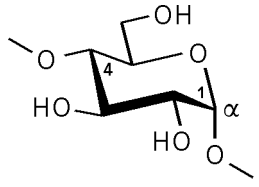
7

Some Reflections on the Multitude of Cyclodextrin Isomers : Astronomic Numbers as a Justification for Computational Studies prior to Synthesis

Abstract: For the cyclodextrins consisting out of 5 – 8 sugar residues, a quick approximation, as well as an exact mathematical procedure for calculating the number of isomers emerging from either chemical modification of hydroxyl groups, or exchange of sugar units is presented. The astronomic numbers of different possibilities found for minor changes only, point towards the relevance of automated computational studies for selecting new compounds, which – in all probability – are expected to exhibit new features that would justify the immense effort of their chemical synthesis.

The native cyclodextrins (CD's) obtained from degradation of starch are small cyclooligosaccharides consisting out of six (**2**, α-CD), seven (**3**, β-CD) or eight (**4**, γ-CD) α(1→4)-linked glucose residues[305-308]. Albeit the fact, that on the high-molecular weight side the cyclodextrin series extends to at least the cycloglucododecaoside (12 sugar units)[440,441], no homologs of lower weight with fewer than six sugar residues could be isolated from natural sources[440], yet the cycloglucopentaoside **1** became available via chemical synthesis[442].



	1	*cyclo*[D-Glc*p*α(1→4)]$_5$
α-CD	**2**	*cyclo*[D-Glc*p*α(1→4)]$_6$
β-CD	**3**	*cyclo*[D-Glc*p*α(1→4)]$_7$
γ-CD	**4**	*cyclo*[D-Glc*p*α(1→4)]$_8$

Due to their unusual loop structures, these compounds are able to incorporate small guest molecules in their center hydrophobic cavities[305-308], and thus have received considerable attention studying inclusion phenomena[374]. The multifarious industrial application profiles of cyclodextrins and their derivatives[312], and the recently increasing number of synthetic approaches towards cyclodextrin derivatives and isomers[427] as, for example, the *manno*-analogs ("cyclomannins")[443-446], necessitates comparative molecular modeling studies to gain a better understanding of the physico-chemical properties of this class of compounds. In the sequel, some reflections on the number of theoretically possible isomers of the cyclodextrins **1** – **4**, emerging

from chemical modification of the hydroxyl groups and / or exchange of individual sugar units for alternative residues will be presented.

Cyclodextrin Isomers from Chemical Modification of Hydroxyl Groups

Due to their intrinsic rotational symmetry properties, the number of theoretically possible isomers is much more difficult to calculate for loop structures than for acyclic compounds. In this context, it seemed appropriate to develop a model by which the number of cyclodextrin isomers becomes computable from the number of acyclic precursors, which are then cyclized in a fictive train of thoughts in a later step.

In cyclodextrins, each of the α(1→4)-linked glucose units possesses three free hydroxyl groups 2-OH, 3-OH, and 6-OH that could be affected by chemical modification, i.e. either by protection or deprotection, and thus, can exhibit $2^3 = 8$ different substitution patterns: one each with all OH groups free or protected, and three with only one hydroxyl modified or unaltered, respectively*. In consequence, for each parent open-chained oligosaccharides consisting out of *n* equal sugar residues with a total of 3*n* hydroxyl groups (excluding the anomeric center and the terminal hydroxyl group used for cyclization) there are $8^n = 2^{3n}$ alternative substitution patterns. However, the starting and the end-point of the acyclic analogs are lost in the cyclic structures: in the case of α-CD (**2**) for example, six different linear hexasaccharides [ABCDEF], [BCDEFA], ... [FABCDE], with each letter denoting a pyranose unit with identical configuration (here [D-Glc*p*α(1→4)]), but an individual *O*-substitution pattern*, yield the very same cyclodextrin isomer (Fig. 7-1). Thus, in a first approximation the total number of cyclodextrin isomers N_{tot} diminishes to the n^{th}-fraction of all acyclic derivatives only (cf. equation 7-1).

$$N_{tot} \approx \frac{1}{n}\left[2^{3n}\right] \qquad (7\text{-}1)$$

---

* Already here an important statistical rule becomes evident: for calculation of the number of isomers, it is immaterial whether modification of the initially free OH groups, or *vice versa*, deprotection of modified hydroxyls is considered; only the relative numbers of hydroxyl groups in each class of chemical state that can be distinguished are relevant.

