.ema.europa.eu/en/medicines/download-medicine-data> (accessed 2020-09-02).

(31) PMDA-approved drugs. <https://www.pmda.go.jp/english/review-services/reviews/approved-information/drugs/0002.html> (accessed 2020-09-02).

(32) To this aim, we used the following Web sites and databases: (a) adisinsight.springer.com (accessed 2021-01-25). (b) www.wikipedia.org (accessed 2021-01-25). (c) [newdrugapprovals.org](https://www.newdrugapprovals.org) (accessed 2021-01-25).

(33) Bargh, J. D.; Isidro-Llobet, A.; Parker, J. S.; Spring, D. R. Cleavable linkers in antibody-drug conjugates. *Chem. Soc. Rev.* **2019**, *48*, 4361–4374.

(34) Casi, G.; Neri, D. Antibody-Drug Conjugates and small molecule-drug conjugates: opportunities and challenges for the

development of selective anticancer cytotoxic agents. *J. Med. Chem.* **2015**, *58*, 8751–61.

(35) RDKit, 2020.09.1 version: Open-Source Cheminformatics Software. <https://www.rdkit.org>. (accessed 2020-08-07).

(36) Meanwell, N. A. Fluorine and fluorinated motifs in the design and application of bioisosteres for drug design. *J. Med. Chem.* **2018**, *61*, 5822–5880.

(37) Bhutani, P.; Joshi, G.; Raja, N.; Bachhav, N.; Rajanna, P. K.; Bhutani, H.; Paul, A. T.; Kumar, R. U.S. FDA approved drugs from 2015–June 2020: a perspective. *J. Med. Chem.* **2021**, *64*, 2339–2381.

(38) Song, S.; Gao, P.; Sun, L.; Kang, D.; Kongsted, J.; Poongavanam, V.; Zhan, P.; Liu, X. Recent developments in the medicinal chemistry of single boron atom-containing compounds. *Acta Pharm. Sin. B* [Online early access]. **2021** DOI: [10.1016/j.apsb.2021.01.010](https://doi.org/10.1016/j.apsb.2021.01.010)

(39) Fernandes, G. F. S.; Denny, W. A.; Dos Santos, J. L. Boron in drug design: Recent advances in the development of new therapeutic agents. *Eur. J. Med. Chem.* **2019**, *179*, 791–804.

(40) Ramesh, R.; Reddy, D. S. Quest for novel chemical entities through incorporation of silicon in drug scaffolds. *J. Med. Chem.* **2018**, *61*, 3779–3798.

(41) Kleemiss, F.; Justies, A.; Duvinage, D.; Watermann, P.; Ehrke, E.; Sugimoto, K.; Fugel, M.; Malaspina, L. A.; Dittmer, A.; Kleemiss, T.; Puylaert, P.; King, N. R.; Staubitz, A.; Tzschentke, T. M.; Dringen, R.; Grabowsky, S.; Beckmann, J. Sila-ibuprofen. *J. Med. Chem.* **2020**, *63*, 12614–12622.

(42) Pirali, T.; Serafini, M.; Cargnin, S.; Genazzani, A. A. Applications of deuterium in medicinal chemistry. *J. Med. Chem.* **2019**, *62*, 5276–5297.

(43) Cargnin, S.; Serafini, M.; Pirali, T. A primer of deuterium in drug design. *Future Med. Chem.* **2019**, *11*, 2039–2042.

(44) Liu, J. F.; Harbeson, S. L.; Brummel, C. L.; Tung, R.; Silverman, R.; Doller, D. A decade of deuteration in medicinal chemistry. *Annu. Rep. Med. Chem.* **2017**, *50*, 519–542.

(45) Gant, T. G. Using deuterium in drug discovery: leaving the label in the drug. *J. Med. Chem.* **2014**, *57*, 3595–3611.

(46) A hydrogen-bearing molecule can potentially exist in several deuterated forms. For example, the d_9 -analogue of ivacaftor has been named deutivacaftor, but a d_{18} -analogue was also reported; therefore, it is possible that, in the future, a prefix alongside the suffix *deu-* will be added to identify the presence of more than one deuterated analogue of a hydrogenated molecule.

