


EMCDDA framework and practical guidance for naming synthetic cannabinoids

Benedikt Pulver¹  | Svenja Fischmann¹ | Ana Gallegos² | Rachel Christie²

¹State Bureau of Criminal Investigation Schleswig-Holstein, Forensic Science Institute, Kiel, Germany

²European Monitoring Centre for Drugs and Drug Addiction, Lisbon, Portugal

Correspondence

Svenja Fischmann and Benedikt Pulver, State Bureau of Criminal Investigation Schleswig-Holstein, Forensic Science Institute, Kiel 24116, Germany.

Email: svenja.dr.fischmann@polizei.landsh.de and benedikt.pulver@projekt-adebar.eu

Funding information

Internal Security Fund of the European Union, Grant/Award Numbers: IZ25-5793-2016-27, IZ25-5793-2019-33

Abstract

Synthetic cannabinoids (SCs), often sold as “legal” replacements for cannabis, are the largest group of new psychoactive substances monitored by the European Monitoring Centre for Drugs and Drug Addiction (EMCDDA). Currently, close to 240 structurally heterogeneous SCs are monitored through the European Union (EU) Early Warning System, and attributing consistent, informative, and user-friendly names to SCs has been a challenge in the past. Over time, several naming conventions have been employed with the aim of making SCs more easily recognizable by non-chemists, including regulators. To achieve this, the names assigned need to contain detailed information on the structural features present in the substance.

This work provides a theoretical framework and a practical hands-on guideline for consistent naming of SCs, which is easy to understand and can be applied by the forensic community, researchers, clinical practitioners, and policy-makers. The proposed framework builds on the established letter code system for molecular building blocks (core, linker, linked group, and tail) implemented by the EMCDDA in 2013 and has been expanded to incorporate additional structural features through substitution. The scope of the issue of attributing semi-systematic code names is illustrated, and earlier approaches used for naming SCs are discussed. The concepts and rules of the EMCDDA framework are described through a flowchart that provides a basis for naming new SCs, a graphical overview of the chemical diversity of SCs, and a detailed list of the SCs identified in the EU by the Early Warning System of the EMCDDA for reference.

KEYWORDS

early warning system, harmonization, new psychoactive substances, semi-systematic naming, synthetic cannabinoids

1 | INTRODUCTION

1.1 | Evolution in the naming of synthetic cannabinoids

Because the first synthetic cannabinoid (SC) JWH-018 was detected in herbal material in 2008,^{1,2} the group of SCs has been growing in number and structural diversity. Many SCs that emerged before 2019 had been previously documented in the scientific and patent

literature. Various classes of substances were identified to bind to the same receptor like tetrahydrocannabinol (THC) and thus investigated for their cannabimimetic effects to be used as medicines.^{3,4} In some instances, the development of new molecules with cannabimimetic properties was pursued using combinatorial chemistry, with chemical libraries prepared through systematic and repetitive covalent bonding of building blocks containing many structurally similar compounds. The compounds were screened against biological targets to assess their properties at the point of interest, generating chemical libraries

of potential SCs to be manufactured and sold on the drug market.⁵ Research into different substance classes by individual researchers like John W. Huffman and Alexandros Makriyannis, as well as the laboratories at the Hebrew University in Jerusalem and the Pfizer company, led to the first generation of SCs, which were notified to the Early Warning System (EWS) on new psychoactive substances of the European Union (EU) from 2008 to 2014. A combination of a prefix and a serial number, for example, JWH-018, has been used in the patent and scientific literature to codify newly synthesized compounds within chemical libraries and for ease of reference. The prefixes JWH and AM are the initials of the names of researchers John W. Huffman and Alexandros Makriyannis; and the letter codes HU, CP, and WIN stand for Hebrew University, Carl Pfizer, and Sterling-Winthrop, respectively.

Besides the codes used in patents and scientific literature, names initially assigned to an SC also depended on the marketed name of the product. Clandestine laboratories and internet vendors used flashy, intriguing, and recognizable names for the products to presumably improve sales and create an emotional connection for the users. Examples include AKB48 and its indole analog 2NE1, named after popular girl bands in Japan and South Korea, respectively.⁶ XLR-11 refers to liquid rocket fuel used in aircraft, and the name of the herbal material in which the SC AB-001 was identified, “atomic bomb,” was used as a street name to refer to this SC later, potentially pointing toward the potency of said substance.^{7–9}

The combination of an appended string and a prefix to highlight the halogenation, for example, was used for the indazole analog of 5F-PB-22. The name “5F-PB-22 indazole analog”¹⁰ highlights how complex the naming of SCs can become if the original name of the most structurally similar SC is continuously modified (see Figure 1). New SCs with different and varied structural moieties appeared on the market, and, as a result, inconsistencies can be observed in the implementation of the naming where the presence of new building blocks is added to the name of the original SC in various ways (prefix, suffix, with and without hyphens or parentheses).¹¹ The continuous modification of an alphanumeric name has its limits and ultimately leads to ambiguous names that contradict the idea of deducing information on the structure from the assigned name.

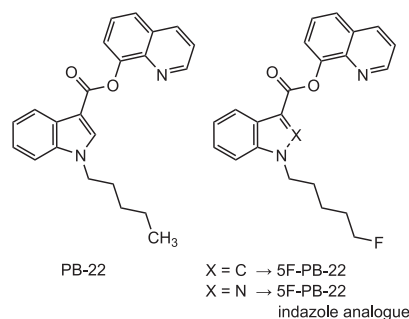


FIGURE 1 Naming of three SCs as examples of former naming approaches

1.2 | Naming SCs detected in Europe

Forensic practitioners and researchers value nomenclatures that allow for the deduction of chemical features and clear differentiation of structurally similar SCs for the most unambiguous communication while avoiding long, nonintuitive, and not user-friendly unambiguous International Union of Pure and Applied Chemistry (IUPAC) names or other descriptors (such as InChIkey). Often, when an SC containing a new moiety emerged in the market for which information was lacking in the literature, deducing a common name would involve considering various sources of input (the name given by the producer, name reported by the laboratory, etc.). Using the names provided by the drug market or by the patent literature does not provide an insight of structural features. Nonetheless, and in particular when these names gain traction, they need to be recorded for completeness and reference.

Semi-systematic naming conventions based on the abbreviation of structural features were developed separately by the EMCDDA¹² and Cayman Chemical^{13,14} close in timeframe to each other. The top-level overview of the EMCDDA naming approach to SCs was made available in 2013 on the EMCDDA public website and considered four fundamental building blocks (core, linker, linked group, and tail). The groundwork was laid by Blaazer et al. in 2011 and Uchiyama et al. in 2012 and 2013.^{6,15,16} Blaazer et al. described the different elements of the SC JWH-018 with side chain, heterocyclic nucleus, linker, and lipophilic substituent, which correspond to tail, core, linker, and linked group, respectively. The elements were designated as sites of modification in a structure–activity relationship study, which is similar to how the chemical libraries are generated in drug design using combinatorial chemistry. This could be considered the first time a stratification of an SC into four elements was mentioned in the scientific literature. Uchiyama et al. described the identification of four SCs, namely, APICA, APINACA, AB-PINACA, and AB-FUBINACA in 2012 and 2013, which was the first time in which the IUPAC name was used to derive a shorter name, which contains information on the SCs' structure. Notably, the name was derived from the complete name without mentioning building blocks as a way to deduce the name. The naming approach made available by the EMCDDA in 2013 can be considered as the culmination of both ideas: the stratification of SCs into four building blocks and the abbreviation of the systematic name through a selection of letters.

The naming approach suggested by Cayman Chemical differed in considering only three elements: a core (equivalent to core + linker in the EMCDDA model), a head (equal to the linked group), and a tail.¹⁷ This model goes as far back as at least 2015 and was revised in February 2022.¹⁴ Implementation of these naming conventions provided the first practical guidelines available allowing for a more consistent naming of SCs. Notably, only the guideline of the EMCDDA is currently accessible. Because the process of deducing a letter code is not straightforward and different approaches to arranging these three or four elements were taken by different naming conventions, the derived names were not consistent across all SCs. Furthermore, the

lack of documentation publicly available for this naming approach may have also led to some inconsistencies.

In this work, structural features present in all historical SCs monitored by the EMCDDA are assigned letter codes derived from the chemical name consistent with the developed EMCDDA framework. The rules used for the deduction of the letter codes are illustrated throughout using several examples. Inconsistencies among the common names used up until now are discussed. The developed framework aims to use the fewest letter codes necessary to name SCs while simultaneously providing names for SCs, which are as rich in structural information as possible, and builds on the previous model developed by the EMCDDA. Thus, 49% of the new harmonized names deviate only slightly from the naming approach formerly employed or are the same.

2 | METHODS

Source data on past SCs were derived from the European Database on New Drugs (EDND) (date of accession: 07.07.2022).

The cannabinoids included in this study are all categorized by the EMCDDA classification system and pharmacological classification system as SCs and are actively monitored by the EMCDDA. A detailed list of the SCs covered by the EMCDDA framework can be found in Table S2 of the Supporting Information. Additional SCs not included in this work are outlined in Table S3.

2.1 | Naming syntax and letter codes

The overarching purpose of the EMCDDA framework is to provide a concise, systematic name reflecting the structural composition/makeup of the SC. The gold standard to achieve this is the IUPAC name in which names and positional descriptors to every element of a molecule are concatenated using a set of clear rules. Hence, a precise but complex name is generated, from which structural information can be derived. The IUPAC name uses the English language, and thus, the codified information, although sometimes complex, can be deduced by researchers without the help of a computer. The IUPAC consortium has produced an additional identifier in collaboration with the National Institute of Standards and Technology (NIST) to communicate molecular structures using computer-interpretable strings, namely, the InChI string and InChI key.^{18,19} An additional and commonly used identifier is the simplified molecular input line entry system (SMILES),²⁰ which allows the specification of structures using small and natural grammar.

2.2 | The basic syntax of the four building blocks

To strike a balance between conveyed information and utility, letter codes are derived for four reoccurring building blocks of SCs. The SCs can be divided into four building blocks: core, linker, linked group, and tail. The generic composition of SCs using these building blocks is displayed in Figure 2. Letter codes have a similar function to acronyms.

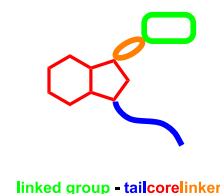


FIGURE 2 Basic syntax stratifying SCs into four building blocks [Colour figure can be viewed at wileyonlinelibrary.com]

They are derived from the chemical name they abbreviate and are a simple representation of complex structural features. The letter code of each building block is joined via the building block syntax “**Linked group-TailCoreLinker**,” representing this generic makeup. The basic syntax uses hyphens to separate the linked group from the three other building blocks within the letter code. Joining the letter codes of the four building blocks under the EMCDDA framework leads to consistent semi-systematic names.