* Throughout this chapter, the abbreviations [XY ... Z] were used for acyclic oligosaccharides, with each letter denoting one sugar residue with its (unless further specified) individual stereochemistry, type of intersaccharidic linkage and / or substitution pattern; different units are represented by alternative letters. The prefix "*cyclo*" is used to indicate loop structures (e.g. *cyclo*[XY ... Z], or *c*[XY ... Z] for short).

For calculation of the number of isomers as a function of the degree of substitution σ, i.e. $N_{tot}(\sigma)$, where σ denotes the number of modified hydroxyl groups per cyclodextrin molecule, the term 2^{3n} – which accounts for all isomers in Eq. 7-1 – has to be replaced to count only those having σ out of 3*n* OH-groups modified*, as represented by the expression $\binom{3n}{\sigma}$ in Eq. 7-2.

$$N_{tot}(\sigma) \approx \frac{1}{n}\left[\binom{3n}{\sigma}\right] = \frac{1}{n}\left[\frac{(3n)!}{(3n-\sigma)!\cdot(\sigma)!}\right] \qquad 0 \le \sigma \le 3n \qquad (7\text{-}2)$$

However, in a strict sense, the equations 7-1 and 7-2 only yield a first estimation for the number of alternative isomers, since they do not properly consider all symmetry properties of loop structures. As exemplified for cyclohexaoside-type structures in Fig. 7-1, the isomers of higher symmetry (2-, 3- or higher-fold rotation axis C_2, C_3, ... C_i) can be traced back to fewer acyclic precursors than the asymmetrically substituted ones, i.e. those, having C_1-symmetry only. α-Cyclodextrin isomers exhibiting C_2-symmetry properties can be constructed starting from three parent acyclic oligosaccharides (second left entry in Fig. 7-1). Moreover, cyclohexaosides of C_3- and C_6-symmetry can be build up from two resp. one open-chained precursors only (right two columns in Fig. 7-1).

In general terms, each cyclodextrin analog of C_i-symmetry, containing *n* sugar units can be traced back to n/i alternative precursors – in the case of α-CD-type structures six (= 6/1) for C_1-compounds, three (= 6/2) for C_2, two (= 6/3) for C_3, and one (= 6/6) for C_6, whereas C_4- and C_5-symmetry is *a priori* impossible. This rule is not only valid for α-CD, but applies in the same fashion to its homologs of lower and higher ring size. For derivatives of **1** and **3**, only C_1 and C_i (i = 5, 7) properties are relevant, but in the case of γ-CD C_1-, C_2-, C_4-, and C_8-symmetry has to be considered.

The mathematical formulation of this procedure is given in eq. 7-3: after classifying all linear saccharides according to the symmetry properties that would result upon cyclization, the total number of cyclodextrin isomers (N_{tot}) carrying an arbitrary number of substituents amounts to the sum over all structures per class (≡ $N(C_i)$), each weighted by the factor n/i. In a similar way, eq. 7-4 counts only those compounds that

---

* Selecting N_2 elements (i.e. hydroxyl groups) out of a total of N_1 different ones for modification, the total number N_{tot} of alternative possibilities amounts to

$$N_{tot} = \binom{N_1}{N_2} = \frac{(N_1)!}{(N_1-N_2)!\cdot(N_2)!} = \frac{N_1\cdot(N_1-1)\cdot\ldots\cdot(N_1-N_2+1)}{N_2\cdot(N_2-1)\cdot\ldots\cdot 1}$$

according to the rules of statistics. The number of isomers with N_2 OH-groups free or protected (i.e. N_1 - N_2 free OH's) is identical, since $N_{tot}(N_1, N_2)$ is equal to $N_{tot}(N_1, N_1 - N_2)$.