(47) Meng, Y.; Yu, B.; Huang, H.; Peng, Y.; Li, E.; Yao, Y.; Song, C.; Yu, W.; Zhu, K.; Wang, K.; Yi, D.; Du, J.; Chang, J. Discovery of dosimertinib, a highly potent, selective, and orally efficacious deuterated EGFR targeting clinical candidate for the treatment of non-small-cell lung cancer. *J. Med. Chem.* **2021**, *64*, 925–937.

(48) Bi, F.; Qin, S.; Gu, S.; Bai, Y.; Chen, Z.; Wang, Z.; Ying, J.; Lu, Y.; Meng, Z.; Pan, H.; Yang, P.; Zhang, H.; Chen, X.; Xu, A.; Liu, X.; Meng, Q.; Wu, L.; Chen, F. Donafenib versus sorafenib as first-line therapy in advanced hepatocellular carcinoma: An open-label, randomized, multicenter phase II/III trial. *J. Clin. Oncol.* **2020**, *38*, 4506–4506.

(49) Feng, M.; Tang, B.; Liang, S. H.; Jiang, X. Sulfur containing scaffolds in drugs: Synthesis and application in medicinal chemistry. *Curr. Top. Med. Chem.* **2016**, *16*, 1200–1216.

(50) Santos, E. S.; Hart, L. Advanced squamous cell carcinoma of the lung: current treatment approaches and the role of afatinib. *Oncotargets Ther.* **2020**, *13*, 9305–9321.

(51) Brown, J. R. PCI-32765, the first BTK (Bruton's tyrosine kinase) inhibitor in clinical trials. *Curr. Hematol. Malig. Rep.* **2013**, *8*, 1–6.

(52) Yver, A. Osimertinib (AZD9291)—a science-driven, collaborative approach to rapid drug design and development. *Ann. Oncol.* **2016**, *27*, 1165–1170.

(53) To calculate pK_a values, the pK_a plugin of Marvin's cxcalc program was used. Marvin 20.19.0, ChemAxon: <https://www.chemaxon.com> (accessed 2020-08-07).

