

Racemic carbohydrates – fact or fiction?

Alexander Senning

Department of Chemistry, Building 207, Technical University of Denmark, DK-2800 Kgs.

Lyngby, Denmark

E-mail: aes@kemi.dtu.dk

Abstract

Chemical Abstracts Service has developed unsound practices in the naming and handling of simple carbohydrates such as aldopentoses **1**, aldohexoses **2**, and ketohexoses **3**. Typically, the common name glucose is sometimes, inappropriately, interpreted as meaning DL-glucose DL-**2d**. Thus, a considerable number of CA names and registry numbers have been created for non-existing racemic carbohydrates and linked to irrelevant references which, moreover, in many cases cannot be retrieved by the SciFinder Scholar program.

Keywords: Aldohexoses, aldopentoses, carbohydrates, chemical abstracts, 2-ketohexoses, racemates, registry numbers

Introduction

Off hand it is easy to dismiss racemic carbohydrates such as DL-glucose DL-**2d** or DL-ribose DL-**1c** as most unlikely objects of chemical endeavor – the non-enantiospecific or partially non-enantiospecific stepwise construction of, say, pentoses or hexoses will never lead to racemates, but rather to complicated diastereomeric mixtures. Should a preexisting single carbohydrate stereoisomer be subjected to non-stereospecific chain elongation a mixture of anomers, but never a racemate, would be produced. As far as theoretical chemistry is concerned it is hard to imagine an uncharted property of a racemic carbohydrate worth the computational effort. Finally, biological studies of racemic carbohydrates would add nothing to the already extensive knowledge of the natural enantiomers and their unnatural counterparts. For accounts of organic stereochemistry in general, cf. Ref.¹ for the systematics and stereochemistry of the common monosaccharides, cf. Ref.² Thus, a database search for racemic carbohydrates should give zero results or nearly so, or should it?

Results and Discussion

All searches mentioned in the following were performed on September 26, 2007 or later with SciFinder Scholar 2007.³ Tables 1, 2, and 3 show how some of the common monosaccharides are treated in the Chemical Abstracts System.⁴

Table 1. Aldopentoses **1** in *Chemical Abstracts*

CA name [other names]	RN	number of references indicated in the Registry File	number of references retrieved in the CA File
D-arabinose D-1a [D(-)-arabinose, (-)-arabinose]	10323-20-3	~2162	2162
L-arabinose L-1a [L(+)-arabinose]	5328-37-0	~2680	2680
arabinose DL-1a [DL-arabinose, (±)-arabinose, dl-arabinose]	147-81-9	~7824	10031
D-lyxose D-1b	1114-34-7	~664	664
L-lyxose L-1b	1949-78-6	~160	160
lyxose DL-1b [(±)-lyxose, dl-lyxose]	65-42-9	~452	472
D-ribose D-1c [ribose]	50-69-1	~6597	9249
L-ribose L-1c	24259-59-4	~201	201
ribose DL-1c [DL-ribose, (±)-ribose, dl-ribose]	34466-20-1	~27	28
D-xylose D-1d [D(+)-xylose, (+)-xylose, wood sugar]	58-86-6	~16623	16626
L-xylose L-1d	609-06-3	~423	423
xylose DL-1d [DL-xylose, (±)-xylose, dl-xylose]	25990-60-7	~95	98

CA's treatment is problematic in a number of cases and for several reasons. Let us first take the CA name arabinose **1a** with RN [147-81-9] which is defined as being synonymous with the non-CA names DL-arabinose, (±)-arabinose, and dl-arabinose (cf. Table 1). As used by mainstream chemists or other professionals in any imaginable rational context 'arabinose' will mean 'D-arabinose', 'L-arabinose' or 'arabinose regardless of its D- or L-configuration', but never the racemate DL-arabinose **DL-1a**. In the extremely rare case where anybody would wish to discuss DL-arabinose **DL-1a** it would, for obvious reasons, be called 'DL-arabinose' and never 'arabinose'. An additional, unexpected problem crops up when one notes that the number of references retrieved in the CA File⁴ for **DL-2c**, **DL-2d**, **DL-2h**, **DL-3a**, and **DL-3c** dramatically outnumbers the number of references shown in the Registry File⁴ for the registry number in question. Thus, the CA name 'fructose' is identified as **DL-3a**, RN [30237-26-4], in the Registry File⁴ with ~76 associated references. This same registry number, however, retrieves an unwieldy

