A Comprehensive Nomenclature System for Cyclodextrins

Amelia Anderson[a], Matthew S. O'Connor[a], James Pipkin[b], Milo Malanga[c], Tamas Sohajda[c], Thorsteinn Loftsson[d], Lajos Szente[e], Rebeca García-Fandiño[f*] and Ángel Piñeiro[g†]

[a]Cyclarity Therapeutics, 8001 Redwood Blvd Novato, CA 94945, USA
[b]Ligand Pharmaceuticals Incorporated, 3911 Sorrento Valley Boulevard, San Diego, CA, USA
[c]CarboHyde, Budapest, Berlini u. 47-49, 1045, Hungary
[d]Faculty of Pharmaceutical Sciences, University of Iceland, Hofsvallagata 53, IS-107 Reykjavik, Iceland
[e]CycloLab Cyclodextrin R&D Laboratory Ltd., Illatos u. 7., Budapest, H-1097, Hungary
[f]Department of Organic Chemistry, Center for Research in Biological Chemistry and Molecular Materials, University of Santiago de Compostela, CIQUS, Spain
[g]Department of Applied Physics, Faculty of Physics, University of Santiago de Compostela, Spain

[*]Corresponding author: rebeca.garcia.fandino@usc.es
[†]Corresponding author: Angel.Pineiro@usc.es

Abstract

Modified cyclodextrins (CDs) are cyclic oligosaccharides with many applications in drug delivery, catalysis, and as active pharmaceutical ingredients. In general, they exist as distributions of structurally diverse molecules rather than single-isomer compounds. Their performance depends on the number of glucopyranose units (GPUs), and the type, number, and position of chemical substitutions in their hydroxyl groups. Effectively targeting individual species within these distributions is essential for optimizing CDs for specific applications. Computational techniques can generate large datasets to AI-driven structural optimization, but the absence of a standardized nomenclature system for modified CDs presents a major barrier to progress in this direction. This lack of consensus limits effective communication, data sharing, automation, and collaboration. To address this, a clear and extensible nomenclature for modified CDs is proposed. In this framework, GPUs are treated like amino-acid residues, with unsubstituted GPUs as reference building-blocks and substituted ones considered as mutations. This approach precisely defines substitution types and patterns, resolves cyclic permutation ambiguities, and offers versatility for both simple and complex modifications, including chiral center alterations and covalently linked CD oligomers. By introducing this standardized nomenclature, we aim to enhance molecular design, improve reproducibility, and streamline both experimental and computational research in the CD field.

Note: To facilitate the adoption of the proposed nomenclature, we have developed a web-based tool named **CyDexID Generator**, which automatically assigns standardized names to modified CD monomers and generates the corresponding 3D structures: https://cydexid-app.lm.r.appspot.com.

Keywords: cyclodextrins; nomenclature; mutations; chemical substitutions; cyclic permutations; sequences.

1.- Introduction

Cyclodextrins (CDs) are cyclic oligosaccharides composed of α-*D*-glucopyranose units (GPUs) linked by α-(1→ 4) glycosidic bonds (Fig. 1). Since their discovery by Villiers in 1891[1], CDs have attracted significant interest across various fields—including pharmaceuticals, food science, cosmetics, and chemistry—due to their unique ability to form inclusion complexes with a wide range of guest molecules[2–4]. This property, combined with their biocompatibility and low toxicity, has made CDs invaluable in drug delivery systems, solubility enhancement, stabilization of volatile substances, and applications in supramolecular catalysis[2,4,5]. The most common natural CDs are α-, β-, and γ-cyclodextrins, consisting of 6, 7, and 8 GPUs, respectively. Chemical modifications at positions C-2, C-3, and C-6 of the glucose units (Fig. 1) have significantly expanded the utility of CDs, allowing for modulation of their physicochemical properties such as solubility, specificity to recognize individual molecules, complexation capacity, and reactivity[6]. These modifications have broadened the spectrum of CD applications, enabling tailored solutions for particular industrial and research needs.

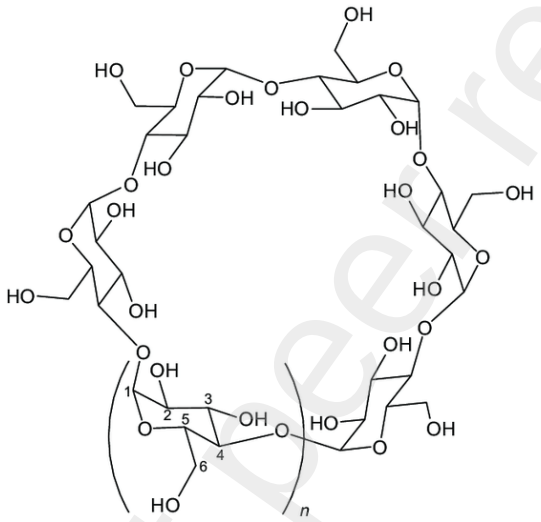

Figure 1.- Chemical structure of an α-cyclodextrin molecule (*n*=1), where each glucopyranose unit is linked via α-1,4-glycosidic bonds, forming a cyclic oligosaccharide. The hydroxyl groups attached to the C-2, C-3, and C-6 positions of each glucose unit are highlighted, representing potential sites for chemical substitutions.

As the complexity and diversity of modified CDs have increased, the opportunity to use specific structures for targeted applications, including as active pharmaceuticals, has become more evident. CDs are already prevalent in food, drug delivery, and environmental applications, where they enhance stability, solubility, and controlled release. However, synthesis typically yields mixtures of isomers, underscoring the need for precise and unambiguous structural descriptions. A standardized CD nomenclature would improve identification, quality control, and cross-industry communication, as well as facilitate intellectual property protections for specific isomers. Current naming systems cannot capture complex CD modifications or oligomeric forms [8,9], leading to inconsistencies, communication barriers, and challenges in building databases or conducting systematic computational studies. Emerging computational techniques like molecular docking, molecular dynamics, and machine learning further highlight the need for such a naming system to maximize research and application potential [8,9].

In this context, we propose a new comprehensive nomenclature and representation system for modified CDs and their oligomers. Drawing inspiration from protein nomenclature, concepts such as *residues*, *mutations*, and *hierarchical levels of structure* (primary, secondary, tertiary and quaternary)[7–9] are adapted to the structural specifics of CDs. The main objectives and applications of this proposal are: (*i*) to provide a unique and unambiguous representation for each modified CD structure; (*ii*) to facilitate precise communication among researchers and regulators as well as the creation of coherent databases; (*iii*) to enable automatic structure generation for

computational studies; (*iv*) to establish a *building-block* approach for representation and even for parameterization in MD simulations, similar to proteins; and (*v*) to improve the reproducibility and comparability of computational, and eventual wet-lab, studies involving CDs. The proposed system encodes modified CD structures in a concise character string, specifying substitution type, position, chirality, and linker presence in oligomers while resolving cyclic permutation ambiguities. A key feature of this approach is its parallelism with protein bioinformatics systems. This not only facilitates understanding and adoption by researchers familiar with bioinformatics, but also paves the way for adapting established algorithms and methodologies from protein studies to the CD domain. Alongside this string-based nomenclature, a structured data format is proposed to simplify CD information processing, and an alternative, less precise notation reduces structures to a lower-dimensional space by indicating only substitution type and count. While this compact notation is less precise than the expanded version, it remains useful for certain comparisons, especially between computational and experimental studies where full structural details are not available.

The following sections detail the proposed nomenclature systems, including descriptions, the algorithm for ensuring uniqueness, and the structured data file format. Several examples illustrating the application of the system to various modified CDs are presented—including complex but theoretically viable structures to push its limits. The system simplifies CD parametrization for computational modeling, facilitating improvements to molecular docking, MD simulations, and machine learning predictions for CDs. Finally, we discuss its limitations and future directions for development. We hope that this nomenclature and representation system will serve as a significant step toward standardization in the field of modified CDs, enhancing communication, computational research, and ultimately the rational and automated design of new CDs with optimized properties for specific applications.

2.- Background

2.1.- Structure and properties of cyclodextrins

Modified CDs are currently referred to simply by the number of GPUs in the ring, the type of substitution, and the average degree of substitution (DS) in the sample, however, this description is incomplete. Each GPU in a CD possesses three hydroxyl groups available for modification: one primary hydroxyl at the C-6 position and two secondary hydroxyls at the C-2 and C-3 positions (Fig. 1) [10]. These hydroxyl groups are the primary sites for chemical modifications that can significantly alter the physicochemical properties of CDs[11,12]. In this work, these substitutions are interpreted as mutations in the native structure of CDs, extrapolating the terminology typically employed for proteins, as will be explained later in more detail. Knowing the number of unique molecular configurations arising from mutations at specific positions is important for understanding the structural diversity, and so the expected variety of functional properties, of CDs. Ignoring the probability of each structure and focusing just on preventing the violation of chemical laws, the number of different possible structures can be determined using *Burnside's Lemma[13,14]*. This lemma allows for consideration of the rotational symmetries of CD monomers and the fact that, due to their chirality, mirror images are not equivalent. According to this proposition, and considering CDs composed of *n* GPUs, where each GPU can undergo one of *k* type of mutations, the number of unique structures (*N*) is given by:

$$N = \frac{1}{|G|}\sum_{d|n} \phi(d) \times k^{\frac{n}{d}} \quad [1]$$

where $|G|$ is the order of the symmetry group, representing the total number of rotational symmetries (for a cyclic molecule with *n* GPUs, $|G| = n$), *d* denotes each divisor of *n*, $\phi(d)$ is Euler's Totient function, which counts the number of integers up to *d* that are coprime with *d*, and *k* is the number of possible *mutations* per GPU. Eight different possibilities for each GPU unit ($k = 8$) will be considered here: native GPUs, GPUs with single modifications at C-2, C-3 or C-6, GPUs with double mutations (at positions C-2 and C-3, C-2 and C-6 or C-3 and C-6), and GPUs with all three hydroxyl groups substituted. This latter configuration, while chemically feasible

and relatively accessible[15,16], is less frequently observed in experimental samples. Under these considerations, the number of unique possible configurations for different CD types can be determined as follows:

