A simple numerical code for polychlorinated biphenyls and bornanes allowing an unequivocal derivation of the steric structure

Michael Oehme[a,b] and Roland Kallenborn[a,b]
[a]Department of Organic Chemistry, University of Basel, St. Johanns-Ring 19, CH-4056 Basel
[b]Norwegian Institute for Air Research (NILU), P.O. Box 100, N-2007 Kjeller, Norway

1. Introduction

The registration and numbering of compounds belonging to groups with a large number of congeners has always been a subject of discussion. Examples are the different classes of polychlorinated aromatics such as polychlorinated biphenyls (PCB), naphthalenes (PCN), dibenzo-p-dioxins (PCDD) and dibenzofurans (PCDF). Different approaches are in use to assign the congeners of these compound classes. In case of PCB a numbering system is generally accepted which is based on the listing of all structures by increasing numbers of Cl-atoms following the rules of IUPAC concerning the hierarchy of substitution[1)]. All congeners in this list are then numbered from 1 to 209[2)]. The advantage of this system is that a number such as CB 105 is shorter and simpler compared to the structure 2,3,3',4,4'-pentachloro biphenyl. However, such a numbering system does not say anything about the structure of the compound.

We feel that an ideal numbering and naming system for compound classes with many isomers and/or congeners should fulfill the following requirements:

- A direct relation between structure and compound code or numbering should exist.
- The numbers or codes should allow to identify compounds with partly similar structures.
- A short number or code should be chosen. Combinations of numbers and characters are more difficult to remember as well as numbers with more than 3 digits.
- The structure should be derived from the identification code without using tables or other aids, and the approach should be easy to recall.

Recently, a numbering system for polychlorinated bornanes was proposed on a poster presented by Tribulovich et al. at the DIOXIN'94 conference in Kyoto, Japan[3)]. Unfortunately, this system was not included in the proceedings and, therefore, the following description has to be given from memory. All positions of the bornane structure which can be occupied by a chlorine atom, were identified by a number from 1 to 13. The presence of a chlorine atom at these positions was marked by 1, a missing chlorine atom by 0. The obtained sequence of numerals 1 and 0 was then interpreted as a binary code and translated into a decimal number. By remembering the numbering of the positions of the bornane structure which can be chlorinated, the decimal number can be re-transformed into the binary number which again gives the number and exact positions of the Cl-atoms present in the congener.

This article describes the potential of a general coding system for classes of polychlorinated compounds with a large number of congeners. It is based on the creation of a binary number which represents each position of the molecule or a molecule substructure which can be substituted by a Cl-atom. This binary number is then transformed into a decimal number by adding the corresponding powers of 2 for each digit of the binary number. By selecting two important substructures which are coded separately, it is possible to include all information necessary for a precise description of the steric structure of the molecule within a double 2- or 3-digits number of the type aa-bb or aaa-bbb.

2. Applied coding and numbering systems

Polychlorinated biphenyls and biphenyl ethers

For polychlorinated biphenyls and biphenyl ethers the same coding system can be applied since the positions which can be substituted by Cl are numbered exactly in the same way. According to the IUPAC nomenclature the biphenyl system is numbered as shown in Figure 1.

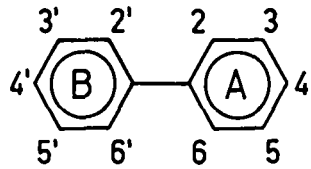

Figure 1: Numbering of the biphenyl structure according to IUPAC

Since the Cl-substitution of each ring is an important structure information, it was decided to code both benzene rings separately as shown in Table 1.

Table 1: Conversion of single ring positions of polychlorinated biphenyls and biphenyl ethers into a binary number and its conversion to the numerical code.