The modular approach includes the majority of SCs listed in the EDND (91%) and represents the dynamic of SC development, which appears to have followed a combinatorial approach, by combining the different moieties. Developments that add entirely new moieties to the list of building blocks, for example, the tosyl side chain, the oxindole core, or the hydrazide linker, can be stratified into this pattern. Using the four building blocks as a basis for syntax and letter codes, the EMCDDA framework detailed in this publication only applies to SCs adhering to this general makeup. Hence, for example, CP-, WIN-, and HU-termed SCs are not classified within this approach (see Supporting Information for a complete list of EDND listed SCs not applicable).

The most important principle within the EMCDDA framework is the unambiguous description of molecular structures. Hence, each letter code must be unique and only abbreviated for one building block. However, the individual letters within a letter code may abbreviate for varying strings. For example, the letter code CA abbreviates the carboxamide linker commonly found in SCs, indicating the carbonyl group and amide moiety. In the letter code CBM (cyclobutyl methyl), on the other hand, the letter C abbreviates for “cyclo.” The use of single letters does not have to be unambiguous.

Basic syntax

- Defines four building blocks: core, linker, linked, and tail group
- Letter codes are unique and abbreviate structural features.
- Letter codes are joined via the building block syntax: Linked group-TailCoreLinker.
- Letters themselves do not have to be unique.

2.3 | Expanded syntax through substitution of building blocks

Substitution of a motif of interest using many different substituents is part of the combinatorial approach to screen drug candidates in

clinical and clandestine/illicit drug development. The result is a plethora of SCs that cannot be univocally described using the basic syntax. Common substitution groups were used to modify the four building blocks further. Within the AM and JWH series, the linked group has been extensively altered (e.g., halogenation and alkylation) to include various functional groups. In Figure 3, the naphthyl scaffold (letter code: NA) and locants assigned to each carbon atom are shown on the left. By substituting a common element, namely, halogenation and alkylation, a halogen atom and methyl (M) group are added, respectively.

The locants now describe the position of the introduced substituent. In theory, substitution can occur at multiple positions throughout the common element creating structural isomers closely resembling each other. The only differentiation is possible via the locant indicating which structural isomer is present. Hence, there exists the need for a modification of the basic syntax to enable the inclusion of the type and position of substitution of a common element in an SC.

Substitution of a building block that promises desirable properties to explore further potential is common in medicinal chemistry and drug design. Hence, the terminology described in this work includes substitution as part of the expanded syntax. Although entire groups of structurally similar SCs altered through substitution have not been encountered recently, the emergence of core-substituted SCs underpins the continued relevance. The previously introduced basic syntax is expanded (see Figure 4). The substitution and their representation in the derived semi-systematic name are clarified in Figure 4 via two examples.

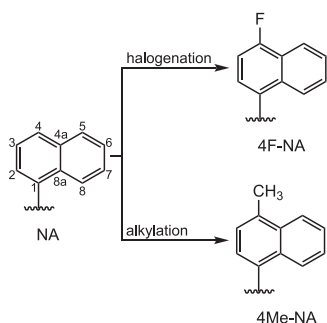
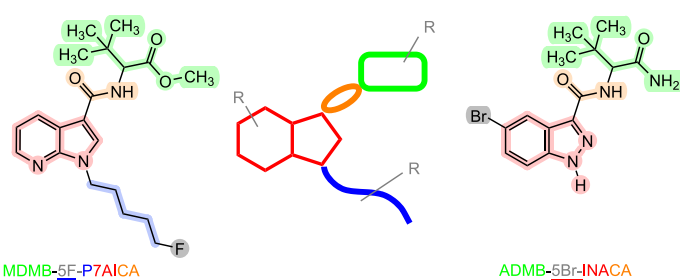


FIGURE 3 Substitution of a naphthyl scaffold



linked group substitution - linked group - tail substitution - tail - core substitution - corelinker

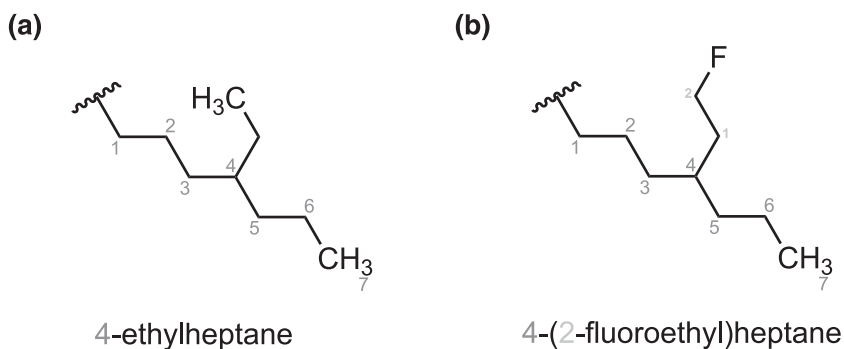
On the left, MDMB-5F-P7AICA (also known as 5F-MDMB-P7AICA) illustrates the substitution of the tail group via a fluorine atom. The location, as well as the type of substitution, is introduced. A hyphen is added to separate the substituent from the building block. Notably, the locants assigned to the carbon atoms of the tail group start at the junction of the tail group and core. Thus, the deduced locant for the fluorine atom is five. On the right, an SC is shown lacking a tail group element, thus only composed of a core, linker, and linked group. The bromination is indicated through the locant and abbreviation “Br” in front of the letter code of the substituted building block, here the core. “Br” is used as the bromination that introduces the element bromine, which has the element symbol “Br” in the periodic table and is commonly used for this substituent. A hyphen is added with the substitution denotation (gray). The hyphen preceding the substitution denotation originates from the basic syntax. Notably, if the core-substituted SC features a tail group, a second hyphen is added to the semi-systematic name due to the substitution. The syntax of the complete structure indicates this, with the gray hyphen preceding the “core substitution.”

The substitution of a building block does not lead to a new letter code but rather an amendment via a prefix, including the substitution type and location. A substituent amended to a building block must not be substituted further to retain the recognizable syntax. A nested substitution is used in a substitute chemical name to clearly describe the position of all the main characteristic groups within a molecule relative to each other. The nesting order highlights which part of the molecule belongs together and is reflected by brackets.²¹ Two theoretical examples for tail groups are shown in Figure 5. The systematic name for the building block is written underneath the molecular structure. The brackets in Figure 5b indicate a nested substitution. An accurate representation of this substitution pattern would require brackets severely compromising the concise nature of the semi-systematic name. A new letter code has to be assigned whenever a new structural feature emerges that requires nested substitution to be described accurately. Still, the focus should always be on describing new structural features with the existing vocabulary of letter codes.

After identifying the most extensive structural feature within a building block and assigning a name from either a trivial name or an IUPAC denotation, the substitution pattern is added as a prefix with the locant and the type of substitution. For a particular structural element to be considered substituted, it has to share a common entity

FIGURE 4 Expanded syntax for deducing the semi-systematic name for a given SC, including the possibility of substitution at the core, tail, or linked group [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE 5 Theoretical examples of building blocks highlighting nested substitution in the systematic name



that has been identified in another SC without any modification. So far, the substitution of building blocks has occurred in the tail group, core, and linked group. The substitution has thus mainly occurred in the periphery of the SCs, where a small change of the molecule can be achieved without impacting the molecule as a whole.

Expanded syntax

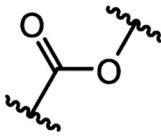
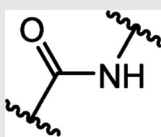
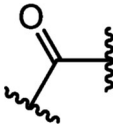
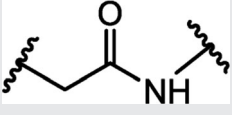
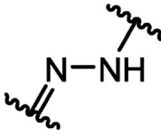
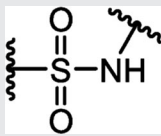
- Substitution leads to minor changes to previous building blocks.
- Prefixes are used to modify the letter codes of building blocks.
- Prefixes amend position and type of the substitution with a hyphen.
- Substituents are represented using commonly used codes, for example, Br for bromine.
- Complex nested substitution is avoided by deriving new building blocks.
- Hyphens are added in front of the prefix when joining the modified letter codes.
- The expanded syntax covers a greater chemical diversity of SCs.

3 | LETTER CODES FOR BUILDING BLOCKS IDENTIFIED SO FAR

All SCs notified to the EMCDDA were analyzed, and the four building blocks in each of the SC were identified. In Sections 3.1–3.4, unique structural elements identified as linker, core, tail, and linked group are displayed together with their molecular structure, systematic name, and derived letter code under the EMCDDA framework. The repository of individual building blocks allows for an overview of the chemical diversity of SCs (see Table 1 to Table 11). In addition, detailed explanations on how the letter codes are derived and why certain aspects of the expanded syntax are necessary are given.

The expanded syntax covers 91% of structures classified as cannabinoids by the EMCDDA in the EDND. The increased coverage, consistency, and uniqueness of letter codes in the EMCDDA framework allow for the visualization and organization of SCs according to the four building blocks, as well as the type and position of substitution. Table S2 can be implemented into databases to make the individual building blocks searchable and allow for a more accessible overview of the chemical diversity of SCs that have emerged on the EU drug market so far. All building blocks described in the tables of

TABLE 1 Structural moieties, IUPAC names, and assigned letter codes for linker elements identified within SCs

Structure element	Letter code	Systematic name
	C	c arboxylate
	CA	c arboxamide
	MO	m ethanone
	ATA*	a cetamide
	ZID	hydrazide
	SA	s ulfonamide

*formerly abbreviated with ACA.

this work have been displayed in Figure 6 to convey the chemical diversity in SCs further.

3.1 | Linker building blocks

In general, SCs are composed of linker, core, tail, and linked group building blocks. The linker is always connected to the linked group and the core, which itself connects to the tail group.

The group of linkers has historically been the most homogenous of all the building block groups because of the prevalence of the CA linker in so many SCs. Compared to the other group of building blocks, it contains relatively few different elements. Up until the introduction of the hydrazide linker, all linkers feature the common carbonyl group. The methanone (MO) linker is the simplest, and changes were made in the α -position to the carbonyl group forming an amide, ester, and acetamide (ATA). The letter codes used by the new EMCDDA framework are shown in Table 1. The letter codes for CA, ATA, hydrazide, and sulfonamide are already utilized to describe the SCs in the EDND. However, the letter codes for MO and carboxylate were newly assigned to generate semi-systematic names for the JWH and AM series.