α-CD, *n* = 6:

$$N_{\alpha\text{-CD}} = \frac{1}{6}(\phi(1) \times 8^6 + \phi(2) \times 8^3 + \phi(3) \times 8^2 + \phi(6) \times 8^1)$$

$$= \frac{1}{6}(1 \times 262{,}144 + 1 \times 512 + 2 \times 64 + 2 \times 8) = \frac{1}{6} \times 262{,}800 = \mathbf{43{,}800}$$

β-CD, *n* = 7:

$$N_{\beta\text{-CD}} = \frac{1}{7}(\phi(1) \times 8^7 + \phi(7) \times 8^1)$$

$$= \frac{1}{7}(1 \times 2{,}097{,}152 + 6 \times 8) = \frac{1}{7} \times 2{,}097{,}200 = \mathbf{299{,}600}$$

γ-CD, *n* = 8:

$$N_{\gamma\text{-CD}} = \frac{1}{8}(\phi(1) \times 8^8 + \phi(2) \times 8^4 + \phi(4) \times 8^2 + \phi(8) \times 8^1)$$

$$= \frac{1}{8}(1 \times 16{,}777{,}216 + 1 \times 4{,}096 + 2 \times 64 + 4 \times 8) = \frac{1}{8} \times 16{,}781{,}472 = \mathbf{2{,}097{,}684}$$

Thus, for monomeric α, β, and γ-CDs with a single substitution type we could, eventually, have up to 43,800, 299,600, and 2,097,684 unique structures, respectively. To illustrate this, the number of possible structures as a function of the total number of substitutions in β-CD, regardless of their location, are presented in Fig. 2, while all possible configurations for several different combinations of substitutions in α-CD are shown in Fig. 3. It is important to note that the probability of each configuration and the distribution of structures with a specific number of substitutions can vary significantly, depending on the specific chemical synthesis pathway and the subsequent purification processes. Note also that the physicochemical properties of each of these structures might significantly depend on the number and location of the substitutions, so identifying them unambiguously is key to optimizing specific applications. Equation [1] has been previously applied to determine the number of distinct substitution patterns in C6 cycles[17]. However, to the best of our knowledge, the total number of possible configurations for CD molecules, accounting for all substitution patterns, has not been reported yet.

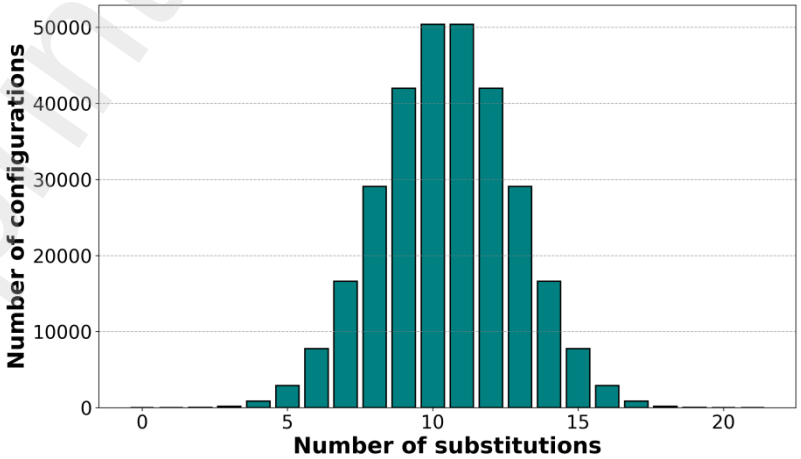

Figure 2.- Number of possible different structures for a β-CD (7 GPUs) as a function of the total number of substitutions at positions C-2, C-3 and C-6.

The previous results apply to CD monomers with a single substitution type. The number of different configurations for covalently linked CD dimers and higher-order oligomers[18,19] can be determined similarly, with a significant increase in structural diversity by several orders of magnitude due to the possible mutation combinations in each CD subunit. It is important to note that the rotational symmetry observed in monomers is lost in oligomers due to the presence of chemical linkers, which disrupt the equivalence of identical substitution configurations upon rotation of individual monomers. However, dimers with symmetric linkers where the attachment points on both CD monomers are equivalent (for instance a saturated alkyl chain bound to two CD units at C-2), lead to a new type of symmetry. In this case, certain configurations become equivalent when the two monomers are exchanged. It is worth mentioning that hybrid modifications or even conjugation of CDs with specific molecules or entities can also be present in monomers or dimers[20], increasing the dimensions of the structural space even more.

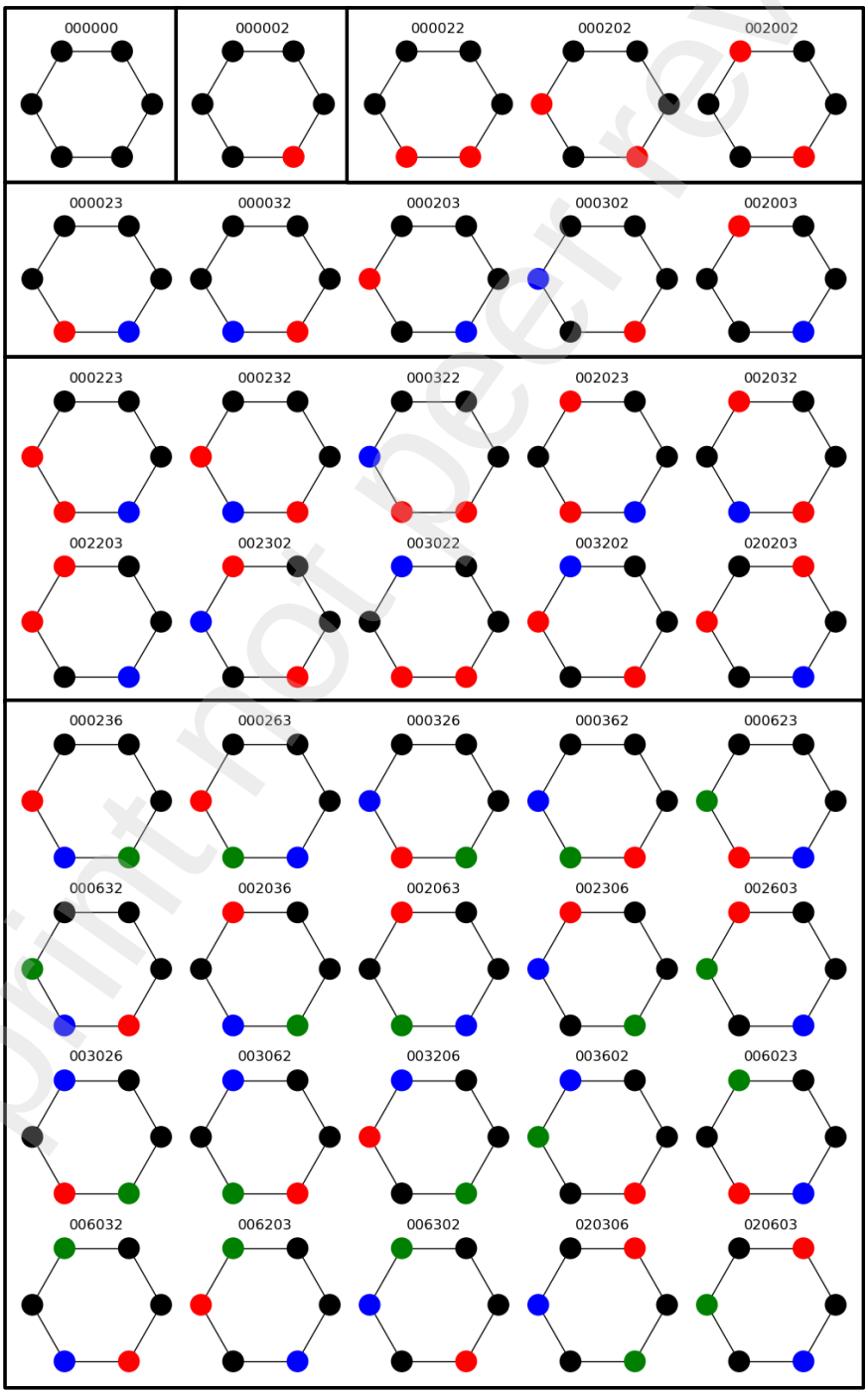

Figure 3.- All possible substitution patterns for α-CD based on specific examples of increasing complexity. Each vertex of the hexagon represents a GPU, with native (unsubstituted) GPUs depicted in black, and substitutions at positions C-2, C-3, and C-6 represented by red, blue, and green spheres, respectively. The native CDs without any substitutions or with a single substitution at one specific position have only one structure each. For the case with two identical substitutions, there are 3 possible structures. The CD with two different substitutions yields 5 structures. For CDs with three substitutions, where two may be identical and one different, or all three are different, the number of possible structures increases to 10 and 20, respectively. Altogether, the figure contains 40 distinct configurations out of the 43,800 possible unique structures for α-CD. The sequence representing the substitution pattern for each structure is indicated above the corresponding diagram (see description of the notation in the "*Description of the nomenclature system"* section). It is worth noting that if all the GPUs had different substitution patterns, considering rotational but not mirror symmetry, there would be 120 unique combinations for α-CD. For β-CD, with distinct mutations on all GPUs, the number of unique structures would increase to 720.