have a fixed number of substituents (σ). For simplicity, the by far largest portions of precursors for asymmetric (C_1) cyclodextrin structures, i.e. the terms $N(C_1)$ and $N(\sigma,C_1)$ in eqs. 7-3 and 7-4, are most conveniently replaced against the total number of all possibilities (cf. eqs. 7-1 and 7-2), diminished by the total amount of oligosaccharides that would lead to cyclodextrins of higher symmetry (i.e. $N(C_2)$, $N(C_3)$, up to $N(C_i)$).

$$N_{tot} = \sum_{i=1,2,\ldots,n} \frac{N(C_i)}{n/i} = \frac{1}{n}\left[2^{3n} - \sum_{i=2,3,\ldots,n} N(C_i)\right] + \sum_{i=2,3,\ldots,n} \frac{N(C_i)}{n/i} \qquad (7\text{-}3)$$

$$N_{tot}(\sigma) = \sum_{i=1,2,\ldots,n} \frac{N(\sigma,C_i)}{n/i} = \frac{1}{n}\left[\binom{3n}{\sigma} - \sum_{i=2,3,\ldots,n} N(\sigma,C_i)\right] + \sum_{i=2,3,\ldots,n} \frac{N(\sigma,C_i)}{n/i} \qquad (7\text{-}4)$$