ADB-FUBIATA, a new SC, featuring a novel linker element, emerged on the European market around mid-2021. It was first identified in Europe by the Bulgarian Customs in August 2021.²² A sample received on the 29th of September 2021, by the EU-project ADEBAR plus, also allowed the detection of ADB-FUBIATA on paper from a seizure of July 2021 by the state police of Schleswig-Holstein. The

chemical identity of ADB-FUBIATA in this seizure was confirmed by NMR analysis in mid-October 2021.^{23,24}

ADB-FUBIATA was also detected in powder and e-liquid samples, seized by police in China, following a class-wide ban on SCs in July 2021. Compared to the previously ubiquitously used CA linker, a methylene group was added to yield the ATA moiety. The name of ATA is deduced from the linker $\text{CH}_2\text{-CONH}_2$ molecule considered by itself. The letter code ATA is used by scientists at Cayman Chemical and NPS Discovery by the Center for Forensic Science Research and Education (CFSRE).²⁵ The letter code ACA has been used in the formal notifications issued by the EMCDDA to refer to the acetamide linker so far. As a result of the present work, the letter code ATA is adopted and used to refer to the ATA linker going forward. The hydrazide and the acetamide moiety have been the latest additions and have been brought onto the market probably in response to a class-wide ban on SCs in China.^{26,27}

3.2 | Core building blocks

The core building block is always connected to the tail and the linker, which in turn connects to the linked group. Most molecular structures found in the group of the core building blocks are bi- and tricyclic aromatic systems. Although the composition of these aromatic systems varies, the connection to the other building blocks has remained simple. Bicyclic core elements are always 1,3-disubstituted with the tail and linker group except for the recently emerged NA structure. The carbazole and γ -carbolinone have also been substituted coherently.

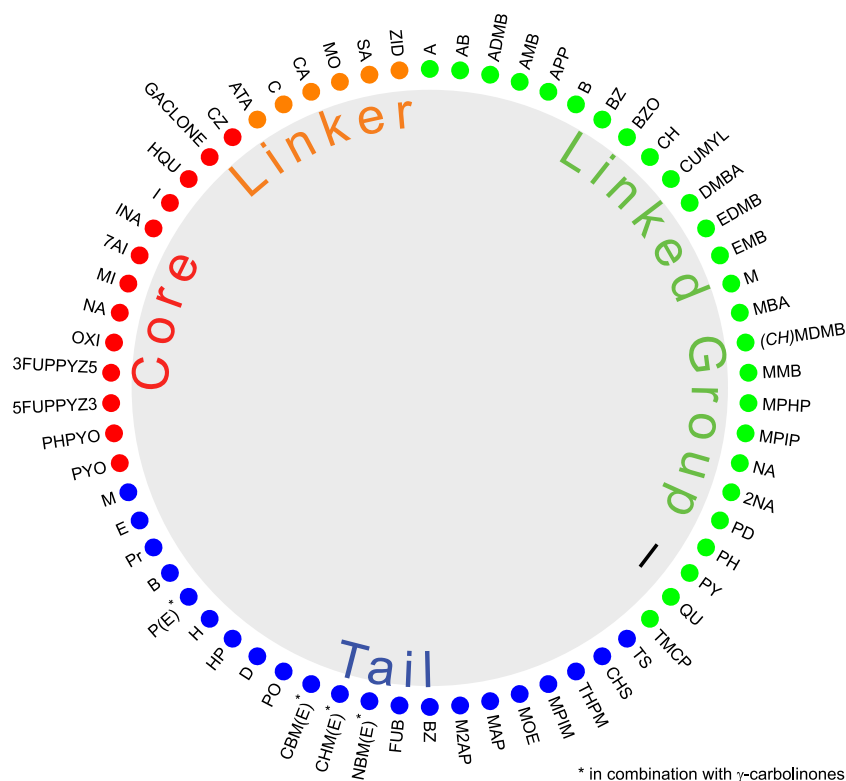


FIGURE 6 Overview of all letter codes used in the EMCDDA framework as of July 7, 2022 [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/terms-and-conditions)]

The locants at which the carbazole and γ -carbolinone are substituted are 3,9- and 2,5-, respectively, leading to similarly oriented SCs (see Section 3.2.1). The last group of structural features found as core elements contains pyrazole and pyrrole (PYO) moieties and introduced variation in the substitution pattern of the core for the first time (see Section 3.2.2). The necessary adjustments to the syntax are derived from Figure 8.

The core as a building block of SCs has been left unsubstituted in SCs notified to the EMCDDA for the most part. Substitution of a core structure has not occurred before the identification of ADB-5Br-INACA in 2021. So far, the brominated indazole core structure has been used in two SCs without a tail group. Before the brominated core, developments of new core structures, which appeared on the European drug market, have consisted of the exchange of atoms in the molecular frame and the introduction of new core elements altogether.

3.2.1 | Bi- and tricyclic core building blocks

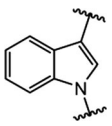
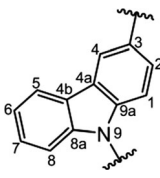
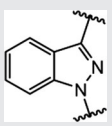
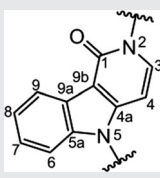
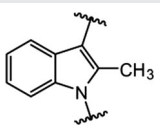
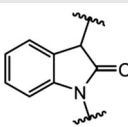
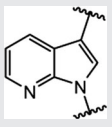
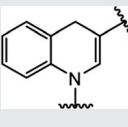
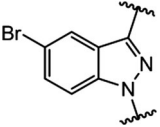
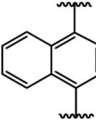
The overall majority of SCs monitored by the EMCDDA features the bicyclic indole and indazole core elements. The most prominent

tricyclic core element has been the γ -carbolinone, and new core structures, namely, oxindole and hydroquinoline, have emerged recently. In Table 2, the most prevalent core structures that have emerged are assigned letter codes under the developed framework. The oxindole core structure has emerged together with the hydrazide linker, probably derived from the work of Diaz et al., which investigated isatin hydrazide-type structures for the treatment of neuropathic pain.²⁸ The connection between the individual building blocks is a single bond apart from the OXIZID type SCs. Here, the joining bond is a double bond as implied by the linker.

3.2.2 | PYO- and pyrazole-type core building blocks

The PYO- and pyrazole-type core elements are unique as they do not easily fit into the four-tiered building block system previously described by the EMCDDA. However, the three building blocks, apart from the core, are easily identified in the PYO- and pyrazole-type SCs. Thus, in some cases, the core building block incorporates not only the PYO and pyrazole moiety but also the phenyl and fluorophenyl element. The core elements of this type that have emerged on the drug market are detailed in Table 3.

TABLE 2 Structural moieties, IUPAC names, and assigned letter codes for core elements identified within SCs

Structure element	Letter code	Systematic name	Structure element	Letter code	Systematic name
	I	1H-indole		CZ	9H-carbazole
	INA	1H-indazole		GACLONE	gamma-mma-carbolinone-1-one
	MI	2-methylindole		OXI	2-oxindole
	7AI	1H-7-azaindole		HQU	1,4-dihydro-quinoline
	5Br-INA	5-bromo-1H-indazole		NA	naphthalene

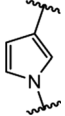
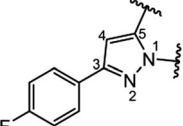
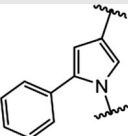
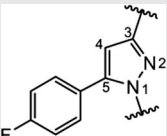
The core elements shown in Table 3 have been derived from the structure of the SCs by following the steps outlined in the flowchart for the naming of SCs (see Figure 16). The elements within the molecular structure of the SC shown in Figure 7a have been sectioned off and assigned building blocks where possible. First, the linker and linked group moiety were identified by comparing the top part of the molecule to all of the linker and linked groups shown in Table 1 and Table 7. The linker building block is a methanone represented by the letter code MO and illustrated in orange. The linked group building block is a naphthyl represented by the letter code NA and illustrated in green. Lastly, a pentyl moiety is connected to the PYO feature and illustrated in blue. Compared with the structural moieties described here, this part of the molecule is identified as a tail group represented by the letter code P.

One of the key challenges facing naming systems for NPS is the dynamic pace at which these substances appear and their chemical diversity. This also makes it difficult to predict which building blocks may occur in the future. The building block that has not been identified yet is the core illustrated in red. There are two ways to proceed in this case: (1) Define the entire unassigned part of the molecule as a novel core building block or (2) only a part is considered the core element. In this case, the residual element (a phenyl) becomes a substitution of the core. There is one SC monitored by the EMCDDA in which a PYO makes up the core building block, JWH-145, and thus it might

sound reasonable to proceed via substitution of the PYO leading to 5PH-PYO. The SC, described as JWH-370 in the primary literature, shown in Figure 7b, provides the basis for why the phenyl-PYO (PHPYO) building block has to be defined rather than using the substitution syntax. If the substitution of a PYO via a phenyl ring was chosen, the SC on the right would then require a nested substitution, which is omitted from the EMCDDA framework to retain a recognizable syntax (see Section 2.1). The core element shown in Figure 7b is one of the four substituted PHPYO core elements that have emerged so far. Table 4 presents an overview of the four derivatives, including their assigned letter code according to the substitution syntax and their systematic name.

Contrary to the other core elements, the pyrazole-type core element is the first example of an otherwise consistent connection to the adjacent building blocks. In its simplest form, the 1*H*-pyrazole is assigned locants starting at the indicated hydrogen in the direction of the second heteroatom (Figure 8). Thus, the locants of the pyrazole, which indicate the connection to the other structural features in the two examples shown in Figure 8, are assigned accordingly. The changes in the arrangement of the pyrazole ring compared to the other parts of the SC require the inclusion of locants into the syntax. The locant of the linker changes from 3 to 5, and the locant of the residual part of the core element, the fluorophenyl, changes from 5 to 3. The locant is joined directly in front of the letter code, representing

TABLE 3 Structural moieties, IUPAC names, and assigned letter codes for pyrrole and pyrazole core elements identified in SCs

Structure element	Letter code	Systematic name	Structure element	Letter code	Systematic name
	PYO	1 <i>H</i> -pyrrole		3FUPPYZ5	3-(4-fluorophenyl)-1 <i>H</i> -pyrazol-5-yl
	PHPYO	5-phenyl-1 <i>H</i> -pyrrole		5FUPPYZ3	5-(4-fluorophenyl)-1 <i>H</i> -pyrazol-3-yl

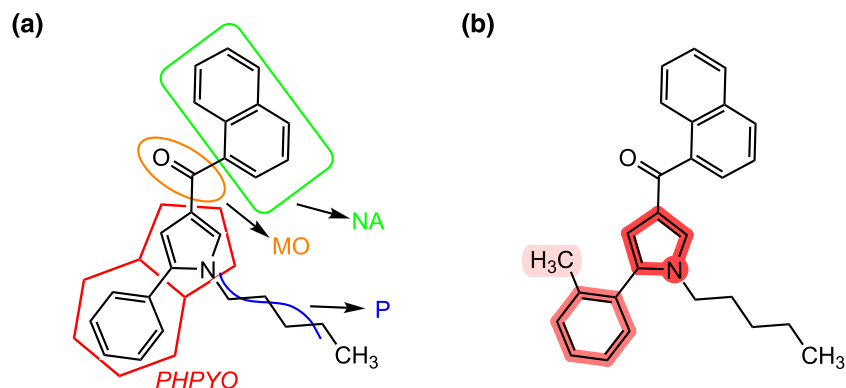


FIGURE 7 Example of the process of deriving a semi-systematic name for a “new” SC (a) and illustration of nested substitution of the phenyl substituent to the pyrrole ring via a methyl group (b) [Colour figure can be viewed at wileyonlinelibrary.com]

the structural feature which position it clarifies. The locant also amends the letter code of the linker. The tail group remains connected to Position 1 of the pyrazole, the position of the indicated hydrogen.