- (54) Charifson, P. S.; Walters, W. P. Acidic and basic drugs in medicinal chemistry: A perspective. *J. Med. Chem.* **2014**, *57*, 9701–9717.
- (55) Alves, V.; Muratov, E.; Capuzzi, S.; Politi, R.; Low, Y.; Braga, R.; Zakharov, A. V.; Sedykh, A.; Mokshyna, E.; Farag, S.; Andrade, C.; Kuz'min, V.; Fourches, D.; Tropsha, A. Alarms about structural alerts. *Green Chem.* **2016**, *18*, 4348–4360.
- (56) Gampe, C.; Verma, V. A. Curse or cure? A perspective on the developability of aldehydes as active pharmaceutical ingredients. *J. Med. Chem.* **2020**, *63*, 14357–14381.
- (57) Mäder, P.; Kattner, L. Sulfoximines as rising stars in modern drug discovery? Current status and perspective on an emerging functional group in medicinal chemistry. *J. Med. Chem.* **2020**, *63*, 14243–14275.
- (58) Taylor, R. D.; MacCoss, M.; Lawson, A. D. G. Rings in drugs. *J. Med. Chem.* **2014**, *57*, 5845–5859.
- (59) Steroid core substructures that were grouped are the following: hexadecahydro-1*H*-cyclopenta[*a*]phenanthrene; 2,3,6,7,8,9,10,11,12,13,14,15,16,17-tetradecahydro-1*H*-cyclopenta[*a*]phenanthrene; 7,8,9,11,12,13,14,15,16,17-decahydro-6*H*-cyclopenta[*a*]phenanthrene; 6,7,8,9,10,11,12,13,14,15,16,17-dodecahydro-1*H*-cyclopenta[*a*]phenanthren-3(2*H*)-one; 6,7,8,11,12,13,14,15,16,17-decahydro-1*H*-cyclopenta[*a*]phenanthren-3(2*H*)-one; 6,7,8,9,10,11,12,13,14,15,16,17-dodecahydro-3*H*-cyclopenta[*a*]phenanthren-3-one.
- (60) Serafini, M.; Piralì, T.; Tron, G. C. Click 1,2,3-triazoles in drug discovery and development: from the flask to the clinic? *Adv. Heterocycl. Chem.* **2020**, [Online early access]. DOI: 10.1016/bs.aihch.2020.10.001.
- (61) Contreras, J.-M.; Sippl, W. Homo and Heterodimer Ligands the Twin Drug Approach. In *The Practice of Medicinal Chemistry*, 3rd ed.; Wermuth, C. G., Ed.; Elsevier, 2008; pp 380–414.
- (62) Gamal, N.; Gitto, S.; Andreone, P. Efficacy and safety of daclatasvir in hepatitis C: an overview. *J. Clin. Transl. Hepatol.* **2016**, *4*, 336–344.
- (63) Flisiak, R.; Flisiak-Jackiewicz, M. Ombitasvir and paritaprevir boosted with ritonavir and combined with dasabuvir for chronic hepatitis C. *Expert Rev. Gastroenterol. Hepatol.* **2017**, *11*, 559–567.
- (64) Liu, X.; Hu, P. Efficacy and safety of glecaprevir/pibrentasvir in patients with chronic HCV infection. *J. Clin. Transl. Hepatol.* **2021**, *9*, 125–132.
- (65) von Kügelgen, I. Pharmacology of P2Y receptors. *Brain Res. Bull.* **2019**, *151*, 12–24.
- (66) Marc, Y.; Boitard, S. E.; Balavoine, F.; Azizi, M.; Llorens-Cortes, C. Targeting brain aminopeptidase A: a new strategy for the treatment of hypertension and heart failure. *Can. J. Cardiol.* **2020**, *36*, 721–731.
- (67) Nomura, M.; Rainusso, N.; Lee, Y. C.; Dawson, B.; Coarfa, C.; Han, R.; Larson, J. L.; Shuck, R.; Kurenbekova, L.; Yustein, J. T. Tegavivint and the β -catenin/ALDH axis in chemotherapy-resistant and metastatic osteosarcoma. *J. Natl. Cancer Inst.* **2019**, *111*, 1216–1227.
- (68) Wein, S.; Maynadier, M.; Bordat, Y.; Perez, J.; Maheshwari, S.; Bette-Bobillo, P.; Tran Van Ba, C.; Penarete-Vargas, D.; Fraisse, L.; Cerdan, R.; Vial, H. Transport and pharmacodynamics of albitalazolium, an antimalarial drug candidate. *Br. J. Pharmacol.* **2012**, *166*, 2263–2276.
- (69) Pepys, M. B. The pentraxins 1975–2018: serendipity, diagnostics and drugs. *Front. Immunol.* **2018**, *9*, 2382.
- (70) Voß, S.; Klewer, L.; Wu, Y.-W. Chemically induced dimerization: reversible and spatiotemporal control of protein function in cells. *Curr. Opin. Chem. Biol.* **2015**, *28*, 194–201.