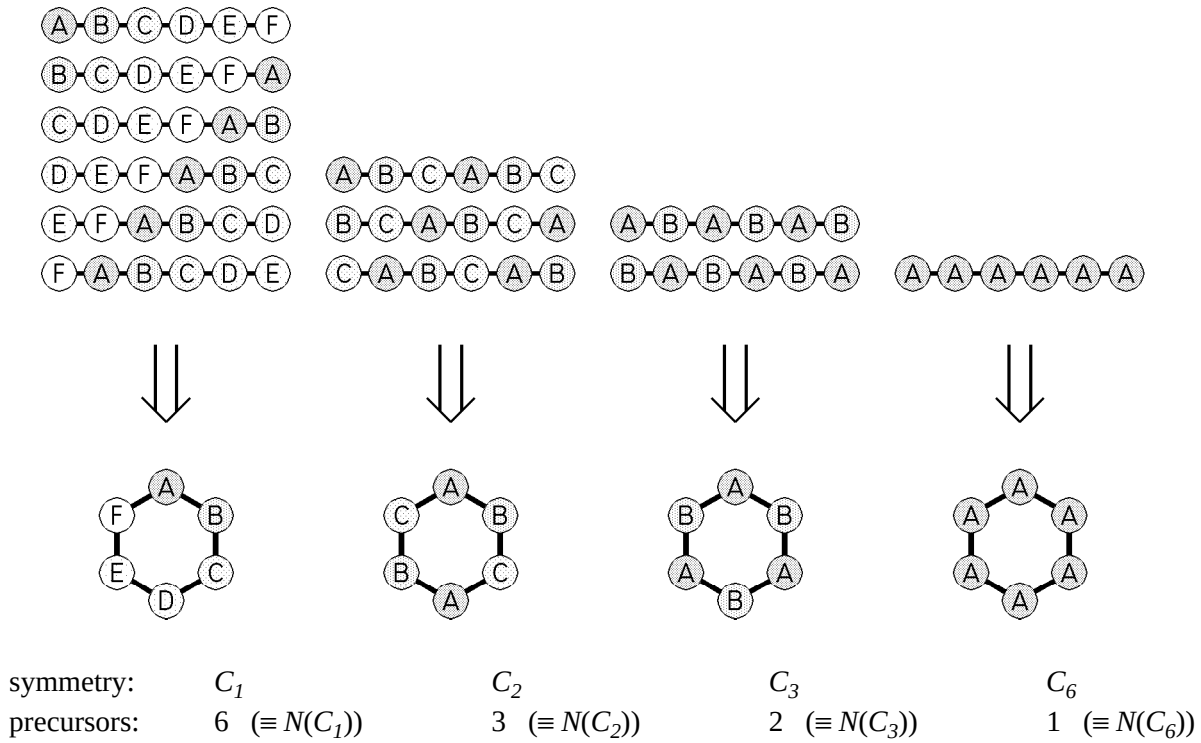


Fig. 7-1. Construction scheme of cyclohexaoside-type (**2**, α-CD) structures starting from acyclic oligosaccharides as precursors: different letters denote sugar units with different substitution patterns, but equal stereochemistry; all intersaccharidic linkages symbolized by solid lines between the circles are considered to be of the same type. Asymmetric C_1-structures derived from α-CD can be build up starting from six different open-chain precursors, for those of higher C_2-, C_3-, and C_6-symmetry (2-, 3-, and 6-fold rotation axis) only three, two, and one parent compound(s) are theoretically possible, respectively. Cyclic structures of the type *cyclo*[AAAABC] with only three different sugar residues, for example, or those exhibiting a mirror plane only instead of rotational symmetry (e.g. *cyclo*[AAABBB] or *cyclo*[ABCCBA]), all belong to the first group with C_1-properties (*left* entry), and thus, can be traced back to six alternative precursors.

The number of isomers as a function of the degree of substitution is listed in Table 7-1 for the cyclodextrins **1** – **4**, the values were all calculated on the basis of eqs. 7-3 and 7-4*, and were counterchecked by straightforward counting and classifying of all combinations of substitution patterns by using a computer program.

Table 7-1. Number of isomers of the cyclodextrins **1** – **4** as a function of the degree of substitution, i.e. the number of modified hydroxyl groups σ.

substituents	cyclodextrin				(continued)	cyclodextrin			
σ	(**1**)	α-CD (**2**)	β-CD (**3**)	γ-CD (**4**)	σ	(**1**)	α-CD (**2**)	β-CD (**3**)	γ-CD (**4**)
0	1	1	1	1	13	21	1,428	29,070	312,018
1	3	3	3	3	14	3	516	16,614	245,256
2	21	27	30	36	15	1	138	7,752	163,438
3	91	138	190	253	16		27	2,907	92,001
4	273	516	855	1,338	17		3	855	43,263
5	603	1,428	2,907	5,313	18		1	190	16,852
6	1,001	3,114	7,752	16,852	19			30	5,313
7	1,287	5,304	16,614	43,263	20			3	1,338
8	1,287	7,314	29,070	92,001	21			1	253
9	1,001	8,110	41,990	163,438	22				36
10	603	7,314	50,388	245,256	23				3
11	273	5,304	50,388	312,018	24				1
12	91	3,114	41,990	338,140					
total:	N_{tot}					**6,560**	**43,800**	**299,600**	**2,097,684**

In a more conspicuous way, a semi-logarithmic plot of the calculated numbers versus the degree of substitution (σ) is provided by Fig. 7-2. The bell-shaped symmetrical curves reach a solitary maximum for half of the accessible OH-groups being modified. On this scale, the respective maxima found for the cyclodextrins **1** – **4** ($n = 5 - 8$ glucose residues per molecule) are correlated with each other by an approximately straight line (cf. Fig. 7-2).