To clarify how to add locants, two additional examples are shown in Figure 9. The linked group present in both examples is an aminodi-methylbutanoate (ADMB), previously referred to as the code ADB. The tail group element of an alkenyl chain is also present in both examples and connected to the core building block at the pyrazole

ring in Position 1, which is the only position a tail group has been joined to a pyrazole-type SC so far. Thus, no further clarification through a locant is required here. In the example on the left, the linker, a CA element, is connected to the pyrazole ring at Position 3. The position of the fluorophenyl feature (FUP) connected to the pyrazole ring is indicated by Locant 5. In contrast, the linker and FUP in the example on the right are connected to the pyrazole ring at Positions 5 and 3, respectively.

TABLE 4 Structural moieties, IUPAC names, and assigned letter codes for substituted pyrrole core elements identified within SCs

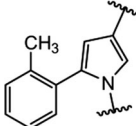
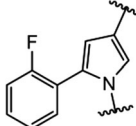
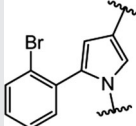
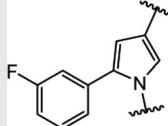
Structure element	Letter code	Systematic name	Structure element	Letter code	Systematic name
	2Me-PHPYO	2 -methyl-(5-phenyl-1H-pyrrole)		2F-PHPYO	2 -fluoro-(5-phenyl-1H-pyrrole)
	2Br-PHPYO	2 -bromo-(5-phenyl-1H-pyrrole)		3F-PHPYO	3 -fluoro-(5-phenyl-1H-pyrrole)

FIGURE 8 Locants assigned to the pyrazole ring within the core building block illustrate the variation in the connection of the adjacent building blocks. [Colour figure can be viewed at wileyonlinelibrary.com]

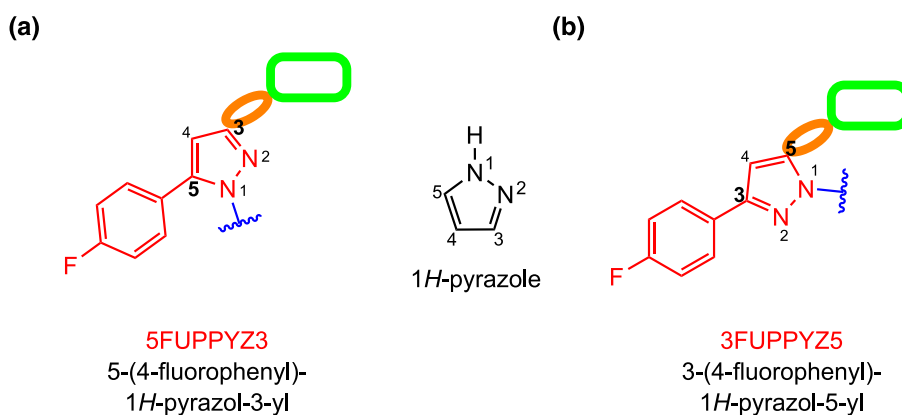


FIGURE 9 Example of the variation in the position in which the building blocks are connected to the pyrazole ring. The SCs shown are also known as 5,3-ADB-4en-PFUPPYCA (a) and 3,5-ADB-4en-PFUPPYCA (b). [Colour figure can be viewed at wileyonlinelibrary.com]

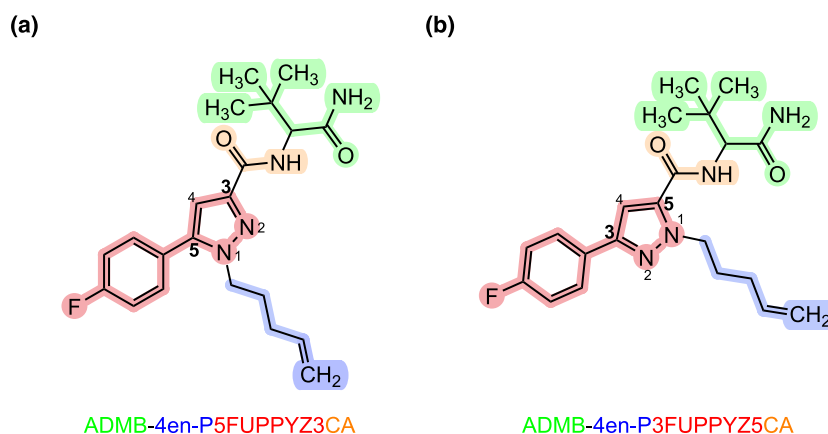

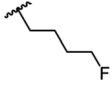

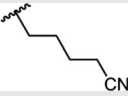

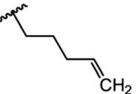

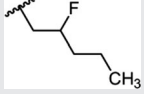
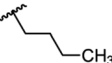


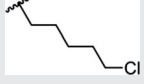
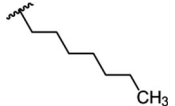

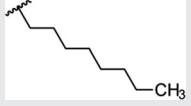
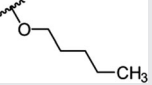
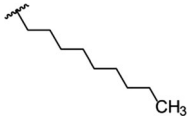
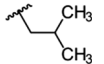
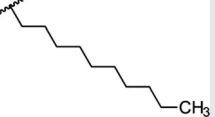
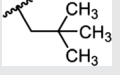
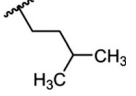


TABLE 5 Structural moieties, IUPAC names, and assigned letter codes for alkyl tails identified within SCs

Structure element	Letter code	Systematic name	Structure element	Letter code	Systematic name
	M	methyl		4F-B	4 -fluoro-butyl
	E	ethyl		4CN-B	4 -cyano-butyl
	PR	propyl		4en-P	pent-4-en
	B	butyl		2F-P	2 -fluoro-pentyl
	P / PE*	pentyl		5F-P / 5F-PE*	5 -fluoro-pentyl
	H	hexyl		5Cl-P	5 -chloro-pentyl
	HP	heptyl		5Br-P	5 -bromo-pentyl
	O [‡]	octyl		PO	pentyloxy
	N [‡]	nonyl		2Me-PR [‡]	2 -methyl-propyl
	D	decyl		2,2Me-PR [‡]	2,2-dimethyl-propyl
				3Me-B [‡]	3 -methyl-butyl

*in combination with carbolinones.

[‡]not detected on the drug market so far.

Pyrrole and pyrazole core elements—A special case

- Pyrrole and pyrazole-type core elements do not easily fit into the four-tiered building block system.
- Pyrazole-type core element broke with a consistent connection to the adjacent building blocks
- For 1H-pyrazole, assigned locants start at the indicated hydrogen in the direction of the second heteroatom
- Locants are necessary to indicate connections of the pyrazoles to other structural features (positions 3 and 5)

3.3 | Tail building blocks

The tail group is always connected to the core element in an SC. The core connects to the linker, which connects to the linked group. The tail group is the building block of SCs, which has seen the most significant evolution in recent years. The letter codes for *N*-alkyl and cyclic tail groups and the letter codes for their substituted analogs are discussed in the following sections. Furthermore, inconsistencies in the letter codes assigned in the past are outlined to highlight the challenges associated with the previously utilized names.

3.3.1 | Alkyl tail building blocks

Alkyl side chains can be found in many SCs. The assigned letter for an unbranched alkyl side chain is derived from its chemical name. Where necessary, in the scope of alkyl side chains, the single letter is extended by one additional letter, where necessary, to enable unambiguous identification. Thus, the hexyl side chain moiety is assigned H as the letter code which has been used in the naming of *N*-(1-Amino-3,3-dimethyl-1-oxobutan-2-yl) 1-hexyl-1*H*-indazole-3-carboxamide (also known as ADB-HEXINACA) already.²⁹ Concomitantly, the heptyl moiety will be assigned the letter code HP. Because the pentyl moiety was first used in the common name of APINACA, *N*-(1-adamantyl)-1-pentyl-1*H*-indazole-3-carboxamide; in 2013, a propyl moiety is abbreviated using PR (Table 5). Because alkyl-type elements can be found in many structural features and are commonly used in trivial and IUPAC naming, the letter codes shown in Table 5 should be used in those instances as well (i.e., a cyclohexyl moiety is termed CH with H abbreviating for the hexyl element similar to the linear chain). To illustrate, using another example, a cyclopropyl moiety will be abbreviated CPR to retain the abbreviation of PR for the propyl element. The alkyl tail elements shown in Table 5 represent the substituted building blocks. The common elements are saturated alkyl chains, which are substituted in different ways. The substitution is added to the letter code of the common element as a prefix together with a hyphen. The introduced atom or functional group is represented according to chemical abbreviations used in organic chemistry. Locants indicating the substitution position are assigned starting at the connection with the core building block outward.

The tail groups (5*F*-)pentyl highlight an exception to the rule, where normally a pentyl side chain is abbreviated with the letter P only. When found in an SC with a γ -carbolinone, the pentyl tail group (substituted or not) is abbreviated using Pe. The exception is made to retain the highly recognizable term “PEGACLONE.” The same concept is applied to the cyclic tail groups in combination with a γ -carbolinone core (see Table 6). Other possible alkyl tail groups in which the first letter of the systematic name is preceded by a consonant, for example, butyl, hexyl, and heptyl, are treated similarly. However, it has to be noted that single letters are used for all other tail groups with alkyl elements.

Small letter denotation within the letter code can be used whenever a single functional group is described by more than one letter (e.g., Tosyl→Ts). Small letters within the name of an SC will help differentiate individual building blocks within the expanded syntax more easily. Again, naming conventions agreed upon in the scientific community are the blueprint to draw from. Ultimately, this should not be treated as a strict rule but rather to increase the readability and interpretability of the semi-systematic names.