The take-home message of the previous calculations and discussion is that, even in the simplest cases, modified CDs can present at least several tens of thousands of structures, and, in more complex scaffolds, this number can reach into the millions or even billions. Each of these structures is expected to display different physicochemical properties that are crucial for their function, affecting their ability to encapsulate specific compounds as well as their self-assembly behavior[18,21]. One of the overarching goals of computational modeling is to enable efficient studies of these structures individually [19,22], with the results ultimately guiding the extraction of the most suitable fractions from real solutions to optimize practical applications.

The extensive number of potential substitution patterns requires a robust and systematic nomenclature framework. Current nomenclature systems for modified CDs vary widely and often lack consistency. Common approaches include abbreviated structural representations and systematic or semi-systematic names. Additionally, a modified CD may obtain patent protection and a measure of commercial use, often under a branded or trademarked name. Such is the case for CAPTISOL®, a distribution of β-cyclodextrin molecules with a variety of sulfobutylether substitutions (https://www.captisol.com/) and ADVASEP™. CAPTISOL® is also referred to as SBE-β-CD, which is an example of a semi-systematic name and abbreviated structural representation[23]. A suffix with the average number of substitutions in the sample, such as SBE-β-CD_DS6.5, is typically added to the semi-systematic names, but they are widely ambiguous because, as explained above, they may contain many thousands or even millions of different structures. The classical IUPAC names are also systematic chemical names[24]. They are precise but difficult to read and to parse into building blocks for computational codes. In 1997, a systematic nomenclature system was proposed for modified CDs[23], to describe their structural modifications, focusing on the type, position, and average number of substituents. This system uses the base CD structure (e.g., α-, β-, or γ-CD) as a starting point, followed by abbreviations for the substituents and their positions when known. For example, HP4-β-CD represents a β-CD molecule with an average of four 2-hydroxypropyl substituents, while 6-SBE1-β-CD specifies a sulfobutylether substituent attached to the 6th position. This nomenclature facilitates quick identification and categorization of modified CDs but cannot fully specify exact substitution patterns or describe the precise distribution of substitution patterns within a sample.

In general, these nomenclature systems suffer from the following limitations: (*i*) lack of uniformity across different modification types; (*ii*) inability to precisely represent the position of modifications on specific GPUs; (*iii*) difficulty in representing complex or hybrid modifications; (*iv*) challenges in computational processing and database storage; and (*v*) ambiguity in

representing dimeric or polymeric CD structures. These shortcomings highlight the need for a more comprehensive and standardized nomenclature system[25].

2.2.- Analogies between oligosaccharides and proteins

There are significant similarities between CDs and other biomolecules, particularly peptides and proteins, which are made from polymerized amino acids while CDs consist of repeating glucose units. The sequence of modifications in CDs can be compared to the amino acid sequence in peptides, suggesting the introduction of a *primary structure* for CDs, defined by the list of GPU units specifying a given substitution pattern. Moreover, the spatial arrangement of these modifications influences their function, much like *secondary structure* in protein folding. CD dimers and higher-order oligomers also parallel protein oligomers, which opens the possibility of defining a *tertiary structure* for CDs based on specific interactions between subunits, such as tail-to-tail, tail-to-head, or head-to-head orientations, among others (Fig. 4). These analogies point to the potential advantages of adopting concepts from protein nomenclature and bioinformatics to develop a more robust system for CDs.

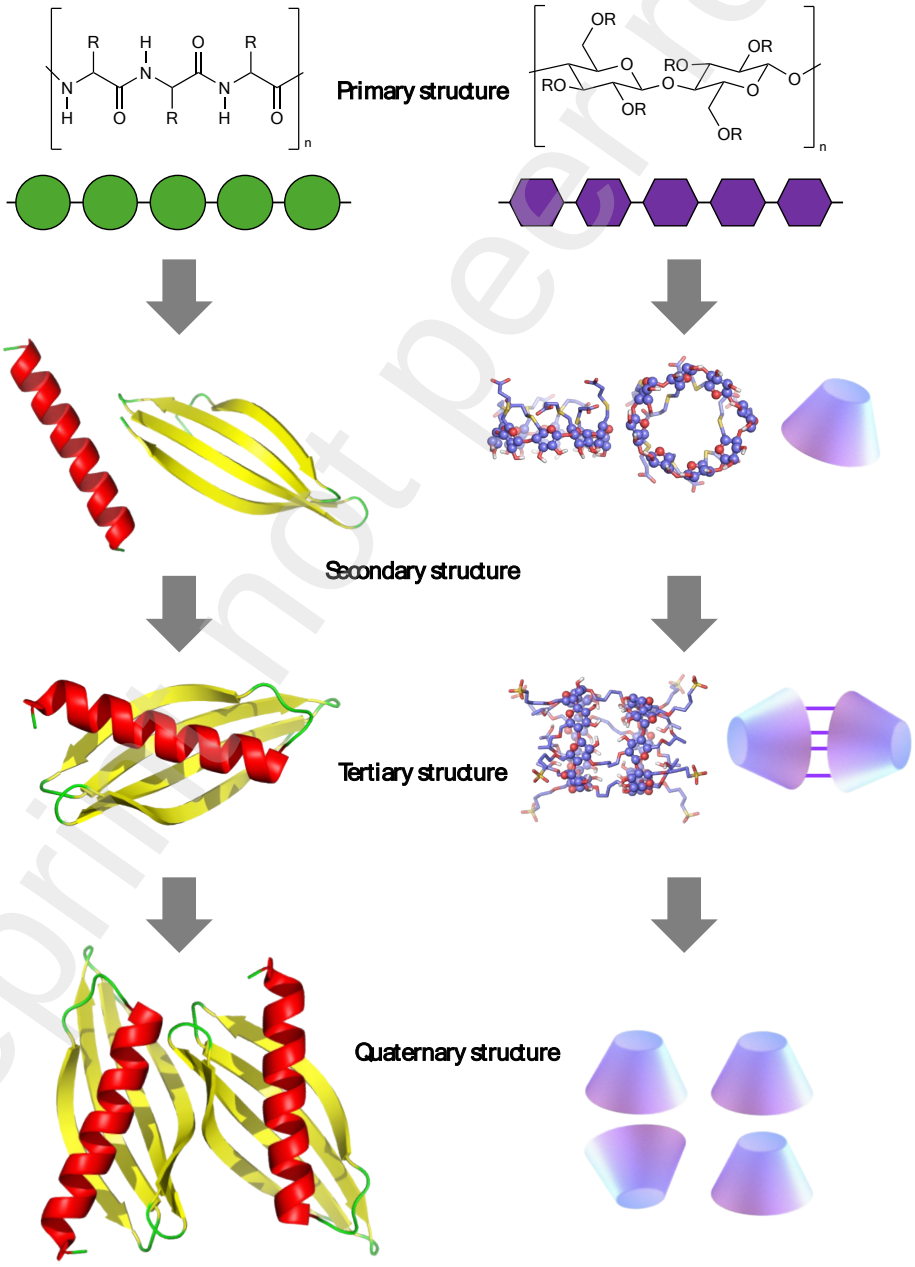

Figure 4.- Structural organization levels in proteins (left column) and cyclodextrins (right column), highlighting the similarities between the two molecular systems. From top to bottom: primary structure (linear chain of amino acids for proteins and GPUs for cyclodextrins), secondary structure (α-helix and β-sheet for proteins, and the arrangement of GPUs into cyclodextrin monomers), and tertiary structure (folded 3D structure of proteins and cyclodextrin dimers or higher order oligomers). A quaternary structure can also be defined for both as the specific but not covalent aggregation of several subunits.

While the similarities are significant and they can be exploited to define a precise nomenclature system taking proteins as a reference, the most significant difference between general protein systems and CDs is the cyclization of the chain. Although cyclic peptides exist in nature[26], they do not represent the most common structural pattern. Of course, non-cyclic oligosaccharides also exist, but here we will focus exclusively on CDs.

The main motivation to propose a new nomenclature is the growing ability to perform large-scale computational studies on CDs, including individual molecules in solution, inclusion complexes, and aggregates of homogeneous or heterogeneous CD systems[11]. While many steps of methods like docking, molecular dynamics, and quantum calculations are automated for proteins, applying them to CDs is challenging due to the lack of automated tools for CD structure building and parameterization. [7–9]. Importantly, computational approaches in the field of CDs are expected to yield more accurate results compared to protein systems, given the lower degrees of freedom of cyclic oligosaccharides[27]. A standardized, building block-based nomenclature for CDs would significantly advance computational research, aiding the design of optimized structures for specific applications through machine learning. Existing protein-focused tools, such as AutoDock Vina[28–36], AMBER[37], and GROMACS[30–36], could be adapted for CDs by incorporating ring-specific constraints and tailored force fields that better describe the unique stereochemistry and flexibility of CDs. Based on this building block scheme, some of us have recently adapted Autodock Vina to specifically consider flexibility just in the hydroxyl and substituted groups of CDs and developed specific software for automatic parameterization of modified CDs and CD dimers for further MD simulations[27]. These adaptations would ensure accurate modeling of CD systems, particularly in scenarios involving host-guest interactions or derivatized CDs with varied substitution patterns.