- (71) Sonpavde, G.; McMannis, J. D.; Bai, Y.; Seethammagari, M. R.; Bull, J. M. C.; Hawkins, V.; Dancsak, T. K.; Lapteva, N.; Levitt, J. M.; Moseley, A.; Spencer, D. M.; Slawin, K. M. Phase I trial of antigen-targeted autologous dendritic cell-based vaccine with in vivo activation of inducible CD40 for advanced prostate cancer. *Cancer Immunol. Immunother.* **2017**, *66*, 1345–1357.
- (72) Gerlach, L. O.; Jakobsen, J. S.; Jensen, K. P.; Rosenkilde, M. R.; Skerlj, R. T.; Ryde, U.; Bridger, G. J.; Schwartz, T. W. Metal ion enhanced binding of AMD3100 to Asp262 in the CXCR4 receptor. *Biochemistry* **2003**, *42*, 710–717.
- (73) Vo, N. H.; Xia, Z.; Hanks, J.; Yun, T.; Bloom, S.; Shen, J.; Koya, K.; Sun, L.; Chen, S. Synthesis, crystallographic characterization and electrochemical property of a copper(II) complex of the anticancer agent elesclomol. *J. Inorg. Biochem.* **2014**, *130*, 69–73.
- (74) Méndez-Lucio, O.; Medina-Franco, J. L. The many roles of molecular complexity in drug discovery. *Drug Discovery Today* **2017**, *22*, 120–126.
- (75) Böttcher, T. An additive definition of molecular complexity. *J. Chem. Inf. Model.* **2016**, *56*, 462–470.
- (76) Ivanenkov, Y. A.; Zagribelnyy, B. A.; Aladinskiy, V. A. Are we opening the door to a new era of medicinal chemistry or being collapsed to a chemical singularity? *J. Med. Chem.* **2019**, *62*, 10026–10043.
- (77) Hopkins, A. L.; Mason, J. S.; Overington, J. P. Can we rationally design promiscuous drugs? *Curr. Opin. Struct. Biol.* **2006**, *16*, 127–136.
- (78) Walters, W. P.; Green, J.; Weiss, J. R.; Murcko, M. A. What do medicinal chemists actually make? A 50-year retrospective. *J. Med. Chem.* **2011**, *54*, 6405–6416.
- (79) Raymer, B.; Bhattacharya, S. K. Lead-like drugs: a perspective. *J. Med. Chem.* **2018**, *61*, 10375–10384.
- (80) Lovering, F.; Bikker, J.; Humblet, C. Escape from flatland: increasing saturation as an approach to improving clinical success. *J. Med. Chem.* **2009**, *52*, 6752–6756.
- (81) Lovering, F. Escape from Flatland 2: complexity and promiscuity. *MedChemComm* **2013**, *4*, 515–519.
- (82) Kombo, D. C.; Tallapragada, K.; Jain, R.; Cheung, J.; Mazurov, A. A.; Speake, J. D.; Hauser, T. A.; Toler, S. 3D molecular descriptors important for clinical success. *J. Chem. Inf. Model.* **2013**, *53*, 327–342.
- (83) FDA's policy statement for the development of new stereoisomeric drugs. *Chirality* **1992**, *4*, 338–340.
- (84) U.S. Food & Drug Administration. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/development-new-stereoisomeric-drugs> (accessed 2021-01-25).
- (85) Murakami, H. From racemates to single enantiomers - Chiral synthetic drugs over the last 20 years. *Top. Curr. Chem.* **2006**, *269*, 273–299.
- (86) Talele, T. T. Opportunities for tapping into three-dimensional chemical space through a quaternary carbon. *J. Med. Chem.* **2020**, *63*, 13291–13315.
- (87) Hiesinger, K.; Dar'in, D.; Proschak, E.; Krasavin, M. Spirocyclic scaffolds in medicinal chemistry. *J. Med. Chem.* **2021**, *64*, 150–183.
- (88) Bauer, M. R.; Di Fruscia, P.; Lucas, S. C. C.; Michaelides, I. N.; Nelson, J. E.; Storer, R. I.; Whitehurst, B. C. Put a ring on it: application of small aliphatic rings in medicinal chemistry. *RSC Med. Chem.* **2021**, [Online early access]. DOI: 10.1039/D0MD00370K.
- (89) Brown, D. G.; Boström, J. Analysis of past and present synthetic methodologies on medicinal chemistry: where have all the new reactions gone? *J. Med. Chem.* **2016**, *59*, 4443–4458.
- (90) Leeson, P. D.; Springthorpe, B. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat. Rev. Drug Discovery* **2007**, *6*, 881–890.
- (91) Lajiness, M. S.; Vieth, M.; Erickson, J. Molecular properties that influence oral drug-like behavior. *Curr. Opin. Drug Discovery Devel.* **2004**, *7*, 470–477.