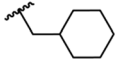
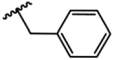
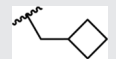
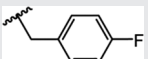
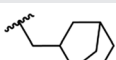
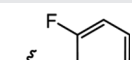
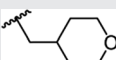
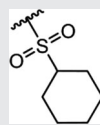
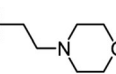
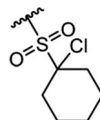
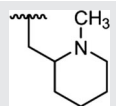
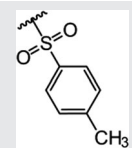
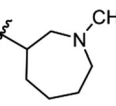
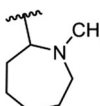
Thus far, only *N*-alkyl side chains have been discussed. However, branched alkyl chains have been observed for many linked groups described in Section 3.4. Three linked groups stand out, requiring guidelines for naming highly branched structural features (see Figure 15). Here, we describe three examples of SCs with branched aliphatic tail groups and how to derive the correct letter code (Figure 10). The naming of branched aliphatic tail groups follows the substitution syntax in terms of abbreviating the additional elements and joining the locants and letter codes together. The locants are assigned to the longest, unbranched aliphatic chain, starting from the atom covalently bonded to the core. The “substituted” M groups are listed together with the individual locant(s) in front of the letter code of the unbranched aliphatic chain. The multiplicative prefix is omitted from the letter code. The locants indicate the number of substitutions of one particular kind.

3.3.2 | Cyclic tail building blocks

The following table (Table 6) highlights tail building blocks in which cyclic elements can be observed—beginning with the core structure, most tail moieties of the past exhibit a M moiety that can be seen as a spacer group between the core structure and the terminal functional group. The first example shown in Table 6 comprises a M spacer and a CH group. The letter code of the spacer is concatenated to the terminal functional group and subsequently joins the letter code of the core structure as per expanded syntax. An ethyl spacer (E) has been observed in the tail group moiety and was assigned the letter code MOE previously. Here a morpholine (MO) group is connected to the core using an E spacer. Consequently, one of the more recently monitored SCs, Cumyl-TsINACA, does not exhibit a spacer-type connector within the side chain.

The 4-fluorobenzyl (FUB) tail group has been found in several SCs notified to the EMCDDA in the past and gained widespread

TABLE 6 Structural moieties, IUPAC names, and assigned letter codes for cyclic tail elements identified within SCs

Structure element	Letter code	Systematic name	Structure element	Letter code	Systematic name
	CHM / CHME*	c yclohexylmethyl		BZ	b enzyI
	CBM / CBME*	c yclobutylmethyl		FUB	4-fluorobenzyl
	NBM / NBME*	n orbonylmethyl		2F-BZ	2 -fluorobenzyl
	THPM	t etrahydropyran-4-yl-methyl		CHS	c yclohexylsulfonyl
	MOE	(4-morpholinyl)-ethyl		1Cl-CHS	1 -c hloro-cyclohexylsulfonyl
	MPIM	(1-methyl-piperid-2-yl)-methyl		TS	t osyl
	MAP	1-methyl-azepan-3-yl		M2AP	1-methyl-azepan-2-yl

*in combination with γ -carbolinones.

recognition, similar to the creation of the “MEGACLONE” abbreviation for the γ -carbolinone. Although the letter code FUB is not consistent with the EMCDDA framework, it is included as an exception because of its established use. Any future modification of a benzyl group via fluorination in the *ortho* or *meta* position will follow the substitution syntax described here and yield 2F-BZ and 3F-BZ, respectively. The SC *N*-(1-carbamoyl-2-methyl-propyl)-1-[(2-fluorophenyl)methyl]indazole-3-carboxamide (AB-FUBINACA 2-fluorobenzyl isomer), notified in 2014, contains such a tail, which would appropriately be abbreviated 2F-BZ, generating the semi-systematic name AMB-2F-BZINACA.

New structural features within an SC can either be assigned a letter code or a concatenated substitution letter code. The decision on which is appropriate is illustrated in Figure 11. The linked group is represented via a concatenated letter code with the benzyl moiety (BZ) and the methoxy substituent (MeO). The locant clarifies the position of the substituent at the BZ. On the other hand, the tail group is not represented via a concatenated letter code. This is because the substitution syntax only applies to a specific common element if

identified without a substitution. Only then is a modification of this common element, via substitution, introduced into the name. For example, although the piperidine methyl moiety can be considered to be modified through methylation, the tail group piperidine methyl has not been identified in an SC before. Consequently, the tail group is represented via the letter code MPIM, abbreviating for methyl piperidine methyl and not termed 1Me-PIM.

3.3.3 | Examples of initial naming approach inconsistencies for the tail building blocks

One of the most prominent examples of inconsistent naming has been the butyl side chain group and the benzyl side chain group. The reference standard for *N*-(1-amino-3,3-dimethyl-1-oxobutan-2-yl)-1-benzyl-1*H*-indazole-3-carboxamide has been available from Cayman Chemical and was used to produce an analytical report utilizing the name ADB-BINACA, denoting the benzyl side chain moiety with the letter B. In November of 2019, the SC *N*-(1-amino-3,3-dimethyl-

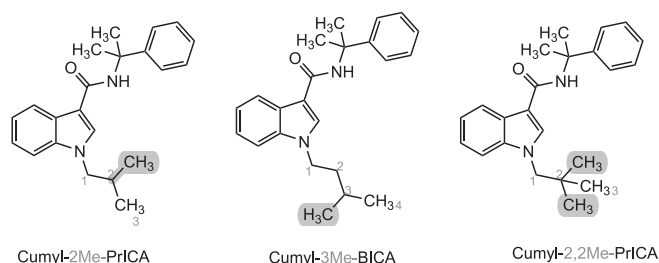


FIGURE 10 Examples of branched alkyl chains as tail groups named using the substitution syntax. These SCs have not been detected on the drug market so far.

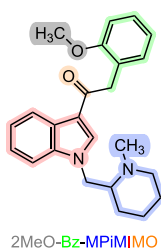


FIGURE 11 Example of the differentiation of substitution and introduction of a new building block of 2MeO-Bz-MPiMIMO (also known as JWH-250 1-[2-methylene-N-methyl-piperidyl] derivative) [Colour figure can be viewed at wileyonlinelibrary.com]

1-oxobutan-2-yl)-1-butyl-1*H*-indazole-3-carboxamide was identified by Sweden and reported to the EDND following a seizure of the pure research chemical. In line with the abbreviation of pentyl side chains with the single letter P in the past, the direct conclusion would have been to choose B for the butyl side chain. Instead, the letter code BUT was selected because B was previously used for benzyl. Under the EMCDDA framework for SCs, a butyl group is abbreviated with the letter B and the benzyl group with the letter code BZ. Since then, the hydrazide SCs have been notified to the EMCDDA, which all feature a benzyl and carbonyl group in the linked group. The letter code established for this linked group was BZO derived from benzoyl (see Table 9). This current naming is in line with the letter code BZ, used for the benzyl group within SCs.

The tail group building block has been modified in three ways in SCs, first identified in Germany. Previously, cyclobutyl methyl, norbornyl methyl, and tosyl tail groups were identified in herbal materials, circumventing the legislation in Germany at the time of occurrence.^{30–34} Upon identification, we proposed the abbreviation of BC[2.2.1]Hp for the newly emerged norbornyl moiety (see Figure 12). The proposed letter-number code provides the greatest possible way to represent small changes within this bicyclic system in the future. However, although this thought is particularly future-proofed, future development is uncertain, and a tail group structure resembling the norbornyl element might not emerge. Therefore, under the framework for naming SCs described in this work, the correct way to abbreviate the norbornyl tail group element is NB, which ties in with the established letter code MEGACLONE to form the semi-

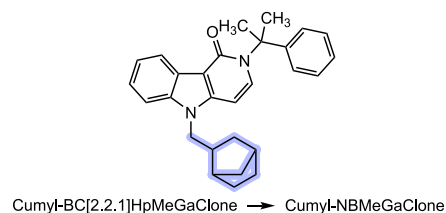


FIGURE 12 Illustration of adjusting previously assigned letter codes to follow the framework [Colour figure can be viewed at wileyonlinelibrary.com]

systematic name for the SC Cumyl-NBMeGaClone shown in Figure 12. Notably, the γ -carbolinone core structures have been excluded from the EMCDDA framework to retain the recognizable “PEGACLONE” and, just like in this case, the “MEGACLONE” letter code. However, it has to be noted that single letters are used for all other tail groups with alkyl elements (see Table 5).

Substitution of an alkyl tail element with a halogen atom yields a haloalkane found extensively in SCs. Commonly, this modification has been amended as a prefix to the name of the unsubstituted SC. As a result, this wrongfully leads to the assumption of the linked group being halogenated. Under the expanded syntax, the additional substitution directly precedes the letter code of the respective building block. For example, the previously named 5F-CUMYL-PINACA should be written CUMYL-5F-PINACA as the fluorination is located at the pentyl tail element, not at the linked group (see Figure 13).

Practical information

- A butyl group is abbreviated with the letter B and the benzyl group with the letter code BZ (e.g., butyl tail group in ADB-BINACA)
- An unsaturated tail and substitution of a tail group are considered in the same way. The location and type of modification/substitution is added as prefix to the modified tail.
- The 4-fluorobenzyl building block is permanently assigned FUB
- New structurally similar building blocks must not build on the letter code FUB (e.g., 3-fluorobenzyl would be assigned the letter code 3F-BZ)

3.4 | Linked group building blocks

The linked group in an SC is always bound to the linker, which acts as the connection with the core and, concomitantly, the tail group. Linked groups are the most heterogeneous group of building blocks present in SCs. Furthermore, the benzyl-linked group is the moiety that has been modified most through substitution. Therefore, this chapter is divided into four sections. The first section details the branched linked groups with the amide and ester functions. Here,

three linked groups are assigned letter codes based on our substitution syntax for the first time. The second section shows the linked groups with cyclic features and their derived letter codes. Information on the linked groups created from the substitution of NA is detailed in

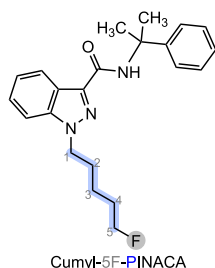


FIGURE 13 Molecular structure of Cumyl-5F-PINACA [Colour figure can be viewed at wileyonlinelibrary.com]

the third section. The fourth section covers the linked groups derived from the common feature of phenyl and benzyl.

3.4.1 | Branched linked groups

The majority of the linked groups in Table 7 is structurally related to the amino acids valine and *tert*-leucine with varying degrees of methylation of the carboxyl group and the exchange for a CA. Abbreviations that have been used initially for an entire SC have evolved into the letter code for a specific linked group moiety. Figure 14 illustrates this using the SC, commonly known as AMB. The fluorinated analog of AMB was termed 5F-AMB, and ultimately, AMB was utilized to describe only the methyl 3-methylbutanoate linked group element. This led to a letter code that does not contain accurate information on the structure of the linked group it represents. Furthermore, using a three-letter code to describe an entire SC does not

TABLE 7 Structural moieties, IUPAC names, and assigned letter codes for aliphatic linked group elements identified within SCs

Structure element	Letter code	Systematic name	Structure element	Letter code	Systematic name
	M	methyl		MBA	3-methylbutanoic acid
	AB	1-amino-1-oxobutane		DMBA	3,3-dimethyl-butanoic acid
	AMB	1-amino-3-methyl-1-oxobutane		MMB	methyl 3-methyl-butanoate
	ADMB	1-amino-3,3-dimethyl-1-oxobutane		MDMB	methyl 3,3-dimethyl-butanoate
	TMCP	tetramethyl-cyclopropane		CHMDMB	cyclohexyl-methyl 3,3-dimethyl-butanoate
	2,3,3Me-1en-B	2,3,3-trimethyl-but-1-en		EMB	ethyl 3-methyl-butanoate
	2,2,3Me-3en-B	2,2,3-trimethyl-but-3-en		EDMB	ethyl 3,3-dimethyl-butanoate
	2,2,3Me-3OH-B	2,2,3-trimethyl-3-hydroxy-butane			

cover the structural features of a core, tail, linker, and linked group. Among the established letter codes utilized for older linked groups, several begin with an A, which indicates the amino function that the linked group described as AMB lacks. The accurate letter code abbreviating for the structural features of methyl 3-methylbutanoate is MMB (see Table 7).