* In cases where no symmetric structures can be constructed, eq. 7-2 applies exactly, e.g. for β-CD (**3**) carrying 10 substituents ($n = 7$, $\sigma = 10$) eq. 7-2 yields $N_{tot}(\sigma) = 50{,}388$ isomers[308] (cf. Table 7-1). The total number of substituted β-CD isomers is readily calculated from eq. 7-3: only C_7 rotational symmetry has to be considered, for which eight different acyclic precursors (= 8 different substitution patterns for all the glucose units) are theoretically possible (i.e. $N(C_7) = 8$, $n = 7$), thus yielding a total of $N_{tot} = 1/7 \cdot [2^{21} - 8] + 8 = 299{,}600$ isomers for substitution of β-CD. Higher symmetry is more difficult to handle. In α-CD (**2**, $n = 6$) with three substituents, C_6-symmetry cannot occur ($N(C_6) = 0$), but six (= $N(C_3)$) acyclic oligosaccharides yield C_3-symmetrical cyclodextrins during cyclization: two precursors of the type [ABABAB], where each sugar unit A or B carries one substituent, and three isomers for each possibility (each glucose residue carries three OH-groups!). Thus, by application of eq. 7-4 it turns out that 138 different α-cyclodextrin isomers with $\sigma = 3$ can be constructed (cf. Table 7-1), whilst the rough (but much simpler) estimation according to eq. 7-2 yields 136 isomers only.

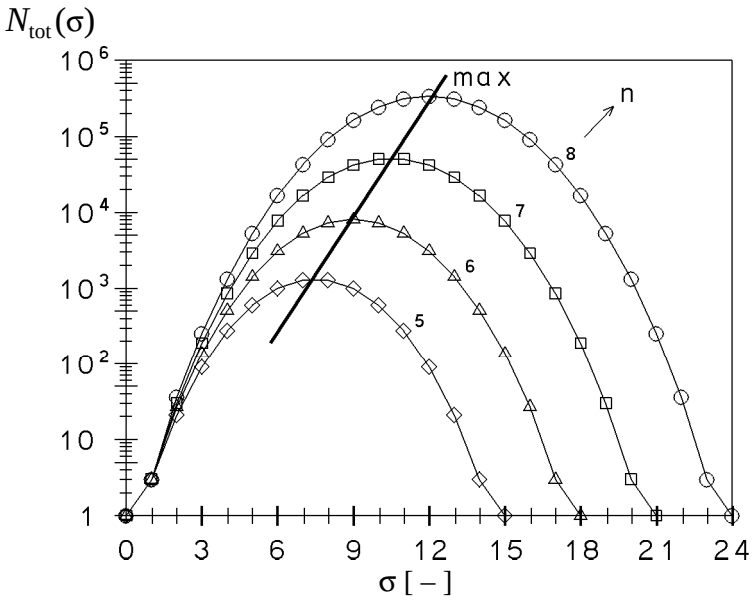


Fig. 7-2. Semi-logarithmic plot of the number of isomers as a function of the degree of substitution σ, as calculated for the cyclodextrins consisting out of five (= *n*, **1**), six (**2**, α-CD), seven (**3**, β-CD), and eight (**4**, γ-CD) sugar residues, respectively. The straight line indicates a least-squares approximation of the individual maxima found for half the hydroxyl groups being modified.

Clearly, every chemical reaction on unsubstituted cyclodextrins must almost instantaneously result in a vast mixture of isomeric compounds. Well-defined products are expected to accumulate in useful yields only if the rate constants for modification of the various hydroxyl groups differ by at least an order of magnitude. However, for most industrial purposes, mixtures of isomers are either satisfactorily or even desirable, the number of isomers contained in such products can be extracted from Table 7-1. In almost all cases, eqs. 7-1 and 7-2 yield sufficiently accurate and convenient quick-estimations for the exact numbers, the latter being invariably larger than their approximations. On the basis of eq. 7-1, the total number of isomers is estimated to ≈ 6,554 for the cyclodextrin **1**, ≈ 43,691 for α-CD (**2**), ≈ 299,593 (**3**, β-CD), and ≈ 2,097,152 (**4**, γ-CD), whilst the exact numbers listed in Table 7-1 are slightly increased by 1 – 2 °/₀₀.