3.- Description of the nomenclature system

The proposed nomenclature system aims to address the limitations of existing alternatives by providing a standardized, unambiguous, and computationally tractable representation of modified CDs and their dimers. The key principles of our proposal are:

- **Uniqueness and completeness:** Each particular CD structure should have a unique identifier and the method should be able to represent any CD structure.
- **Comprehensiveness:** The CD expression should capture all relevant structural information, including the type, number, and position of modifications.
- **Scalability:** It should accommodate simple modifications, complex hybrid structures with multiple substitution types, and even oligomeric CDs with several branches or cross-links.
- **Computational friendliness:** The format should be easily parsed and processed by computer algorithms.
- **Human readability:** The system should also be easily interpretable by researchers with minimal training at a glance.
- **Analogy to protein nomenclature:** Leveraging established protein nomenclature conventions to adapt bioinformatics concepts and tools. This includes a building-block strategy for nomenclature that can also be used for digital synthesis methods, i.e. building 3D structures and obtaining parameterizations for computational experiments such as molecular dynamics simulations.

Next, the proposal will be described step by step, starting with the simplest structures and demonstrating how it meets the prerequisites outlined in the previous list of key principles.

3.1.- Encoding single and multiple substitutions of a single type for CD monomers

Each CD structure is represented by a string formed by four concatenated fields:

`[Type]_[#Substitutions]_[#GPUs]x[Positions]`

where

- `Type`: it represents an abbreviation for the substitution type (e.g., SBE for sulfobutylether, HP for 2-hydroxypropyl, ME for methyl, etc).
- `#Substitutions`: it is the total count of this modification type in the CD, the minimum number is zero for any native CDs, and the maximum would be 18, 21, and 24 for α, β, and γ-CD monomers, respectively.
- `#GPUs`: it is the number of glucose units in the CD (e.g., 6 for α-CD, 7 for β-CD, 8 for γ-CD). Nomenclature for CDs with less than 6 or more than 8 GPUs can be easily created with this system.
- `Positions`: it is a sequence indicating substitution positions on each GPU. By convention, when the primary face of the CD is oriented upwards, the sequence is traversed in a clockwise direction (Fig. 5). In order to indicate the substitution positions, the following encoding is proposed (Fig. 6):
  - **0**: for native GPUs in the corresponding position.
  - **2**, **3**, **6**: for substitutions at positions C-2, C-3, or C-6.
  - **5** (2+3), **8** (2+6), **9** (3+6): for double substitutions at the positions indicated by the numbers between parentheses.
  - **1**: for the triple substitution (2+3+6) within the same GPU.
- To indicate the stereochemistry of chiral positions, additional strings can be introduced if needed. By default, all GPU residues are assumed to have natural chirality, and their stereochemistry is not specified.
  - Chiral modifications at positions 2, 3, and/or 5 will be marked with a string starting with RS236, where the "chiral mutations" are identified by the same numbers as those used for the substitutions at positions 2, 3 and/or 6. Although the chiral center is formally at position **5**, we use **6** to denote modifications in its stereochemistry. This choice ensures consistency with the substitution notation while avoiding ambiguity, as **2+3=5** is already used to represent a different type of substitution. Additionally, since carbon 5 is covalently bound to carbon 6, this representation remains chemically intuitive (Fig. 6).
  - Chiral modifications at positions 1 and/or 4 will be marked with an extra string starting by RS14, followed by a sequence of digits 0, 1, 4, or 5, depending on the number and location of chiral mutations at positions 1, 4 or both in each residue.

Note that the fields `#Substitutions` and `#GPUs` are redundant, as they can be easily computed from the `Positions` field; however, they do not significantly increase the length of the string, as only one character is needed to include this information, and they enhance human readability as well as comprehension. Additionally, the code used to label the substitution position at each GPU is clear, concise, and easy to process. A single character represents both the number of substitutions and their locations at each GPU, while the character's position in the sequence indicates the corresponding GPU's order within the cyclic structure. The digits used to indicate the number and location of substitutions within each GPU are intuitive: **0** represents no substitution, while **2**, **3**, and **6** represent single substitutions at the corresponding positions. Double substitutions are indicated by digits resulting from the sum of the substituted positions (**5**, **8**, **9**) and, by convention, **1** is used for triple substitutions since the sum of the three digits is 11, and it is not practical to use two digits to represent a single GPU in the sequence.

SBE_5_7x0323066

which represents a β-CD with five SBE substitutions at specific positions indicated by the sequence 0323066 (Fig. 5).

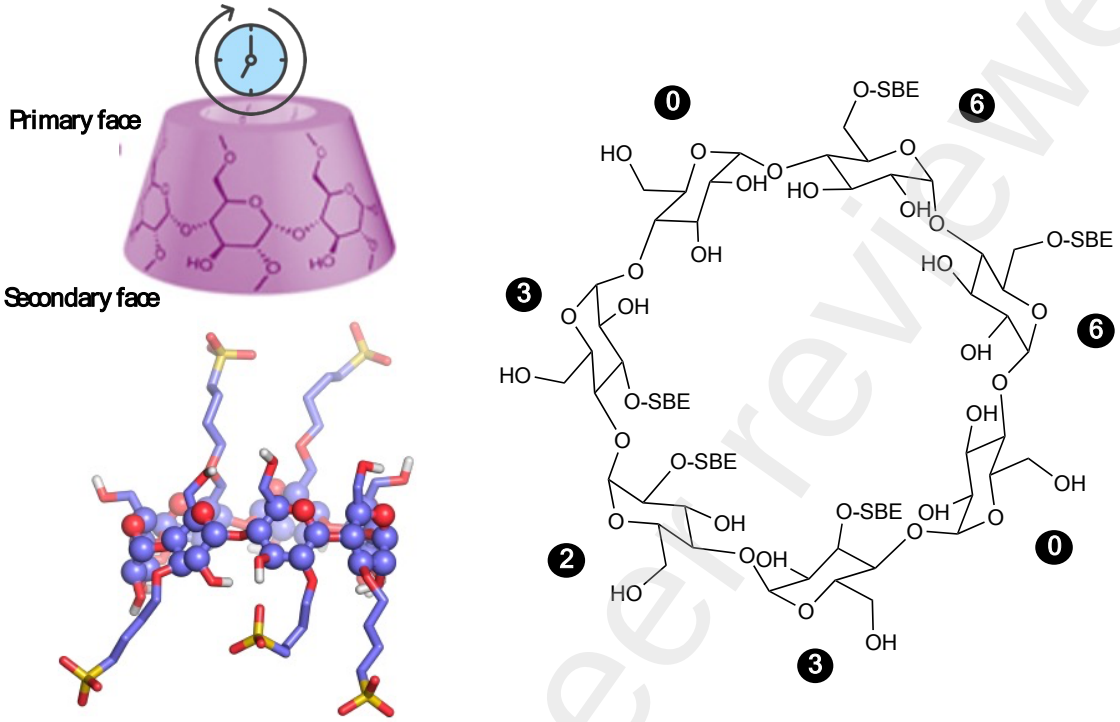

Figure 5.- Structure corresponding to SBE_5_7x0323066.

Eventual chiral modifications at different positions of this structure would be represented by a longer string:

SBE_5_7x0323066_RS236_6_7x0015200_RS14_2_7x0001004

In this last example, although the string is relatively long for a monomer, it provides a complete and unambiguous specification of the structure, detailing the type and location of all substitutions as well as the stereochemical configurations of all chiral centers.

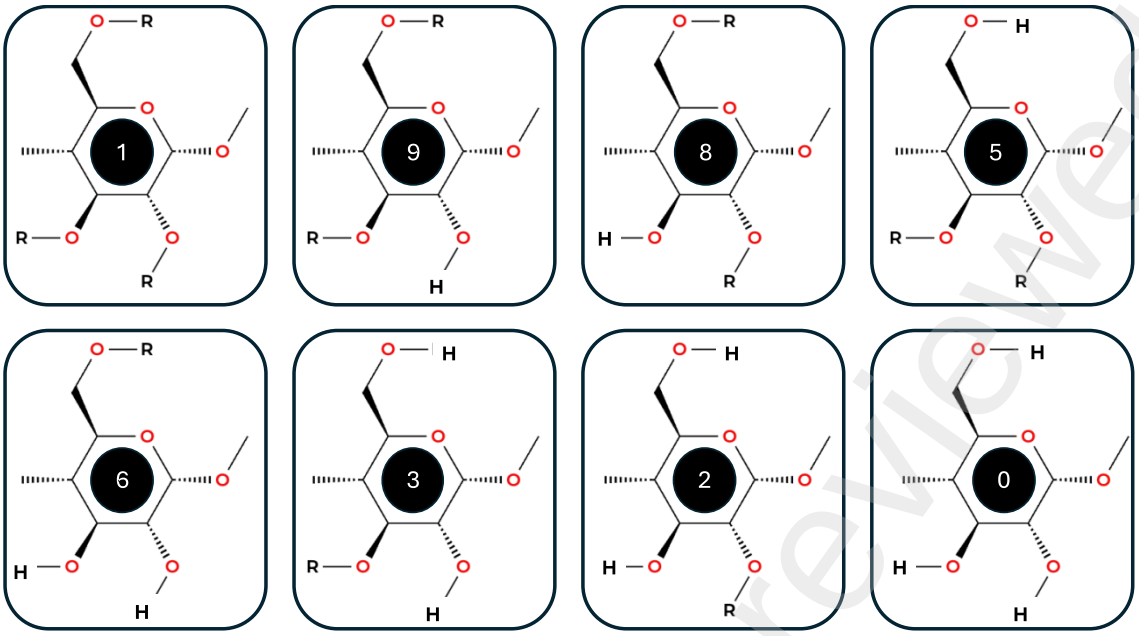

Figure 6.- Schematic representation of the encoding system for chemical modifications on a glucopyranoside ring. Each ring is labeled with a single-digit code displayed in a black circle, representing specific substitution patterns at positions 2, 3, and 6. A ring without modifications (all positions with hydrogen) is assigned the number 0. Single substitutions are encoded as 2 (position 2), 3 (position 3), or 6 (position 6). Double substitutions are summed as follows: 5 for positions 2 and 3 (2+3), 8 for positions 2 and 6 (2+6), and 9 for positions 3 and 6 (3+6). A triple substitution at positions 2, 3, and 6 is represented by the number 1, ensuring single-digit representation despite the sum 2+3+6 = 11. An equivalent numerical code is used for chiral modifications around the same positions.