Other letter codes of the past that resulted in inconsistencies were AB abbreviating for 1-amino-3-methyl-1-oxobutanoate, ABO abbreviating for 1-amino-1-oxobutane, and ADB abbreviating for 1-amino-3,3-dimethyl-1-oxobutane. Furthermore, the two letter codes AMB and MMB have been used synonymously for the same structural feature, which is not helpful when accurately referring to a single SC. The letter code AB does not include the M feature included in the other two letter codes. Also, the letter code ABO does not accurately represent the systematic name it is supposed to represent. The naming order of the structural elements is different in the systematic and letter code representations. The semi-systematic name assigned under the EMCDDA framework aims to consistently represent structural features in the order in which they are found in the systematic name. Consequently, the assigned letter codes of the abovementioned structural features were renamed to obtain letter codes that accurately correlate with the systematic names (see Table 8).

Some letter codes shown in Table 7 have been used interchangeably in the past. The letter code AB now describes a different structural feature, namely, 1-amino-1-oxobutane. Compared to AB, AMB adds the letter code M and thus indicates the M group present in the linked group. The addition of another M group is subsequently indicated by the multiplicative prefix “di-” represented in the letter code ADMB. Lastly, methyl 3-methylbutanoate is abbreviated using MMB, which was formerly used synonymously with AMB.³⁵ Here, the carbonyl group is part of the ester. The systematic name does not imply a particular representation of the carbonyl group compared to the amide functional group in AB, AMB, and ADMB. Because of this, the carbonyl group is not reflected in the letter code of the butanamides either. These changes are not entirely new, and the resulting semi-systematic names can be found in the literature already.^{36,37}

For many linked groups, the IUPAC name exhibited relatively few words that had to be abbreviated to reflect the structural features. However, the three linked group moieties shown in Figure 15 pose a more significant challenge as the systematic name is complex.

Therefore, these linked groups fall under the case of substituting a principal characteristic group. The principal characteristic group is a C₄-chain, substituted with M and hydroxy functional groups, in all three cases. The locants, starting at the bond connecting the linked group to the linker, are included to indicate the number and position of functional groups of one particular kind, as well as the position of unsaturation.

3.4.2 | Cyclic linked groups

The letter codes for the linked groups that contain cyclic elements are shown in Table 9. Notably, the NA-linked group has been extensively substituted in SCs that have appeared on the drug market (see Table 10). The unsubstituted NA-linked group element is an example of simplifying letter codes of structural features due to the time of occurrence. The first NA moiety, JWH-018, which was formally notified in 2008, was connected to the linker at Position 1. In the later identified SC 2NA-5F-PICA (also known as 5F NNEI 2'-naphthyl isomer), the NA moiety was attached to the linker at Position 2. To differentiate the new SC from the previously detected isomer, the new configuration of the linked group to the linker is described with a locant in front of the letter code of the linked group. The rule applied here states that the first identification of a unique building block element does not include locants within its letter code. However, when a structural isomer of the same building block occurs, its derived letter code must contain a locant to retain uniqueness within the letter codes. In Table 10, substituted NA-linked group elements that have emerged are presented with assigned letter codes under the developed framework. So, the common abbreviation of the substituted functional group is prefixed with a hyphen to the letter code NA.

3.4.3 | Substituted phenyl- and benzyl-derived linked groups

In Table 11, substituted benzyl and phenyl-linked group elements that have emerged up until now are presented with assigned letter codes under the developed framework. The abbreviation of the functional group substituted to the benzyl or phenyl element is added with a

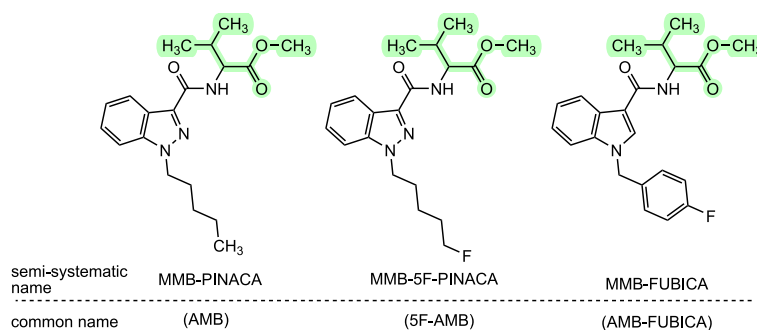
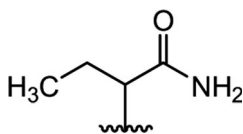
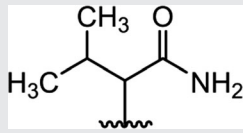
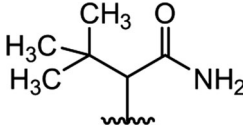
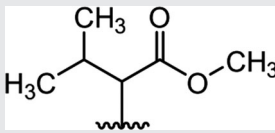


FIGURE 14 Molecular structures for MMB-PINACA, MMB-5F-PINACA, and MMB-FUBICA, also known as AMB, 5F-AMB, and AMB-FUBICA [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 8 Challenging letter codes of linked group elements due to old and new naming concepts

Structure element	Systematic name	Old letter code	New letter code
	1-amino-1-oxobutane	ABO	AB
	1-amino-3-methyl-1-oxobutane	AB	AMB
	1-amino-3,3-dimethyl-1-oxobutane	ADB (MAB)	ADMB
	methyl 3-methylbutanoate	MMB (AMB)	MMB

hyphen as a prefix to the letter code BZ or PH, respectively. The locant is added directly in front of the abbreviation for the substituent without an additional hyphen.

Practical information on building blocks

- Deriving any new letter code should be done, including the principle of previous abbreviations
- The acetamide linker is abbreviated using ATA under the EMCDDA framework
- Bicyclic core building blocks are always 1,3-disubstituted with the tail and linker group; no locants are required
- Locants are required to indicate the connection of the pyrazole-type cores to other building blocks
- Locants are always introduced without a hyphen in front of a substitution or in front of a building block when the connection to other parts of the SC has to be clarified
- Branched alkyl chains are considered substitutions
- Multiplicative prefixes are omitted from the letter code
- The letter code AB now describes 1-amino-1-oxobutane
- 1-Amino-3-methyl-1-oxobutane (formerly known as AB) is now referred to as AMB
- The letter code MMB is used to describe methyl 3-methylbutanoate (formerly sometimes referred to as AMB)
- Substitution can only occur if the modified building block has been identified without substitution before
- Substitution always requires the locant and type of substitution as a prefix to the letter code abbreviating the modified building block

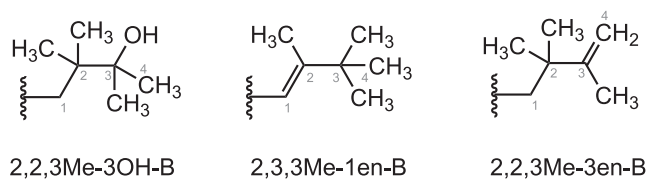
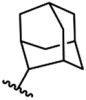
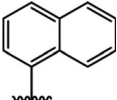

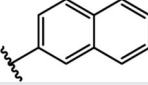
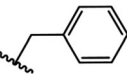
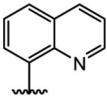
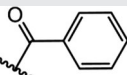

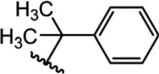

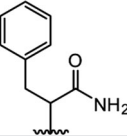
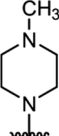
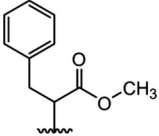
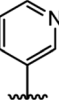


FIGURE 15 Three highly branched linked groups have occurred on the illicit drug market in the past. Due to the lack of trivial names, which accurately describe these linked groups, the building block is abbreviated using the substitution syntax. Assignment of locants starts at the atom connecting the linked group to the linker.

4 | FLOWCHART FOR THE NAMING OF AN SC FOLLOWING THE EMCDDA FRAMEWORK GUIDELINES

The process of deriving a name for an SC with a given molecular structure is outlined in Figure 16. The first step is to identify the framework's elements: core, linker, linked group, and tail. Substitution of a building block is included in the second step and always builds on an already established building block identified before. If an identified building block cannot be abbreviated using already assigned letter codes, it may be considered substituted, or a new building block is defined. The decision between substitution and introducing a new building block is also influenced by the closest building block identified before. For example, suppose the building block with the most similarity is substituted, the new SC would

TABLE 9 Structural moieties, IUPAC names, and assigned letter codes for cyclic linked group elements identified within SCs

Structure element	Letter code	Systematic name	Structure element	Letter code	Systematic name
	A	adamantyl		NA	1-naphthyl
	PH	phenyl		2NA	2-naphthyl
	BZ	benzyl		QU	8-quinolinyl
	BZO	benzoyl		CH	cyclohexyl
	CUMYL	functional group: Cumen		PY	1-pyrrolidinyl
	APP	1-amino-1-oxo-3-phenylpropane		MPIP	4-methyl-piperazin-1-yl
	MPHP	methyl 3-phenylpropanoate		PD	pyridin-3-yl

now be described through the substitution of the substitution. To avoid nested substitution in this case, a new letter code should be assigned.

Exceptions to the framework

- Exception for γ -carbolinones: single letter code tail groups with a consonant after the first letter in the systematic name are abbreviated with the first two letters (e.g., Cumyl-5F-PeGaClone).
- The 4-fluorobenzyl building block is permanently assigned FUB.

The name of an entirely new individual building block is deduced from common trivial names for the element or the IUPAC name, and each building block is considered individually when deriving a name. Previously assigned letter codes should be used wherever possible. Trivial names are favored over the IUPAC denotation to deduce the

name of an entirely new moiety. The same rule applies to modifying building blocks through substitution where the previously utilized letter codes are used where possible. The overall number of existing letter codes, which abbreviate building blocks, should be kept as low as possible. The generation of letter codes and substituents should also be informed, where possible, from standard abbreviations of organic chemistry, where abbreviated structural features are integral to communicating reactions of molecules concisely. The last step is the combination of the letter code of the four building blocks according to the naming syntax.