Position of ring system A or B:	**2 or 2'**	**3 or 3'**	**4 or 4'**	**5 or 5'**	**6 or 6'**
Binary number, (Cl present = 1, not = 0)					
Corresponds to the n-th power of 2	0	1	2	3	4
Value of the ring position as 2^n	1	2	4	8	16

Consequently, the conversion of the binary number 11111 for the fully chlorinated ring system 2,3,4,5,6 or 2',3',4',5',6' into the corresponding decimal number will give 1+2+4+8+16=31. This means that 2-digits decimal numbers will describe completely the chlorine substitution of one benzene ring. The last rule needed is that the decimal number for ring A (positions 2,3,4,5,6) is written first and that for ring B (positions 2',3',4',5',6') is placed after a hyphen giving codes of the type A-B. Decoding is done as shown for the numerical code 21-17. 21 is the combination

of the powers of 2 of 1+4+16 which is equal to 2,4,6 (see Table 1), and 17 corresponds to 1+16 which gives the positions 2',6'. The final structure is then 2,2',4,6,6'-PeCB (CB 104).

Polychlorinated bornanes

As Figure 2 shows, in principle 18 positions can be occupied by Cl in the bornane structure. By using ordinary numbers for exo- and primed numbers for endo-substitutions, one can list them as follows: 2,2',3,3',4,5,5',6,6',8α,8β,8γ,9α,9β,9γ,10α,10β,10γ.

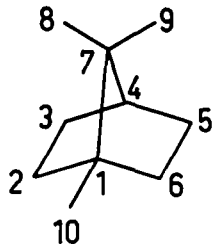

Figure 2: Carbon atom skeleton of the bornane structure

A binary number which includes totally 18 positions would correspond to a decimal number with 6 digits, and the highest number would be 2^{18}-1 = 262143. This number is much higher than the sum of all possible polychlorinated bornane congeners according to Vetter[4] and such long numbers are difficult to remember. In addition, compounds with the same ring substitution would have completely different numbers. Therefore, the coding of the ring and the methyl groups as the two most important substructures was carried out separately. The first 3-digits number represents the ring system and the second describes the Cl-substitution of the three methyl groups (see Table 2).

Consequently, the numerical codes consist of a 3-digits number < 512 for the ring structure and a 3-digits number for the methyl groups, the first digit represents C8, the second C9 and the third C10. The number and the steric position of the Cl-atoms at the methyl C-atoms are converted to a decimal number of 0 to 7 by adding the powers of 2^n for the occupied positions. The numbers representing different Cl substitutions are as follows: no Cl = 0, Cl in α = 1, β = 2, γ = 4,α,β = 3,α,γ = 5, β,γ = 6, α,β,γ = 7.

The polychlorinated bornane 2-exo,3-endo,5-exo,6-endo, 8α, 8β, 10α,10β-octachlorobornane has the binary code 100101001 for the ring system which corresponds to the decimal number 1+8+32+256=297. The binary code for the methyl groups is 1+2|0|1+2 which gives the code 303. Both codes are combined with a hyphen to 297-303. Decoding is as easy as for PCB. The bornane structure of the numerical code 237-071 can be reconstructed as follows. 237 consists of the following numbers which are the n-th power of 2, 1+4+8+32+64+128. This corresponds to the binary number 101101110 which is transformed to the ring structure 2-exo,3-exo,3-endo,5-exo,5-endo,6-exo. The methyl group code 071 is converted into 0|1+2+4|1 which gives 9α,9β,9γ,10α. The final structure is then 2-exo,3-exo,3-endo,5-exo,5-endo,6-exo,9α,9β,9γ,10α-decachlorobornane.

Table 2: Conversion algorithm for the binary numbers describing the positions of Cl-atoms at the C-atoms of the ring and the methyl groups of polychlorinated bornanes. Non-primed numbers assign exo- and primed number endo positions.

Position at the ring:	**2**	**2'**	**3**	**3'**	**4**	**5**	**5'**	**6**	**6'**
Binary number (Cl present = 1, not = 0)									
Corresponds to the n-th power of 2	0	1	2	3	4	5	6	7	8
Value of the ring position as 2^n	1	2	4	8	16	32	64	128	256

Cl-substitution of the methyl groups	**8α**	**8β**	**8γ**	**9α**	**9β**	**9γ**	**10α**	**10β**	**10γ**
Binary number (Cl present = 1, not = 0)									
Corresponds to the n-th power of 2	0	1	2	0	1	2	0	1	2
Value of methyl group position as 2^n	1	2	4	1	2	4	1	2	4
	Digit 1			Digit 2			Digit 3		