and for all practical purposes irrelevant 10623 references in the CA File, a remarkable flaw in the search capabilities of SciFinder Scholar which renders the user unable to inspect the implied 76 'genuine' references. At the same time one should note that the non-CA name 'fructose' is by CA, and rightly so, regarded as synonymous with the CA name 'D-fructose' [57-48-7] (cf. Table 3).

Table 2. Aldohexoses **2** in *Chemical Abstracts*

CA name [other names]	RN	number of references indicated in the Registry File	number of references retrieved in the CA File
D-allose D-2a	2595-97-3	~371	371
L-allose L-2a	7635-11-2	~31	31
allose DL-2a [dl-allose]	6038-51-3	~263	307
D-altrose D-2b	1990-29-0	~187	187
L-altrose L-2b	1949-88-8	~36	36
altrose DL-2b [dl-altrose]	5987-68-8	~247	247
D-galactose D-2c [D-(+)-galactose, (+)-galactose, galactose]	59-23-4	~25280	25292
L-galactose L-2c	15572-79-9	~395	395
galactose DL-2c [DL-galactose, (±)-galactose, dl-galactose]	26566-61-0	~81	12876
D-glucose D-2d [D-(+)-glucose, (+)-glucose, glucose, dextrose, meritose, corn sugar, grape sugar]	50-99-7	~209901	314248
L-glucose L-2d [L-(-)-glucose, l-glucose]	921-60-8	~889	889
glucose DL-2d [DL-glucose, (±)-glucose]	58367-01-4	~33	736
D-gulose D-2e	4205-23-6	~150	150
L-gulose L-2e	6027-89-0	~81	81
gulose DL-2f [DL-gulose]	19163-87-2	~236	248
D-idose D-2g	5978-95-0	~100	100
L-idose L-2g	5934-56-5	~53	53
DL-idose DL-2g	2152-76-3	~187	195
D-mannose D-2h [D-(+)-mannose, (+)-mannose, mannose, carubinose, seminose]	3458-28-4	~16306	16310
L-mannose L-2h	10030-80-5	~216	216
mannose DL-2h [DL-mannose, (±)-mannose, dl-mannose]	31103-86-3	~82	6865
D-talose D-2i [D-(+)-talose]	2595-98-4	~286	286
L-talose L-2i	23567-25-1	~41	41
talose DL-2i [dl-talose]	30077-17-9	~283	283

Table 3. 2-Ketohexoses **3** in *Chemical Abstracts*

CA name [other names]	RN	number of references indicated in the Registry File	number of references retrieved in the CA File
D-fructose D- 3a [D-(–)-fructose, D-(–)-levulose, fructose, levulose, fruit sugar, D- <i>arabino</i> -2-hexulose]	57-48-7	~34145	34200
L-fructose L- 3a	7776-48-9	~167	167
fructose DL- 3a [DL-fructose, (±)-fructose, dl-fructose, methose, <i>arabino</i> -2-hexulose]	30237-26-4	~76	10623
D-psicose D- 3b [D-allulose, D- <i>ribo</i> -2-hexulose]	551-68-8	~218	218
L-psicose L- 3b [L- <i>ribo</i> -2-hexulose]	16354-64-6	~35	35
psicose DL- 3b [DL-psicose, allulose, erythrohexulose, pseudofructose, <i>ribo</i> -2-hexulose]	23140-52-5	~148	174
D-sorbose D- 3c [D-(+)-sorbose, sorbinose, D- <i>xylo</i> -2-hexulose]	3615-56-3	~25280	25292
L-sorbose L- 3c [L-(–)-sorbose, L-sorbinose, L- <i>xylo</i> -2-hexulose, L-1,3,4,5,6-pentahydroxyhex-an-2-one]	87-79-6	~2491	2492
L-sorbose (incompletely defined substance) L- 3c	97806-30-9	0	0
sorbose DL- 3c [DL-sorbose, (±)-sorbose, dl-sorbose, <i>xylo</i> -2-hexulose]	3615-39-2	~35	441
D-tagatose D- 3d [D- <i>lyxo</i> -2-hexulose]	87-81-0	~365	365
L-tagatose L- 3d [L- <i>lyxo</i> -2-hexulose]	17598-82-2	~43	43
tagatose DL- 3d [DL-tagatose, <i>lyxo</i> -2-hexulose]	17598-81-1	~325	400