3.2.- Representation of hybrid modifications for CD monomers

For CD monomers with multiple modification types, each is represented separately, ordered by the number of substitutions (descending) and alphabetically if such numbers are equal. For example:

`SBE_6_7x2306608_HP_4_7x0030010`

denotes a β-CD with both SBE and HP modifications. In the given example, all GPUs have at least one substitution. GPUs 1, 2, 4, 5, and 7 have SBE substitutions at positions C-2, C-3, C-6, C-6, and C-2+C-6, respectively, while GPUs 3 and 6 have HP substitutions at positions C-3 and C-2+C-3+C-6, respectively. This notation allows for hybrid mutations even within the same GPU, provided that the two substitution types do not occur at exactly the same position. Furthermore, this encoding is not only easy to validate computationally but also easy for a human to interpret. Another example with just one substitution per GPU is given by (Fig. 7):

`SBE_5_7x2306206_HP_2_7x0030060`

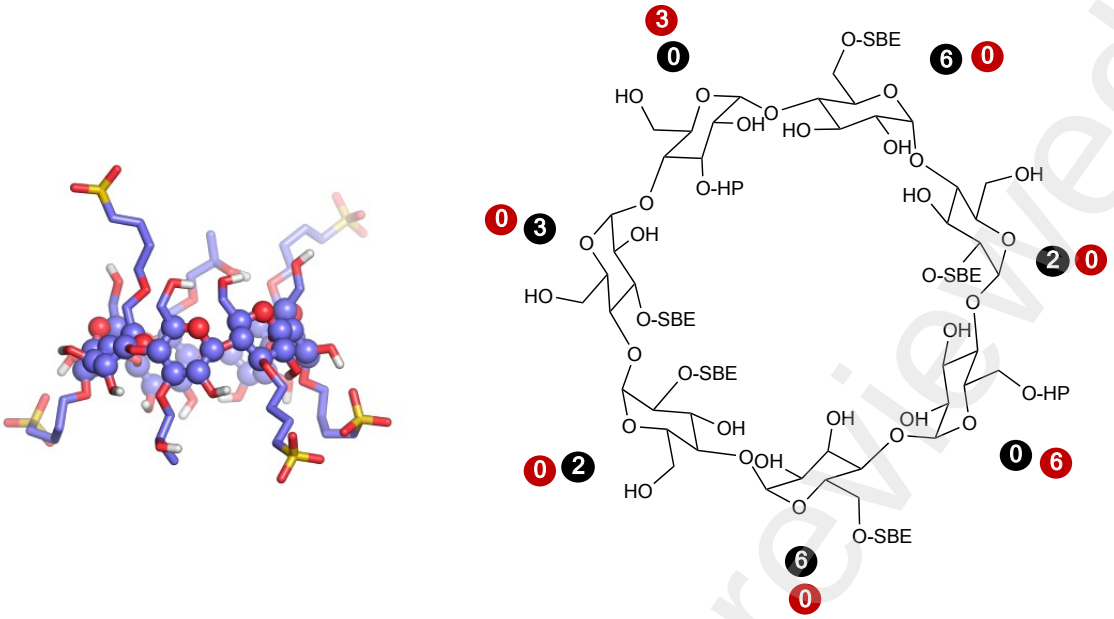

Figure 7.- Structure corresponding to `SBE_5_7x2306206_HP_2_7x0030060`.

3.3.- Algorithm for ensuring uniqueness for CD monomers

As previously discussed, CD monomers have cyclic symmetry, meaning that for a CD composed of *n* GPUs, there are *n* possible sequences that describe the same structure. This applies to CDs with either homogeneous or heterogeneous substitution patterns. To eliminate ambiguities caused by cyclic permutations, a straightforward algorithm that can be easily implemented in a computational code has been developed. The algorithm generates all possible cyclic permutations of a given sequence and assigns a unique numerical value to each permutation. This value is simply the number obtained by concatenating the digits describing each permutation. In mathematical terms, it would be obtained as follows:

$$P_i = \sum_{j=1}^{n} d_j \times 10^{(n-j)} \qquad [2]$$

where d_j is the digit at position j in the permutation. The sequence with the smallest resulting value is then selected as the canonical representation for the structure. When several substitution types are present, this process is applied to the first substitution type, and the same permutation is then used for the remaining modifications.

Example: Consider the following sequence of substitutions for a CD monomer with 6 GPUs:

230660

The cyclic permutations and the corresponding values of the metric given by equation [2] are:

P_1: 230660
P_2: 306602
P_3: 066023
P_4: 660230
P_5: 602306
P_6: 023066

The smallest value corresponds to P_6 = 023066. Therefore, the canonical representation for the sequence is:

023066

3.4.- Description of oligomers and linkers

For dimeric, and potentially multimeric, CD structures covalently linked together, the complete string consists of the independent contributions from each CD monomer, along with an additional segment representing one or more linkers between subunits. These three components are concatenated together. The definition of each monomer follows the approach outlined in the previous section, which already accounts for any possible substitution pattern in each sequence. The substring corresponding to the linker(s) is placed between the substrings of the monomers it connects. The format for this linker substring is as follows:

`_[Residue ID][Position]_[Linker Name]_[Residue ID][Position]_`

where the first `[Residue ID][Position]` refers to the GPU number and the position it is linked to in the first monomer, while the second equivalent chain provides the same information for the second monomer. Thus, for a general oligomer of CDs the nomenclature would be:

`[CD1]_[Linker(s) 1-2]_[CD2]_[Linker(s) 2-3]_[CD3]_...`

where `[CDi]` represents the string corresponding to the i-th CD, and `[Linker(s) i-j]` represents the string corresponding to the linker(s) joining `CDi` with `CDj`.

Example:

`SBE_6_7x2306603_HP_2_7x0030060`
`_13_BUT_32_`
`HP_2_6x200006_ME_2_6x006030`

This represents a highly complex dimeric structure, where the first CD is connected to the second via a BUT (butyl) linker, attached at residue 1, position 3 of the first CD, and residue 3, position 2 of the second CD (Fig. 8). The first CD has the substitution pattern described in section 3.2, while the second subunit is a CD composed of six GPUs with two HP substitutions (2-hydroxypropyl), one at position 2 of the first GPU and another at position 6 of GPU 6. Additionally, there are two ME substitutions at positions 6 and 3 of GPUs 3 and 5, respectively.

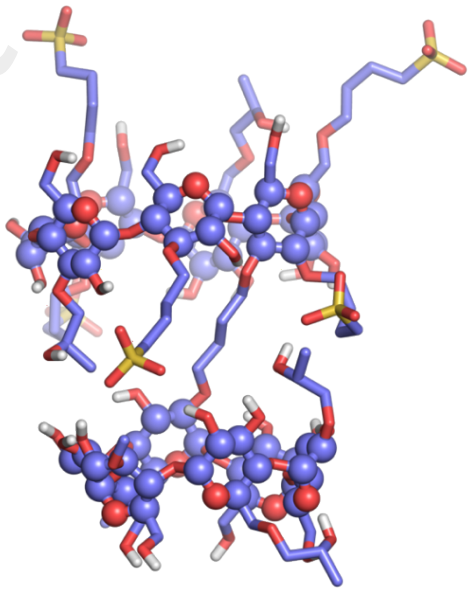

Figure 8.- Structure corresponding to `SBE_6_7x2306603_HP_4_7x0030060_13_BUT_32_HP_2_6x200006_ME_2_6x006030`.

3.5.- Complementary short notation

In addition to the full or expanded version of the nomenclature, a more compact version is introduced for cases where detailed structural information is not required. This compact version of the notation simplifies the description of modified CDs by focusing just on the type and number of substitutions without specifying their exact positions on the GPUs. Although this approach results in a loss of precision compared to the expanded notation, it remains useful in experimental contexts where the exact locations of substitutions are either unknown or considered irrelevant. This situation is common when using CDs as excipients, where specific recognition of a single ligand is less important than the ability to increase the availability and solubility of different ligands. The full version of the notation will be referred to as **CyDexID-E**, while the compact or short version will be referred to as **CyDexID-S**. The compact notation condenses the structural information into a single string per subunit, indicating the total number of substitutions and the type(s) of chemical groups involved, but omitting details about their distribution across the CD ring.