As described previously, the elements within the molecular structure of the SC shown in Figure 17 are first sectioned off and assigned building blocks, where possible. The linker and linked group moiety were identified by comparing the top part of the molecule to all the linker and linked groups shown in Table 1 and Table 7. The linker building block is methanone represented by the letter code MO and illustrated in orange. The linked group building block

TABLE 10 Structural moieties, IUPAC names, and assigned letter codes for substituted naphthyl-linked group elements identified within SCs

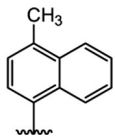
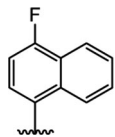
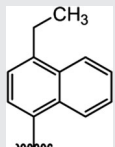
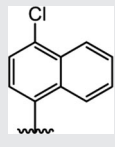
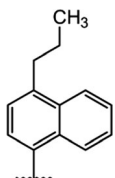
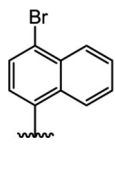
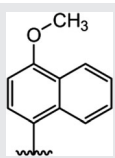
Structure element	Letter code	Systematic name	Structure element	Letter code	Systematic name
	4Me-NA	4 -methyl-naphthalen-1-yl		4F-NA	4 -fluoro-naphthalen-1-yl
	4Et-NA	4 -ethyl-naphthalen-1-yl		4Cl-NA	4 -chloro-naphthalen-1-yl
	4Pr-NA	4 -propyl-naphthalen-1-yl		4Br-NA	4 -bromo-naphthalen-1-yl
	4MeO-NA	4 -methoxy-naphthalen-1-yl			

TABLE 11 Structural moieties, IUPAC names, and assigned letter codes for substituted benzyl- and phenyl-linked group elements identified within SCs

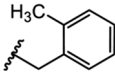
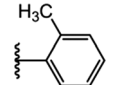
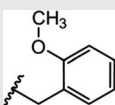
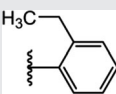
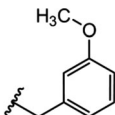
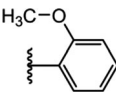
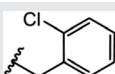
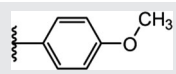
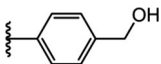

Structure element	Letter code	Systematic name	Structure element	Letter code	Systematic name
	2Me-BZ	2-methyl-benzyl		2Me-PH	2-methyl-phenyl
	2MeO-BZ	2-methoxy-benzyl		2Et-PH	2-ethyl-phenyl
	3MeO-BZ	3-methoxy-benzyl		2MeO-PH	2-methoxy-phenyl
	2Cl-BZ	2-chloro-benzyl		4MeO-PH	4-methoxy-phenyl

TABLE 11 (Continued)

Structure element	Letter code	Systematic name	Structure element	Letter code	Systematic name
	4HOME-PH	4-hydroxymethyl-phenyl		2I-PH	2-iodo-phenyl

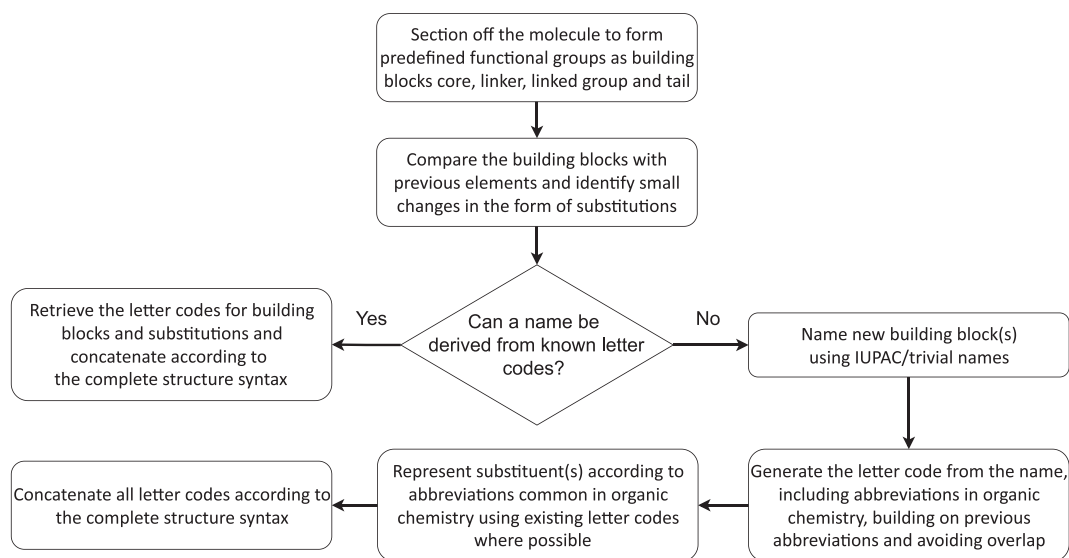


FIGURE 16 Flowchart for the naming of SCs following the EMCDDA framework guidelines

is naphthyl represented by the letter code NA and illustrated in green. Lastly, a pentyl moiety is connected to the PYO feature and illustrated in blue. Compared with the structural moieties described here, this part of the molecule is identified as a tail group represented by the letter code P.

The building block that has not been identified yet is the core illustrated in red. There are two ways to proceed in this case. Either one defines the entire unassigned part of the molecule as a novel core building block or only a part is considered the core element. In this case, the residual element (a phenyl) becomes a substitution of the core. There has been one SC notified to the EMCDDA in which a PYO makes up the core building block (also known as JWH-145), and thus, it might sound reasonable to proceed via substitution of the PYO leading to 5PH-PYO. Because in the SC 1-naphthyl-[5-(*o*-tolyl)-1-pentyl-pyrrol-3-yl]methanone (JWH-370), the phenyl-PYO (PHPYO) building block is substituted with a M group, a new core building block has to be defined in this case. If the substitution of a PYO via a phenyl ring was chosen, the SC on the right would then

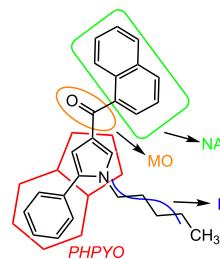


FIGURE 17 Example of the process of deriving a semi-systematic name for a “new” SC [Colour figure can be viewed at wileyonlinelibrary.com]

require a nested substitution, which is omitted from the EMCDDA framework to retain a recognizable syntax (see Section 2.1). After concatenating all letter codes, the semi-systematic name NA-PHPPYOMO is generated.

Most important rules

- Divide the SC into the four building blocks: core, linker, linked group, and tail.
- Limit the vocabulary of letter codes; reuse the previous abbreviation.
- Modify building blocks by single substitution when only minor changes occur.
- Letter codes are joined in the order: Linked group-TailCoreLinker (prefixes for substituents added with hyphens where necessary).
- Letter codes originated from trivial or systematic names (order and abbreviation of characteristic groups).
- Rules for assigning locants to characteristic groups follow the IUPAC nomenclature.
- The number of substituents is derived from the locants; multiplicative prefixes are not used.
- The EMCDDA framework is case-insensitive; uppercase and lowercase letters can be used to increase readability.

5 | CONCLUSION

The chemical diversity of SCs continues to grow dynamically making necessary the establishment of a consistent framework for the semi-systematic naming, bringing consistency for all actors facing this phenomenon.

The EMCDDA model for the naming of SCs, developed in 2013, was used as the blueprint for the new EMCDDA framework. As a result, formerly assigned names remain largely the same, and approximately half of the names used by the EMCDDA to monitor the SCs identified in the EU do not deviate from the names derived by this framework. The highly recognizable terms PEGACLONE, MEGACLONE, and FUB were retained in the EMCDDA framework and are explicitly mentioned as exceptions. Pentyl and M elements connected to γ -carbolinones and fluorinated benzyl elements were not changed to be consistent with other letter codes.

All SCs monitored by the EMCDDA at the time of publication were used as a basis to deduce consistent letter codes for all four building blocks (core, linker, linked group, and tail) as defined by the basic syntax. The expanded syntax includes additional flexibility to cover the substitution of each building block. Substitution of building blocks is part of the letter code of a building block to reflect minor changes without entirely defining a new abbreviation for the building block. The combination of letter codes for all elements results in an accurate representation of the complete SC. Ultimately, the objective is to use a letter code derived from the EMCDDA framework that is as unambiguous as possible. This framework aims to provide user-friendly names that facilitate the work of professionals in the field of NPS, including researchers and policy-makers.

The scope of application for the EMCDDA framework has intentionally been limited to SCs, which have emerged on the drug market. The framework would have to be as precise as possible and account for potential chemical features that have not yet appeared on the drug market. The syntax's overall complexity can be reduced by defining the set of structures that have to be differentiated.

The developed framework streamlines the process of assigning letter codes to newly emerging SCs. The main principle of the EMCDDA framework is to establish clear rules to assign common names and to reuse previously assigned letter codes, making it easier to recognize the molecular structures. The overall vocabulary should be limited by sharing abbreviations of common features across building blocks. Whenever a letter code has to be derived for a new moiety, all previous building blocks and the derived letter codes have to be considered to maximize consistency and minimize complexity.

This framework provides the rationale on how names for SCs are derived and, more importantly, practical guidance and examples on how semi-systematic names for SCs should be derived. With the globalization of the market in SCs, there is a need for a concerted effort and worldwide harmonization in the naming of emerging SCs.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Folker Westphal for the continued support, productive discussions, and review and Dr. Felix Bächle for support on the graphical design of building blocks. The authors also wish to thank EMCDDA: Dr. Roumen Sedefov, Andrew Cunningham, Michael Evans-Brown, Dr. Rita Jorge, Joanna de Morais, and Dr. Gregorio Planchuelo. This work was created under EMCDDA contract CT.21.SAS.0108.1.0. Svenja Fischmann gratefully acknowledges the Internal Security Fund of the European Union (grant no. IZ25-5793-2016-27) for the funding of the project ADEBAR. Benedikt Pulver gratefully acknowledges the Internal Security Fund of the European Union (grant no. IZ25-5793-2019-33) for the funding of the project ADEBAR *plus*. The ADEBAR projects provided information essential to the contract/work. Additionally, the authors wish to extend their thanks to Dr. Barry Logan and Dr. Alex Krotulski, NMS Labs, and CFSRE, for fruitful discussions and support of the EMCDDA framework.

DATA AVAILABILITY STATEMENT

Data available on request from the authors.