Cyclooligosaccharide Isomers from Exchange of Sugar Units

Cyclodextrin isomers not only result from chemical substitution of hydroxyl groups, but also – at least in principle – from alteration of each of the five stereocenters within every sugar residue, and / or both types of modification simultaneously. Since it neither seems useful to consider all varieties of generating isomers separately, nor would such reflections yield meaningful numbers of general validity, a more global

approach to assess the number of possibilities to generate cyclodextrin isomers was chosen.

If the above restriction of uniform stereochemistry and / or equal types of intersaccharidic linkages is discarded, it turns out to be most convenient to represent the individual sugar moieties by different letters, and abbreviate the cyclodextrins as *cyclo*[XY ... Z] (cf. above). In this context here, alternative letters denote distinguishable residues, irrespective of the reasons for their differences, as for example, their *O*-substitution patterns, inversion of stereocenters, types of linkages, and / or exchange of pyranoid units versus furanoid ones.

For cyclooligosaccharides consisting out of five, six, seven, and eight sugar residues, the various possibilities and the principle number of isomers emerging from these considerations are listed in Table 7-2. The data includes various positional isomers of sugar units relative to each other, the compounds have been arranged in groups according to their sum formulas, respectively. In the case of cyclohexaosides for example, three isomers of the general formula A_4B_2 can be constructed: *cyclo*[AAAABB], *cyclo*[AAABAB], and *cyclo*[AABAAB]. The latter, exhibiting C_2-symmetry, would be identical to *cyclo*[ABAABA] and *cyclo*[BAABAA], and thus, these possibilities were removed from the listing in Table 7-2. In addition, the entry *cyclo*[AAAABB] includes compounds of the type *cyclo*[BBBBAA], since the alternative form does not result in new isomers unless the nature of the residues A and B is further specified. In a similar way, *cyclo*[AABBCC] is equal to compounds such as *cyclo*[AACCBB] and *cyclo*[BBAACC]. The number of isomers increases enormously with increasing ring size of the macrocycle, out of the 143 and 575 different types of cyclohepta- and cyclooctaosides only those of higher rotational symmetry were included in Table 7-2.

Depending on the types of changes allowed for each sugar residue within the macrocycles, the number of alternative possibilities can be calculated from the data of Table 7-2*. Considering for example all different pyranoid sugars (5 stereocenters C_1 to C_5 – including the anomeric center – yield $2^5 = 32$ different units) with four possible types of linkages for each residue (1→2, 1→3, 1→4, and 1→6), results in $32 \cdot 4 = 128$ alternative building blocks as constituent components for cyclooligosaccharides.

---

* The number of cyclodextrin isomers emerging from derivatization of hydroxyl groups (data of Table 7-1) can also be derived from the formula abbreviations contained in Table 7-2, when characterizing the eight different glucose substitution patterns by letters A – H. However, this procedure is more lengthy and difficult than the one presented above, since some classes of isomers again exhibit rotational symmetry properties.

Table 7-2. Principal classes of isomers for cyclooligosaccharides containing 5 – 8 sugar residues listed according to their sum formulas. For cyclopentaosides and cyclohexaosides the different types of isomers are further specified by abbreviated formulas, whilst for heptamers and octamers only those of higher symmetry are listed, superscripts indicate the type of C_i-symmetry ($i \geq 2$), respectively.