For instance, the expanded or full notation for a SBE-substituted β-CD with five modifications at specific positions on the GPUs might be expressed as `SBE_5_7x2306603` (CyDexID-E). In the short form, this could simply be written as `SBE5_7u` (CyDexID-S), where `SBE5` indicates the number and type of substitutions, `7` denotes the number of glucopyranose units, and `u` serves as a placeholder to indicate that the exact positions are unspecified. Another example with two different types of substitutions would be `SBE_5_7x2306206_HP_2_7x0030060` in the expanded form (CyDexID-E), which becomes `SBE5HP2_7u` in the compact form (CyDexID-S). Finally, the heterogeneous dimer given by `SBE_6_7x2306603_HP_4_7x0030060_13_BUT_32_HP_2_6x200006_ME_2_6x006030` in the expanded form (CyDexID-E) is written as `SBE6HP4_7u_BUT1_HP4ME2_6u` in the compact format (CyDexID-S).

The compact notation is especially useful when comparing distributions of modified CDs, such as in experimental samples, where the average DS is more relevant than the precise substitution pattern, or in cases where experimental (distribution of isomers) and computational (specific isomers) results are being compared. Despite its ambiguity, the short notation retains enough information to facilitate comparison to some extent between samples and computational predictions, while providing a simpler and more manageable format. One could even imagine defining mixtures of CDs for specific uses by their percentages in an actual sample and concatenating the names together either for clear labeling and/or more robust computational modeling of mixtures.

4.- Structured data file format

Although the defined string satisfies the requirements of an ideal nomenclature, it remains difficult to generate manually. To simplify this process, a structured, human-readable data file in a YAML-like format[38] was developed to define the CD structure. This file can be easily parsed by a Python script to automatically generate the full string. This structured format not only aids in generating the string but is also designed for computational analysis, providing clear and easily parsed sections for both human and software interpretation. The data file for a CD dimer has the following format:

```
CD1: [Modification Details for First Monomer]
CD2: [Modification Details for Second Monomer (if applicable)]
Linker_name: [Name(s) of Linker(s)]
```

```
ini_res/pos: [Initial Residue/Position for Each Linker]
end_res/pos: [End Residue/Position for Each Linker]
```

4.1.- Encoding CD sequences

The sequence of each CD monomer is detailed on a separate line, with modifications listed as space-separated entries for each residue:

```
[Type][Position] [Type][Position] [Type][Position]…
```

where `Type` represents again the type of modification and position the number and location of the substitutions in the corresponding GPU.

```
CD1: SBE6 SBE6 SBE6 SBE6 SBE6 SBE6 SBE6
```

represents a β-CD with SBE modifications at position 6 on all seven residues. The extension for hybrid modifications is trivial:

```
CD1: SBE6 HP8 SBE2 ME1 HP5 ME9
```

while hybrid modifications within the same residue can also easily be specified:

```
CD1: SBE6 SBE3_HP8 SBE2_HP6 ME1 HP5 ME9
```

which can be translated in the following CyDexID-E code:

```
HP_5_6x086050_ME_5_6x000109_SBE_3_6x632000
```

which collapses to:

```
HP5ME5SBE3_6u
```

when using the CyDexID-S nomenclature.

4.2.- Encoding linker information

Linkers are only present in dimers and higher order oligomers. As in the case of substitution types, different chemical groups can be employed to covalently join multiple CD subunits. As explained above, the chemical group and the position to which it is bound to both CD monomers must be described. Additionally, it is possible to have several linkers joining the same subunits. This information is supplied by the `Linker_name`, the `ini_res/pos` and the `end_res/pos` fields of the file.

Example:

```
Linker_name: BUT AMD HEP
ini_res/pos: 1/2 4/3 6/2
end_res/pos: 1/2 2/3 6/6
```

where we would have three different linkers: a butyl (`BUT`) linker connecting residue 1, position 2 of CD1 to residue 1, position 2 of CD2; an amide (`AMD`) linker connecting residue 4, position 3 of CD1 to residue 2, position 3 of CD2; and a heptyl (`HEP`) linker connecting residue 6, position 2 of CD1 to residue 6, position 6 of CD2. Writing/reading this structure file as well as parsing it to the string-based nomenclature, and vice versa, is straightforward.

Let's see a complete example that could even be interesting from the practical point of view (Fig. 9):

File Format Representation:

```
CD1: SBE6 HP6 SBE6 HP6 SBE6 HP6 SBE6 HP6
CD2: SBE6 HP6 SBE6 HP6 SBE6 HP6 SBE6 HP6
Linker_name: BUT BUT BUT BUT
ini_res/pos: 1/3 3/2 5/3 7/2
end_res/pos: 6/2 4/3 2/2 8/3
```

which can be mapped to the following CyDexID-E string:

```
HP_4_8x06060606_SBE_4_8x60606060
_13_32_53_72_BUT_BUT_BUT_BUT_62_43_22_83_
HP_4_8x06060606_SBE_4_8x60606060
```

and to the more compact CyDexID-S string:

```
HP4SBE4_8u_BUT4_HP4SBE4_8u
```

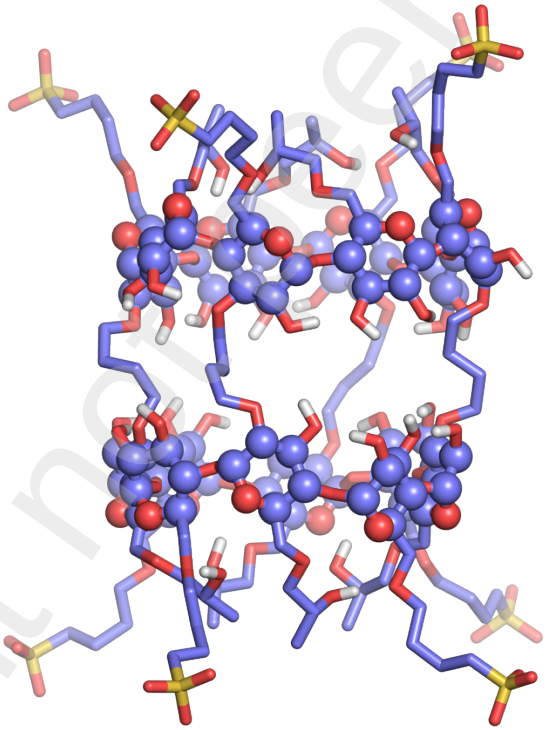

Figure 9.- Structure corresponding to `HP_4_8x06060606_SBE_4_8x60606060_13_32_53_72_BUT_BUT_BUT_BUT_62_43_22_83_HP_4_8x06060606_SBE_4_8x60606060` in the CyDexID-E format, which can be projected to the CyDexID-S notation as `HP4SBE4_8u_BUT4_HP4SBE4_8u`.

The structure shown in Figure 9 is elaborate, as it includes four linkers between the two subunits and eight substitutions of two different types (SBE and HP) on each monomer. Both the structure file and the CyDexID-E code describe the structure unambiguously, including the specific connections between the CD subunits for each linker. This structure could potentially be useful as a host for a variety of molecules. Notably, the CyDexID-S code for this molecule applies to many structures with the same components but different linker and substitution distributions.

Changes, such as placing SBE or HP substitutions at different positions or altering the symmetry of sequences, would significantly affect the molecule's properties, including cavity dynamics and encapsulation specificity. This highlights the critical importance of specifying the precise locations of all substitutions and linkers for functional accuracy.

5.- Functional groups

One of the challenges in implementing the proposed nomenclature is the need for a concise and standardized code to represent functional groups. A unique three-letter code for each functional group would greatly enhance the clarity and usability of the nomenclature, allowing for the systematic, universal, and compact description of modified CDs. This is particularly important given the diversity of functional groups that can be introduced in CDs, as well as the need to maintain consistency and brevity in both experimental and computational contexts. While existing molecular notations, such as SMILES and SMARTS, provide robust frameworks for describing entire molecular structures or querying substructures, they do not offer standalone, standardized representations for functional groups. A three-letter coding system would streamline classification and enhance data interoperability across software, databases, and theoretical chemistry workflows.

A review of the literature highlights active efforts in computational functional group recognition and annotation using rule-based systems, hierarchical classifications, and semantic annotations (e.g., Toxtree, CheckMol, FGO, OWL-DL) [39–41] [42,43]. However, these approaches lack standardized alphanumeric codes. Tools like BiSSCat [44] [45,46] and OCHEM [39] provide useful frameworks for specific contexts, but they rely on database-specific or numerical identifiers rather than interoperable codes. This fragmentation underscores a significant opportunity to develop standardized alphanumeric coding systems, tailored to facilitate interoperability and streamline functional group applications across computational chemistry, cheminformatics tools, and broader chemical databases. We propose a preliminary set of two and three-letter abbreviations to encode commonly used functional groups in modified CDs (Table 1). The table below lists the functional groups along with their proposed abbreviations:

Table 1.- List of representative functional groups along with their proposed abbreviations.

Substitution	Abbreviation
2-Hydroxypropyl	HP
4-Sulfobutyl	SBE
Carboxythioether	CTE
Carboxymethyl	CME
Quaternary Ammonium	QA
Acetyl	AC
Tosyl	TOS
Succinyl	SUC
Benzoyl	BZ
Methyl	ME
Ethyl	ETH
1-Propyl	PRP
1-Butyl	BUT

These codes serve as a foundation for further refinement and standardization, addressing the gap in the literature and enabling seamless integration into both experimental workflows and computational tools.