3. Discussion

Polychlorinated biphenyls

Does the proposed numerical code for PCB code really represent an improvement compared to the numbering system based on the listing of all PCB congeners[2)]? For a first answer, let us have a look on some of the important PCB congeners present in biota. The World Health Organization recommends to analyse non-ortho substituted PCB congeners due to their toxic properties which are comparable to those of 2,3,7,8-chlorine substituted PCDD/PCDF[5)]. CB 77 (3,3',4,4'-TeCB), 126 (3,3',4,4',5-PeCB) and 169 (3,3',4,4',5,5'-HxCB) are most frequently determined. CB 81 (3,4,4',5-TeCB) is also sometimes quantified. Their binary codes are given in Table 3.

Table 3: Binary codes of CB 77, 81, 126, 169

Cl-position in ring A:	**2**	**3**	**4**	**5**	**6**
Value of the ring position as 2^n	1	2	4	8	16
Binary code (Cl present = 1, missing = 0)					
CB 77	0	1	1	0	0
CB 81	0	1	1	1	0
CB 126	0	1	1	1	0
CB 169	0	1	1	1	0
Cl-position in ring B:	**2'**	**3'**	**4'**	**5'**	**6'**
Binary code (Cl present = 1, missing = 0)					
CB 77	0	1	1	0	0
CB 81	0	0	1	0	0
CB 126	0	1	1	0	0
CB 169	0	1	1	1	0

This gives for CB 77 the numerical code 6-6 (ring A: 0+2+4+0+0, ring B: 0+2+4+0+0), 14-4 for CB 81 (ring A: 0+2+4+8+0, ring B: 0+0+4+0+0), 14-6 for CB 126 and 14-14 for PCB 169. Is this now a much better numbering which allows us to receive information about the structure? First,

the position 2 is not occupied at non-ortho substituted rings and, therefore, $2^0 = 1$ is never added. This means that all codes with two even numbers designate unequivocally non-ortho substituted PCB. Second, the substitution of ring B is the same for CB 77 and 126 since the second code number is 6. Third, ring A is identical (=14) for CB 81, 126 and 169. All this information is not available from the official numbers CB 77, CB 81, CB 126 and 169.

Now let us go one step further and have a look on the PCB indication congeners which should be quantified according to international recommendations[6]. These are CB 28, 52, 101, 138, 153 and 180. With the exception of CB 28, all are di-ortho substituted. Furthermore, some mono-ortho substituted congeners and a few di-ortho substituted PCB have dioxin-like properties, and dioxin toxicity equivalency factors (TEFs) have been proposed[5]. Their numbers according to Ballschmiter and Zell[2] are listed in Table 4 together with their numerical codes according to the system proposed in this article.

Table 4: Numerical codes of selected mono- and di-ortho substituted PCB

Mono-ortho substituted PCB					
CB 28	2,4,4'-TrCB	5-4	CB 156	2,3,3',4,4',5-HxCB	15-6
CB 105	2,3,3',4,4'-PeCB	7-6	CB 157	2,3,3',4,4',5'-HxCB	7-14
CB 114	2,3,4,4',5-PeCB	15-4	CB 167	2,3',4,4',5,5'-HxCB	13-14
CB 118	2,3',4,4',5-HxCB	13-6	CB 189	2,3,3',4,4',5,5'-HpCB	15-14
CB 123	2',3,4,4',5-PeCB	14-5			
Di-ortho substituted PCB					
CB 52	2,2',5,5'-TeCB	9-9	CB 153	2,2',4,4',5,5'-HxCB	13-13
CB 101	2,2',4,5,5'-PeCB	13-9	CB 170	2,2',3,3',4,4',5-HpCB	15-7
CB 138	2,2',3,4,4',5'-HxCB	7-13	CB 180	2,2',3,4,4',5,5'-HpCB	15-13

As can be seen from Table 4 the code numbers of the di-ortho substituted PCB are all odd-odd numbers, those of the mono-ortho substituted are odd-even numbers. Is this valid for all congeners in both groups? Di-ortho substituted PCB have Cl-atoms at either the 2,2', 2,6 or 2',6' positions. 6,6' would give an even number, but this substitution is not possible due to the numbering rules. Therefore, all di-ortho substituted PCB will have odd numbered codes since 2 or 2' corresponds to 2^0=1. Mono-ortho substituted PCB will always have a Cl-atom at the position 2 or 2' since 6 or 6' is excluded as explained before. Therefore, odd-even or even-odd numbered codes represent mono-ortho substituted PCB congeners.