sum formula	number of isomers	formulas of structural isomers
5 sugar residues:		
5	1	[ ] ^{5}
4 1	1	[ ]
3 2	2	[ ] [ ]
3 1 1	2	[ ] [ ]
2 2 1	3	[ ] [ ] [ ]
2 1 1 1	2	[ ] [ ]
1 1 1 1 1	1	[ ]
total:	**12**	
6 sugar residues:		
6	1	[ ] ^{6}
5 1	1	[ ]
4 2	3	[ ] [ ] [ ] ^{2}
3 3	3	[ ] [ ] [ ] ^{3}
4 1 1	3	[ ] [ ] [ ]
3 2 1	10	[ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ]
2 2 2	5	[ ] [ ] [ ] [ ] [ ] [ ] ^{2}
3 1 1 1	4	[ ] [ ] [ ] [ ]
2 2 1 1	9	[ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ]
2 1 1 1 1	3	[ ] [ ] [ ]
1 1 1 1 1 1	1	[ ]
total:	**43**	

sum formula	number of isomers	isomers (higher symmetry only)	sum formula	number of isomers	isomers (higher symmetry only)
7 sugar residues:			**8 sugar residues:**		
7	1	[ ] ^{7}	8	1	[ ] ^{8}
6 1	1		7 1	1	
5 2	3		6 2	4	[ ] ^{2}
4 3	5		5 3	7	
5 1 1	3		4 4	7	[ ] [ ] ^{2} ^{4}
4 2 1	15		6 1 1	4	
3 3 1	10		5 2 1	21	
3 2 2	15		4 3 1	35	
4 1 1 1	5		4 2 2	29	[ ] [ ] ^{2} ^{2}
3 2 1	30		3 3 2	37	
2 2 2 1	21		5 1 1 1	7	
3 1 1 1 1	5		4 2 1 1	54	
2 2 1 1	15		3 3 1 1	38	
2 1 1 1 1	4		3 2 1 1 1	105	
1 1 1 1 1 1 1	1		2 2 2 2	23	[ ] ^{2}
total:	**143**		4 1 1 1 1	10	
			3 2 1 1	71	
			2 2 2 1	76	
			3 1 1 1 1	8	
			2 2 1 1	31	
			2 1 1 1 1	5	
			1 1 1 1 1 1	1	
			total:	**575**	

Exchange of only two sugar units in β-CD (sum formula: A_7) for two different ones ($\rightarrow A_5B_2$, three types of isomers cf. Table 7-2), leads to 3·128·127 = 48,768 isomers (128 possibilities for A, and 127 remaining alternatives for B). Introduction of a third type of sugar by exchange of one further unit A against C ($\rightarrow A_4B_2C$, 15 classes of isomers) increases the number of isomers to 30,723,840.

Selecting Relevant Isomers prior to Synthesis

However, when considering not only pyranoid sugar units, but also furanoid ones, as realized for example in the cyclofructins (≡ *cyclo*[D-Fru*f*β(1→2)]$_n$, "cycloinulins")[447,448], the number of isomers increases to astronomic dimensions and finally reaches infinity*. Even when considering only a few relevant exchange patterns out of all possible ones, the number of isomeric compounds is still far beyond what synthetic chemistry could afford.

The quest for gaining new insights and a deeper comprehension of the conformations and complexation properties of novel cyclooligosaccharides, requires a rigorous restriction of synthetic efforts to a few targets only. Their selection inevitably must rely on computational methods rather than intuition, and only automated computational strategies represent an adequate tool to select those isomers, for which – in all probability – basically new features are to be expected. It is also likely that molecular properties will change only gradually when going from *cyclo*[A_n] to *cyclo*[B_n]-type compounds, independently from the pathway of exchanging A → B at the various positions. This even furthermore indicates how restrictions of synthetic efforts to interesting compounds can be made.

Out of all these possibilities, the small ring analogs of the natural cyclodextrins, containing five, four, and three glucose residues, as well as the all-*manno* (≡ *cyclo*[D-Glc*p*α(1→4)]$_n$, "cyclomannins") and all-*galacto* (≡ *cyclo*[D-Gal*p*β(1→4)]$_n$, "cyclogalactins") analogs of **1** and **2** (*n* = 5 and 6 sugar residues, respectively) must be considered as relevant cyclodextrin isomers, the cyclomannins were even already synthesized[443-446].

* In general, the rules outlined here for cyclooligosaccharides, apply for all loop structures.