Racemic carbohydrates in CAS

To sum it up the CA names and registry numbers assigned to the ‘racemates’ arabinose [147-81-9], lyxose [65-42-9], ribose [34466-20-1], xylose [25990-60-7] (cf. Table 1), allose [5038-51-3], altrose [5987-68-8], galactose [26566-61-0], glucose [58367-01-4], gulose [19163,87-2], idose [2152-76-3], mannose [31103-86-3], talose [30077-17-9] (cf. Table 2), fructose [30237-26-4], psicose [23140-52-5], sorbose [3615-39-2], and tagatose [17598-81-1] (cf. Table 3) do not and

cannot refer to any racemates and are thus unsuitable to retrieve any useful information (cf. Tables 1, 2, and 3).

Typical hits for 'glucose' [58367-01-4] retrieve references dealing with subjects such as 'glucose tolerance test' or glucose as pharmaceutical excipient, without any even remote involvement of DL-glucose DL-**2d**.

The corresponding sugar hemiacetals such as glucopyranose [54-17-1] with 89 associated references and β -galactofuranose [131064-96-5] with 4 associated references receive a corresponding unsound treatment as undocumented racemates. For the sake of brevity we refrain from elaborating further details here.

Conclusions

It appears that poor documentalistic practices have been allowed to create a niche of virtual carbohydrate chemistry in the CA Registry File which is apt to confuse novices and to baffle practicing organic chemists or documentalists. As could be expected a priori and as confirmed by numerous spot checks none of the references linked to the racemic carbohydrates and their registry numbers do in fact mention them, explicitly or implicitly.

References

1. Eliel, E. L.; Wilen, S. H. *Stereochemistry of Organic Compounds*, John Wiley & Sons, Inc.: New York etc., 1994.
2. Belitz, H.-D.; Schieberle, P.; Grosch, W. *Food Chemistry*, 3rd Ed., Springer: Berlin, 2004.
3. <http://www.cas.org/products/sfacad/index.html>.
4. <http://www.cas.org/support/academic>.

Authors' biographical data



Alexander Senning was born in 1936 in Riga, Latvia. He studied chemistry at Munich, Germany (1954-59) and Uppsala, Sweden (1960-62). He obtained a Ph.D. in organic chemistry from Uppsala University (1962), joined the Department of Chemistry, Aarhus University, Denmark as assistant professor (1962-65) and served as associate professor during 1965-93. During a sabbatical leave (1973-75) he was head of the research laboratory of the drug company A/S Alfred Benzon, Copenhagen, Denmark. He joined the Danish Engineering Academy (DIA), Lyngby, Denmark, later part of the Technical University of Denmark (DTU), Kgs. Lyngby, Denmark, as professor of organic chemistry in 1993, until his retirement in 2003. Alexander Senning is the author of *Elsevier's Dictionary of Chemoetymology*, published in 2006. His research interests include organic sulfur chemistry and medicinal chemistry. He has been intensively involved as a journal and handbook editor.