6.- Conclusions

There is much evidence indicating that the properties of modified CD molecules depend on the number of GPUs, as well as on the type, number, and even the specific location of substituted groups[47]. However, standard nomenclature systems, which only account for the number of GPUs, the type of substitution, and the average number of substitutions, can represent thousands or even millions of different structures. This is particularly limiting when designing and optimizing new CD applications. This work provides detailed calculations and examples showcasing the variety of structures that can arise in a sample with a given DS, including cases where the same number and types of substitutions occur at different locations. Because these details may be more or less relevant depending on the context, two forms of notation are proposed: an expanded version (CyDexID-E), which provides detailed information and can unambiguously describe any CD structure, and a compact version (CyDexID-S), which is useful in contexts where the exact position of substitutions is either unknown or unnecessary. This dual approach allows researchers to choose the level of detail appropriate for their specific application, ensuring both accuracy and simplicity where needed. CyDexID-E codes can be paired with structured data files for enhanced clarity and computational workflows, enabling seamless representation and scalability for diverse CD modifications, including stereochemical and polymerizable groups.

Several examples of CD structures with increasing complexity are provided to demonstrate the method's ability to describe them unambiguously. Some examples intentionally push the boundaries of structural complexity—while remaining within the bounds of chemical laws—to rigorously test and challenge the proposed nomenclature system. Furthermore, to maximize standardization, a list of three-letter abbreviations for functional groups typically employed in modified CDs is proposed. In the present work, our discussion is limited to native glucose-based CD units. Thus, 3,6-monoanhydro-CD or per-anhydro forms are not considered within the CD scaffold. Furthermore, the nomenclature proposed here only considers structures where all units are glucose-based. The current proposal should be extended in order to explicitly consider these groups. A similar strategy used to label substitutions could be employed for this aim.

The expanded version of the proposed nomenclature system for CD derivatives satisfies all the essential requirements aimed at solving the problems of current nomenclature systems: uniqueness, completeness, comprehensiveness, scalability, computational efficiency, human readability, and alignment with established protein nomenclature systems. A key strength of this nomenclature system is its flexibility, enabling it to accommodate any kind of CD modifications, from simple single substitutions in CD monomers to complex hybrid structures and covalently linked oligomers with any number and location of the linkers between subunits. Additionally, the proposed nomenclature considers stereochemical modifications on any of the chiral centers of the glucose units. Functional groups that can polymerize are denoted by a suffix indicating the polymerization degree. Its scalability ensures that it can adapt to the growing diversity of CD-based applications, whether in drug delivery, specific molecular recognition, catalysis, or supramolecular chemistry.

The proposed system is highly suitable for computational applications, offering a format that is easy to parse and manipulate. It enables automatic generation of CD structures, parameterization for molecular docking and dynamics simulations, and the creation of databases to support machine learning models for predicting CD properties. The structured data file format complements the string-based nomenclature, streamlining workflows and facilitating efficient analysis of complex CD structures.

One of the standout features of this system is its parallelism with protein nomenclature, allowing the use of familiar terms such as residues, mutations, and sequences. Moreover, it facilitates the classification of structural hierarchies, from primary sequences to secondary intramolecular patterns, tertiary oligomerization through linkers, and even quaternary non-covalent aggregates.

Just as the universe of proteins is described by the term proteome[48], that of lipids by lipidome[49], and that of saccharides by the glycome[50], the entire collection of cyclodextrins (as a unique subset of oligosaccharides within the glycome) can be distinguished as the cyclodextrinome, with each compound unambiguously identified by its CyDexID-E.

To encourage wide adoption of this nomenclature, we have developed an interactive web-based tool named CyDexID Generator, which automatically assigns standardized names to modified CD monomers. The tool is accessible at: https://cydexid-app.lm.r.appspot.com. This tool can also convert between SMILES, IUPAC, and CyDexID E/S nomenclatures, and even produce corresponding 3D structures. We hope the clear advantages of our proposal presented in this work and the availability of an automated tool contribute to rapid and widespread adoption of our proposed nomenclature. Additionally, we commit to using this nomenclature in our future publications, which will help to demonstrate its utility and facilitate its broader acceptance. Finally, although we believe our proposal is solid, we cannot disregard future optimizations. Thus, in order to ensure future adaptability, a version identifier could be included as a prefix at the start of the CyDex-ID string, allowing for backward compatibility and seamless integration of potential extensions without disrupting existing representations.

In summary, the proposed nomenclature system addresses key limitations in the current landscape of CD research while opening new opportunities for both experimental and computational studies. Its ability to facilitate communication, reproducibility, and computational efficiency, as well as its integration with AI-driven drug design and other predictive modeling techniques, opens exciting possibilities for future research. By providing a more systematic and standardized approach, this system is expected to accelerate the discovery of novel CD-based compounds with optimized properties for various applications. Despite the potential challenges in gaining acceptance, we believe its long-term impact will lead to significant advancements in the rational design and optimization of CDs for diverse applications.

7. Acknowledgments

This work was supported by the European Union's Horizon Europe Research and Innovation Programme (Marie Sklodowska-Curie grant agreement Bicyclos N 101130235), the Spanish Agencia Estatal de Investigación (AEI) and the ERDF (PID2019-111327GB-I00, PDC2022-133402-I00, PID2022-141534OB-I00 and CNS2023-144353), by Xunta de Galicia (ED431C 2021/21 and Centro de investigación do Sistema universitario de Galicia accreditation 2023-2027, ED431G 2023/03) and the European Union (European Regional Development Fund – ERDF). We acknowledge CESGA for providing computational support.

During the preparation of this work the author(s) used ChatGPT 4 in order to improve readability and flow of the language. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

References

[1] A. Villiers, Sur la fermentation de la fécule par l'action du ferment butyrique, Compte Rendus Des Séances de l'Académie Des Sciences (France) CXII (1891) 536–538.

[2] T. Loftsson, M.E. Brewster, Pharmaceutical Applications of Cyclodextrins. 1. Drug Solubilization and Stabilization, J Pharm Sci 85 (1996) 1017–1025. https://doi.org/10.1021/JS950534B.

[3] P.C. Manor, W. Saenger, Topography of Cyclodextrin Inclusion Complexes. III. Crystal and Molecular Structure of Cyclohexaamylose Hexahydrate, the (H2O)2 Inclusion Complex, J Am Chem Soc 96 (1974) 3630–3639. https://doi.org/10.1021/JA00818A042.

[4] E.M.M. Del Valle, Cyclodextrins and their uses: a review, Process Biochemistry 39 (2004) 1033–1046. https://doi.org/10.1016/S0032-9592(03)00258-9.

[5] Fenyvesi, M. Vikmon, L. Szente, Cyclodextrins in Food Technology and Human Nutrition: Benefits and Limitations, Crit Rev Food Sci Nutr 56 (2016) 1981–2004. https://doi.org/10.1080/10408398.2013.809513.

[6] G. Crini, Review: A history of cyclodextrins, Chem Rev 114 (2014) 10940–10975. https://doi.org/10.1021/CR500081P.

[7] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S.A.A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold, Nature 2021 596:7873 596 (2021) 583–589. https://doi.org/10.1038/s41586-021-03819-2.

[8] E.J. Stollar, D.P. Smith, Uncovering protein structure, Essays Biochem 64 (2020) 649–680. https://doi.org/10.1042/EBC20190042.

[9] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The Protein Data Bank, Nucleic Acids Res 28 (2000) 235–242. https://doi.org/10.1093/NAR/28.1.235.

[10] J. Szejtli, Introduction and general overview of cyclodextrin chemistry, Chem Rev 98 (1998) 1743–1753. https://doi.org/10.1021/CR970022C.

[11] Á. Piñeiro, J. Pipkin, V. Antle, R. Garcia-Fandino, Aggregation versus inclusion complexes to solubilize drugs with cyclodextrins. A case study using sulphobutylether-β-cyclodextrins and remdesivir, J Mol Liq 343 (2021) 117588. https://doi.org/10.1016/J.MOLLIQ.2021.117588.

[12] M.E. Davis, M.E. Brewster, Cyclodextrin-based pharmaceutics: past, present and future, Nature Reviews Drug Discovery 2004 3:12 3 (2004) 1023–1035. https://doi.org/10.1038/nrd1576.

[13] T. Hao, Burnside's Lemma and Its Applications in Combinatorics Problems, Highlights in Science, Engineering and Technology 47 (2023) 126–130. https://doi.org/10.54097/HSET.V47I.8175.

[14] W. Burnside, Title: Theory of Groups of Finite Order, (2012). https://www.gutenberg.org/ebooks/40395 (accessed December 4, 2024).

[15] T. Kraus, M. Buděšínský, J. Závada, General approach to the synthesis of persubstituted hydrophilic and amphiphilic β-cyclodextrin derivatives, Journal of Organic Chemistry 66 (2001) 4595–4600. https://doi.org/10.1021/jo010046q.

[16] X.J. Chen, G.L. Yang, X.D. Xu, J.J. Sheng, J. Shen, H.X. Dong, Preparation and chromatographic evaluation of β-cyclodextrin derivative CSPs bearing substituted phenylcarbamate groups for HPLC, J Liq Chromatogr Relat Technol 39 (2016) 647–657. https://doi.org/10.1080/10826076.2016.1227993.