One can also identify tri- and tetra-ortho substituted PCB congeners. Tri-ortho PCB have Cl-atoms at the positions 2 or 2' and 6 or 6' of one ring which corresponds to the code 1+16=17 (see Table 1). Consequently, one ring code of all tri-ortho substituted PCB must be ≥17. The 2 or 2' position of the other ring has also a Cl-atom resulting in a code which must be odd-numbered and between 1 and 15 (the second 6 or 6' position (=16) has no Cl-atom). For tetra-ortho substituted PCB both ring codes must be odd-numbered and ≥17.

A big advantage of the proposed coding system is the easy conversion of the numerical code into the structure as demonstrated in the chapter "applied coding systems". One has only to remember that the ring positions correspond to powers of 2 starting with n = 0 for 2 (2^0=1) and n = 4 for 6 (2^4=16) as described in Table 1.

Numerical codes for polychlorinated bornanes

The advantages of the coding system for polychlorinated bornanes are as follows. Two 3-digits numbers combined with a hyphen describe the structures of all possible congeners. The first one defines the number of Cl-atoms in the ring and their positions (exo or endo). Rings with the same numerical code have identical structures. Furthermore, a 3 digit number describes completely the number of Cl-atoms and their steric positions at all 3 methyl groups. Also in this case coding and decoding can be carried out by only remembering the approach shown in Table 2. Enantiomers can be marked by an additional small letter. Work is in progress to develop a similar system for chlordane congeners and related compounds.

Of course, the most important question is: Has this system a future compared to the established numbering according to Ballschmiter and Zell[2] and IUPAC? The answer could be yes, if promotion is carried out over a longer period, if the system is accepted by international committees and by IUPAC and if the congeners are marked simultaneously by both the IUPAC number and the numerical code over a longer period of time. The replacement of a measuring unit by another one represents a similar situation. As numerous examples have shown, this does not pose any serious problem as long as the new system is more logical, more practical or more in agreement with the daily needs.

Acknowledgement

R. Kallenborn thanks for a DEMINEX research fellowship. This work was part of the institute programme "Organochlorines in the Arctic" which is financially supported by the Norwegian Ministry of Environment.

4. References

1) Hutzinger, O., S. Safe and V. Zitko (1974). The chemistry of PCB's. CRC Press, Cleveland.
2) Ballschmiter, K. and M. Zell (1980). Analysis of polychlorinated biphenyls (PCB) by glass capillary gas chromatography. Fresenius Z. Anal. Chem. **302**, 20-31.
3) Tribulovich, V.G., V.A. Nikiforov, V.S. Karavan, S.A. Miltsov and S. Bolshakov (1994). Synthesis and characterisation of toxaphene congeners. In: Proceedings 14th Int. Symp. on Chlorinated Dioxins and Related Compounds (Eds. H. Fiedler, O. Hutzinger, R. Clement, S. Sakai), Organohalogen compounds, Vol. 19, Kyoto University, Kyoto, Japan, ISBN 4-924960-94-2.
4) Vetter, W. (1993). Toxaphene. Theoretical aspects of the distribution of chlorinated bornanes including symmetrical aspects. Chemosphere, **26**, 1079-1084.
5) Ahlborg, U.G., Becking, G.C., Birnbaum L.S., Brouwer, A., Derks, H.J.G.M., Feeley, M., Golor, G., Hanberg, A., Larsen, J.C., Liem, AKD, Safe, S.H., Schlatter, C., Wærn, F., Younes, M. & Yrjänheikki, E. (1994). Toxic equicalency factors for dioxin-like PCBs. Chemosphere **28**, 1049-1067.
6) Duinker, J.C., D.E. Schultz and G. Petrick (1988). Selection of chlorinated biphenyl congeners for analysis in environmental samples. Marine Pollut. Bull. **19**, 19-25.