ORCID

Benedikt Pulver  <https://orcid.org/0000-0002-7772-2111>

REFERENCES

1. Auwärter V, Dresen S, Weinmann W, Müller M, Pütz M, Ferreirós N. Spice' and other herbal blends: harmless incense or cannabinoid designer drugs? *J Mass Spectrom*. 2009;44(5):832-837. doi:10.1002/jms.1558
2. European Monitoring Centre for Drugs and Drug Addiction. *Understanding the 'spice' phenomenon*. Publications Office; 2009. Accessed April 3, 2022. <https://data.europa.eu/doi/10.2810/27063>
3. Huffman J, Padgett L. Recent developments in the medicinal chemistry of cannabimimetic indoles, pyrroles and indenenes. *Curr Med Chem*. 2005;12(12):1395-1411. doi:10.2174/0929867054020864
4. Makriyannis A, Deng H. Cannabimimetic indole derivatives. Published online July 10, 2007. Accessed April 27, 2022. <https://patents.google.com/patent/US7241799B2/en>

5. Liu R, Li X, Lam KS. Combinatorial chemistry in drug discovery. *Curr Opin Chem Biol*. 2017;38:117-126. doi:10.1016/j.cbpa.2017.03.017
6. Uchiyama N, Kawamura M, Kikura-Hanajiri R, Goda Y. Identification of two new-type synthetic cannabinoids, N-(1-adamantyl)-1-pentyl-1H-indole-3-carboxamide (APICA) and N-(1-adamantyl)-1-pentyl-1H-indazole-3-carboxamide (APINACA), and detection of five synthetic cannabinoids, AM-1220, AM-2233, AM-1241, CB-13 (CRA-13), and AM-1248, as designer drugs in illegal products. *Forensic Toxicol*. 2012;30(2):114-125. doi:10.1007/s11419-012-0136-7
7. Grigoryev A, Kavanagh P, Melnik A. The detection of the urinary metabolites of 3-[(adamantan-1-yl)carbonyl]-1-pentylindole (AB-001), a novel cannabimimetic, by gas chromatography-mass spectrometry. *Drug Test Anal*. 2012;4(6):519-524. doi:10.1002/dta.350
8. Banister SD, Wilkinson SM, Longworth M, et al. The synthesis and pharmacological evaluation of adamantane-derived indoles: cannabimimetic drugs of abuse. *ACS Chem Neurosci*. 2013;4(7):1081-1092. doi:10.1021/cn400035r
9. Frost JM, Dart MJ, Tietje KR, et al. Indol-3-ylcycloalkyl ketones: effects of N1 substituted indole side chain variations on CB₂ cannabinoid receptor activity. *J Med Chem*. 2010;53(1):295-315. doi:10.1021/jm901214q
10. Sekuła K, Zuba D, Lorek K. Analysis of fragmentation pathways of new-type synthetic cannabinoids using electrospray ionization. *J Am Soc Mass Spectrom*. 2018;29(10):1941-1950. doi:10.1007/s13361-018-2008-9
11. Uchiyama N, Kawamura M, Kikura-Hanajiri R, Goda Y. URB-754: a new class of designer drug and 12 synthetic cannabinoids detected in illegal products. *Forensic Sci Int*. 2013;227(1-3):21-32. doi:10.1016/j.foresciint.2012.08.047
12. EMCDDA. *Synthetic cannabinoids in Europe*. European Monitoring Centre for Drugs and Drug Addiction; 2017. Accessed April 3, 2022. https://www.emcdda.europa.eu/system/files/publications/2753/POD_Synthetic%20cannabinoids_0.pdf
13. Cayman Chemical. Synthetic cannabinoid flipbook. Accessed April 22, 2022. https://www.caymanchem.com/forensics/search/flipbook?p1=PH00040&p2=PH00004&p3=PH00084&fq1=*&fq2=*&fq3=*
14. Schelkun, R.M., Iula, D.M. Laboratory guide for synthetic cannabinoid identification and naming. Published February 2022. Accessed April 22, 2022. <https://www.caymanchem.com/literature/laboratory-guide-for-synthetic-cannabinoid-identification-and-naming>
15. Blaazer AR, Lange JHM, van der Neut MAW, et al. Novel indole and azaindole (pyrrolopyridine) cannabinoid (CB) receptor agonists: design, synthesis, structure-activity relationships, physicochemical properties and biological activity. *Eur J Med Chem*. 2011;46(10):5086-5098. doi:10.1016/j.ejmech.2011.08.021
16. Uchiyama N, Matsuda S, Wakana D, Kikura-Hanajiri R, Goda Y. New cannabimimetic indazole derivatives, N-(1-amino-3-methyl-1-oxobutan-2-yl)-1-pentyl-1H-indazole-3-carboxamide (AB-PINACA) and N-(1-amino-3-methyl-1-oxobutan-2-yl)-1-(4-fluorobenzyl)-1H-indazole-3-carboxamide (AB-FUBINACA) identified as designer drugs in illegal products. *Forensic Toxicol*. 2013;31(1):93-100. doi:10.1007/s11419-012-0171-4
17. Krotulski AJ, Mohr ALA, Logan BK. Emerging synthetic cannabinoids: development and validation of a novel liquid chromatography quadrupole time-of-flight mass spectrometry assay for real-time detection. *J Anal Toxicol*. 2020;44(3):207-217. doi:10.1093/jat/bkz084
18. Heller SR, McNaught A, Pletnev I, Stein S, Tchekhovskoi D. InChI, the IUPAC international chemical identifier. *J Chem*. 2015;7(1):23. doi:10.1186/s13321-015-0068-4
19. Heller S, McNaught A, Stein S, Tchekhovskoi D, Pletnev I. InChI—the worldwide chemical structure identifier standard. *J Chem*. 2013;5(1):7. doi:10.1186/1758-2946-5-7
20. Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J Chem Inf Model*. 1988;28(1):31-36. doi:10.1021/ci00057a005
21. Hellwich KH, Hartshorn RM, Yerin A, Damhus T, Hutton AT. Brief guide to the nomenclature of organic chemistry (IUPAC technical report). *Pure Appl Chem*. 2020;92(3):527-539. doi:10.1515/pac-2019-0104
22. Bulgarian Focal Point on Drugs and Drug Addictions. ADB-FUBIACA (N2-([1-[(4-Fluorophenyl)Methyl]-1H-Indol-3-Yl]acetyl)-3-Methylvalinamide); 2021. Accessed September 18, 2022. <https://ednd2.emcdda.europa.eu/ednd/report/view/8023>
23. EU-project ADEBAR plus. ADB-FUBIACA in Germany; 2021. Accessed September 18, 2022. <https://ednd2.emcdda.europa.eu/ednd/report/view/7939>
24. Pulver B, Fischmann S, Westphal F, et al. The ADEBAR project: European and international provision of analytical data from structure elucidation and analytical characterization of NPS. *Drug Test Anal*. Published online May 24, 2022. 14(8):1491-1502. doi:10.1002/dta.3280
25. ADB-FUBIATA. Center for Forensic Science Research and Education (CFSRE); 2021:8. Accessed February 27, 2022. https://www.npsdiscovery.org/wp-content/uploads/2021/11/ADB-FUBIATA_111721_CFSRE-Chemistry_Report.pdf
26. Liu C, Hua Z, Jia W, Li T. Identification of AD-18, 5F-MDA-19, and pentyl MDA-19 in seized materials after the class-wide ban of synthetic cannabinoids in China. *Drug Test Anal*. 2022;14(2):307-316. doi:10.1002/dta.3185
27. Schelkun RM, Krotulski AJ, Iula DM, Logan BK. *New systematic naming for synthetic cannabinoid “MDA-19” and its related analogues: BZO-HEXOXIZID, 5F-BZO-POXIZID, and BZO-POXIZID*. Center for Forensic Science Research and Education (CFSRE); 2021:1. Accessed February 7, 2020. https://www.npsdiscovery.org/wp-content/uploads/2021/08/New-Systematic-Naming-for-MDA-19-and-Related-Analogues_NPS-Discovery_083121.pdf
28. Diaz P, Xu J, Astruc-Diaz F, Pan HM, Brown DL, Naguib M. Design and synthesis of a novel series of N-alkyl isatin acylhydrazone derivatives that act as selective cannabinoid receptor 2 agonists for the treatment of neuropathic pain. *J Med Chem*. 2008;51(16):4932-4947. doi:10.1021/jm8002203
29. Gilbert N, Costello A, Ellison JR, et al. Synthesis, characterisation, detection and quantification of a novel hexyl-substituted synthetic cannabinoid receptor agonist: (S)-N-(1-amino-3,3-dimethyl-1-oxobutan-2-yl)-1-hexyl-1H-indazole-3-carboxamide (ADB-HINACA). *Forensic Chem*. 2021;26:100354. doi:10.1016/j.forc.2021.100354
30. Pulver B, Riedel J, Schönberger T, et al. Comprehensive structural characterisation of the newly emerged synthetic cannabimimetics Cumyl-BC[2.2.1]HpMeGaClone, Cumyl-BC[2.2.1]HpMINACA, and Cumyl-BC[2.2.1]HpMICA featuring a norbornyl methyl side chain. *Forensic Chem*. 2021;26:100371. doi:10.1016/j.forc.2021.100371
31. Pulver B, Riedel J, Schönberger T, et al. Dataset allowing for the identification of three new synthetic cannabimimetics featuring a norbornyl methyl side chain by spectrometric and spectroscopic techniques. *Data Brief*. 2021;39:107628. doi:10.1016/j.dib.2021.107628
32. Halter S, Pulver B, Wilde M, et al. Cumyl-CBMICA: a new synthetic cannabinoid receptor agonist containing a cyclobutyl methyl side chain. *Drug Test Anal*. 2021;13(1):208-216. doi:10.1002/dta.2942
33. Haschimi B, Grafinger KE, Pulver B, et al. New synthetic cannabinoids carrying a cyclobutyl methyl side chain: human phase I metabolism

- and data on human cannabinoid receptor 1 binding and activation of Cumyl-CBMICA and Cumyl-CBMINACA. *Drug Test Anal.* 2021;13(8): 1499-1515. doi:10.1002/dta.3038
34. Pulver B, Schönberger T, Weigel D, et al. Structure elucidation of the novel synthetic cannabinoid Cumyl-Tosyl-Indazole-3-Carboxamide (Cumyl-TsINACA) found in illicit products in Germany. *Drug Test Anal Published Online April.* 2022;9(8):1387-1406. doi:10.1002/dta.3261
 35. AMB (CAS 1890250-13-1). Accessed April 11, 2022. <https://www.caymanchem.com/product/15488>
 36. Explore ADMB-HEXINACA|PIHKAL · info. Accessed April 11, 2022. <https://isomerdesign.com/PIHKAL/explore.php?domain=pk&id=14089>
 37. Center for Forensic Science Research and Education. *ADB-HEXINACA*. Center for Forensic Science Research and Education; 2021:8. Accessed April 11, 2022. https://www.npsdiscovery.org/wp-content/uploads/2021/04/ADB-HEXINACA_042921_CFSRE-Chemistry_Report.pdf

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Pulver B, Fischmann S, Gallegos A, Christie R. EMCDDA framework and practical guidance for naming synthetic cannabinoids. *Drug Test Anal.* 2022;1-22. doi:10.1002/dta.3403