[17] B. Wang, E. Zaborova, S. Guieu, M. Petrillo, M. Guitet, Y. Blériot, M. Ménand, Y. Zhang, M. Sollogoub, Site-selective hexa-hetero-functionalization of α-cyclodextrin an archetypical C6-symmetric concave cycle, Nat Commun 5 (2014). https://doi.org/10.1038/ncomms6354.

[18] Z. Fülöp, S. Kurkov, T. Nielsen, K. Larsen, T. Loftsson, Self-assembly of cyclodextrins: formation of cyclodextrin polymer based nanoparticles, 22 (n.d.) 2012.
[19] A.M. Anderson, I. Manet, M. Malanga, D.M. Clemens, K. Sadrerafi, Á. Piñeiro, R. García-Fandiño, M.S. O'Connor, Addressing the complexities in measuring cyclodextrin-sterol binding constants: A multidimensional study, Carbohydr Polym 323 (2024) 121360. https://doi.org/10.1016/J.CARBPOL.2023.121360.
[20] P.F. Garrido, M. Calvelo, A. Blanco-González, U. Veleiro, F. Suárez, D. Conde, A. Cabezón, Á. Piñeiro, R. Garcia-Fandino, The Lord of the NanoRings: Cyclodextrins and the battle against SARS-CoV-2, Int J Pharm 588 (2020) 119689. https://doi.org/10.1016/J.IJPHARM.2020.119689.
[21] P. Saokham, A. Sá Couto, A. Ryzhakov, T. Loftsson, The self-assemble of natural cyclodextrins in aqueous solutions: Application of miniature permeation studies for critical aggregation concentration (cac) determinations, Int J Pharm 505 (2016) 187–193. https://doi.org/10.1016/j.ijpharm.2016.03.049.
[22] A. Anderson, Á. Piñeiro, R. García-Fandiño, M.S. O'Connor, Cyclodextrins: Establishing building blocks for AI-driven drug design by determining affinity constants in silico, Comput Struct Biotechnol J 23 (2024) 1117–1128. https://doi.org/10.1016/J.CSBJ.2024.02.011.
[23] D.O. Thompson, Cyclodextrins-Enabling Excipients: Their Present and Future Use in Pharmaceuticals, Critical Reviews in Therapeutic Drug Carrier Systems 14 (1997) 1-104. https://doi.org/10.1615/CritRevTherDrugCarrierSyst.v14.i1.10.
[24] H.A. Favre, W.H. Powell, Nomenclature of Organic Chemistry: IUPAC Recommendations and Preferred Names 2013, Royal Society of Chemistry, 2014.
[25] C.J. Easton, S.F. Lincoln, Modified Cyclodextrins: Scaffolds And Templates For Supramolecular Chemistry, (1999). https://books.google.es/books?id=cM42DwAAQBAJ&printsec=frontcover&hl=es#v=onepage&q&f=false (accessed October 25, 2024).
[26] N.L. Daly, D.T. Wilson, Plant derived cyclic peptides, Biochem Soc Trans 49 (2021) 1279–1285. https://doi.org/10.1042/BST20200881.
[27] A. Anderson, R. García-Fandiño, Á. Piñeiro, M.S. O'Connor, Unraveling the molecular dynamics of sugammadex-rocuronium complexation: A blueprint for cyclodextrin drug design, Carbohydr Polym 334 (2024) 122018. https://doi.org/10.1016/J.CARBPOL.2024.122018.
[28] O. Trott, A.J. Olson, AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, J Comput Chem 31 (2010) 455–461. https://doi.org/10.1002/jcc.21334.
[29] J. Eberhardt, D. Santos-Martins, A.F. Tillack, S. Forli, AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings, J Chem Inf Model 61 (2021) 3891–3898. https://doi.org/10.1021/acs.jcim.1c00203.
[30] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J.C. Berendsen, GROMACS: Fast, flexible, and free, J Comput Chem 26 (2005) 1701–1718. https://doi.org/10.1002/jcc.20291.
[31] H.J.C. Berendsen, D. van der Spoel, R. van Drunen, GROMACS: A message-passing parallel molecular dynamics implementation, Comput Phys Commun 91 (1995) 43–56. https://doi.org/10.1016/0010-4655(95)00042-E.
[32] S. Páll, A. Zhmurov, P. Bauer, M. Abraham, M. Lundborg, A. Gray, B. Hess, E. Lindahl, Heterogeneous parallelization and acceleration of molecular dynamics simulations in GROMACS, Journal of Chemical Physics 153 (2020). https://doi.org/10.1063/5.0018516.
[33] E. Lindahl, B. Hess, D. van der Spoel, GROMACS 3.0: A package for molecular simulation and trajectory analysis, J Mol Model 7 (2001) 306–317. https://doi.org/10.1007/S008940100045.
[34] B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation, J Chem Theory Comput 4 (2008) 435–447. https://doi.org/10.1021/ct700301q.

[35] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M.R. Shirts, J.C. Smith, P.M. Kasson, D. Van Der Spoel, B. Hess, E. Lindahl, GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit, Bioinformatics 29 (2013) 845–854. https://doi.org/10.1093/bioinformatics/btt055.

[36] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, E. Lindah, Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, SoftwareX 1–2 (2015) 19–25. https://doi.org/10.1016/j.softx.2015.06.001.

[37] D.A. Case, H.M. Aktulga, K. Belfon, D.S. Cerutti, G.A. Cisneros, V.W.D. Cruzeiro, N. Forouzesh, T.J. Giese, A.W. Götz, H. Gohlke, S. Izadi, K. Kasavajhala, M.C. Kaymak, E. King, T. Kurtzman, T.S. Lee, P. Li, J. Liu, T. Luchko, R. Luo, M. Manathunga, M.R. Machado, H.M. Nguyen, K.A. O'Hearn, A. V. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, A. Risheh, S. Schott-Verdugo, A. Shajan, J. Swails, J. Wang, H. Wei, X. Wu, Y. Wu, S. Zhang, S. Zhao, Q. Zhu, T.E. Cheatham, D.R. Roe, A. Roitberg, C. Simmerling, D.M. York, M.C. Nagan, K.M. Merz, AmberTools, J Chem Inf Model 63 (2023) 6183–6191. https://doi.org/10.1021/acs.jcim.3c01153.

[38] Yet Another Markup Language (YAML) 1.0, (n.d.). https://yaml.org/spec/history/2001-12-10.html (accessed October 25, 2024).

[39] E.S. Salmina, N. Haider, I. V. Tetko, Extended functional groups (EFG): An efficient set for chemical characterization and structure-activity relationship studies of chemical compounds, Molecules 21 (2016). https://doi.org/10.3390/molecules21010001.

[40] R. Benigni, O. Tcheremenskaia, A. Worth, Computational Characterisation of Chemicals and Datasets in Terms of Organic Functional Groups-a New Toxtree Rulebase, (n.d.). https://doi.org/10.2788/33281.

[41] I. Sushko, E. Salmina, V.A. Potemkin, G. Poda, I. V. Tetko, ToxAlerts: A web server of structural alerts for toxic chemicals and compounds with potential adverse reactions, J Chem Inf Model 52 (2012) 2310–2316. https://doi.org/10.1021/ci300245q.

[42] N. Villanueva-Rosales, M. Dumontier, Describing chemical functional groups in OWL-DL for the classification of chemical compounds, in: CEUR Workshop Proc, 2007. http://ontology.dumontierlab.com/cfg-owled-2007.

[43] P. Kumar Varadwaj, T. Lahiri, Hypothesis FGO: A novel ontology for identification of ligand functional group, Bioinformation 2 (2007) 113–118. www.bioinformation.net.

[44] M. Kotera, A.G. McDonald, S. Boyce, K.F. Tipton, Functional group and substructure searching as a tool in metabolomics, PLoS One 3 (2008). https://doi.org/10.1371/journal.pone.0001537.

[45] M. Korichi, V. Gerbaud, P. Floquet, A.H. Meniai, S. Nacef, X. Joulia, Computer aided aroma design I–Molecular knowledge framework, Chemical Engineering and Processing: Process Intensification 47 (2008) 1902–1911. https://doi.org/10.1016/J.CEP.2008.02.008.

[46] D. Qv, J. Su, M. Muraki, T. Hay Aka Wa, A Decoding System for a Group Contribution Method, 1992. https://pubs.acs.org/sharingguidelines.

[47] G. Kali, S. Haddadzadegan, A. Bernkop-Schnürch, Cyclodextrins and derivatives in drug delivery: New developments, relevant clinical trials, and advanced products, Carbohydr Polym 324 (2024). https://doi.org/10.1016/j.carbpol.2023.121500.

[48] M.R. Wilkins, J.C. Sanchez, A.A. Gooley, R.D. Appel, I. Humphery-Smith, D.F. Hochstrasser, K.L. Williams, Progress with proteome projects: Why all proteins expressed by a genome should be identified and how to do it, Biotechnol Genet Eng Rev 13 (1996) 19–50. https://doi.org/10.1080/02648725.1996.10647923.

[49] M.R. Wenk, The emerging field of lipidomics, Nature Reviews Drug Discovery 2005 4:7 4 (2005) 594–610. https://doi.org/10.1038/nrd1776.

[50] A. Varki, R.D. Cummings, J.D. Esko, P. Stanley, G.W. Hart, M. Aebi, D. Mohnen, T. Kinoshita, N.H. Packer, J.H. Prestegard, R.L. Schnaar, P.H. Seeberger, Essentials of Glycobiology, Cold Spring Harbor (NY) (2022) 892. https://doi.org/10.1101/9781621824213.