

Paivi Riikonen
ORC-J

August 10, 2006

09006-

29. HSC Formula Syntax for Hydrocarbon Species

Identification of species in HSC Chemistry is based on unique chemical formulae for each different species. This idea works fine with basic inorganic chemicals, but there are some problems with complicated organic compounds. The structural formula may be too long and inconvenient to use and there may be several different species for the same cross formula. Another problem is the large number of synonyms for many organic compounds. Therefore **cross formulas with specific suffix** have been used in the HSC database for the most of the organic compounds, see Chapter 21.2.1. This appendix will give more information on the syntax for organic species and instructions how to find a specific compound from the database.

Chemical and common names, as well as CAS numbers, are given for the most of the species in the HSC database. They will help a lot to identify the compounds. The species in the HSC database are arranged in alphabetical order by chemical formulae and suffixes. For example, 4-Ethyl-1,2-dimethylbenzen C10H14(4E12DMB) is before the Butylbenzene C10H14(BB) in the species list.

The chemical names of the species are usually based on IUPAC¹ rules. These may be summarized as follows: A) find the longest carbon chain in the compound, B) name each appendage group which is attached to this principal chain, C) alphabetize the appendage groups and D) number the principal chain from one end in such a way that the lower number is used at the first point of difference in the two possible series of locates.

A functional group in the hydrocarbon, either a double bond, a hydroxy group or an amino group will determine both the characteristics and the name of the compound. The functional group will have the lowest number in the principal chain of the hydrocarbon. If there are several functional groups in the compound, the name of the compound is determined according to the strongest functional group. For example, we have aminoacids, which contain both amino and acid groups, but they are called acids because the acid group is stronger than the amino group.

29.1 Basic hydrocarbons, C_xH_y

Naming and marking a basic hydrocarbon begins from the alkanes homologues series. Alkanes, alkenes and alkynes are marked by similar letters, but cannot be mixed up because their chemical formulas differ in the number of hydrogen atoms. Similarly the appendage groups derived from alkanes use the very same letters.

Remember that the number of appendage groups does not affect the alphabetical order of the appendage groups in the chemical name. Numbers are also marked. For example 3-Ethyl-2,4-dimethylpentane is marked **3E24DMP**. Notice also that in the chemical name there is a comma separating the different numbers not a point. "Mono" is seldom used in the names of hydrocarbons and exists often only in the deuterium compound names.

Sometimes straight chain alkanes have the **n** character in their name, like the n-butane, which means a normal butane, so the compound is not the isobutane. In the HSC **n** is not used in the compound names. Cyclocompounds are marked with the **C** character, for example, cyclobutane is marked **CB**. There are also deuterium compounds in the HSC database. Their formula is the same as the corresponding hydrogen formula, but deuterium is marked with the **D** character.

If there is a double bond in the compound there may be a chance that there are two different stereoisomers, cis and trans or **Z** and **E**. These isomers are named and marked as different compounds and the **c**, **t**, **Z** or **E** character is located before the actual compound name. If there is a chiral C-atom in the hydrocarbon then the compound is optically active. The absolute configuration of the compound is determined by **D** and **R** characters before the actual name of the compound. Optically active isomers interact with plane polarized light a different way, and that is marked by - and + in the isomer name.

Large hydrocarbon compounds can be very complicated. The appendage group may have its own appendage groups and there might be parentheses in the compound name; parentheses are not, however, used in the suffix.

Paivi Riikonen
ORC-J

August 10, 2006

09006-

Table 1.

Chemical Name	Suffix	Formula
1-Butyl-2-methylbenzene	1B2MB	C11H16
3-Ethyl-2,2-dimethylpentane	3E22DM	C9H20
Tridecylcyclohexane	TCH	C19H38

29.1.1 Appendage groups

Common alkane type appendage groups are the iso-group, sec-group and tert-group. These are used in the common name, but not in the chemical name.

Table 1.1.

Chemical Name	Common Name	Suffix	Formula	Structural formula
1,1-Dimethylethyl-	tert-Butyl-	TB-		-C(CH ₃) ₃
1-Methylethyl-	Isopropyl-	IP-		-CH(CH ₃) ₂
2-Chlorobutane	sec-Butyl chloride	SBC	C ₄ H ₉ Cl	CH ₃ CH ₂ CHClCH ₃

29.1.2 Aromatic compounds

Benzenes are marked **B**. In the large compounds there might be a need to consider benzene as an appendage group in which case it is marked **P**, phenyl-.

If there are only two appendage groups, the name of the benzene compound can be formed by the ortho- meta- para-system. Ortho- (shortened **-o-**), appendage groups are in the 1 and 2 positions, in meta- (shortened **-m-**), they are in the 1 and 3 positions and in para- (shortened **-p-**) they are in the 1 and 4 positions. In the HSC database ortho-meta-para derived names are used only in the common names. Many aromatic compounds have specific common names.

Table 1.2.

Chemical Name	Common Name	Suffix	Formula	Structural formula
1,2-(1,8-Naphthalene)benzene	Fluoranthene	FLU	C ₁₆ H ₁₀	
1,2-Dimethylbenzene	o-Xylene	OXY	C ₈ H ₁₀	C(CH ₃)C(CH ₃)CHCHCH
1,3-Dimethylbenzene	m-Xylene	MXY	C ₈ H ₁₀	C(CH ₃)CHC(CH ₃)CHCHCH
1,4-Dimethylbenzene	p-Xylene	PXY	C ₈ H ₁₀	H ₃ C(C ₆ H ₄)CH ₃
1H-Indene	Indene	IN	C ₉ H ₈	(C ₆ H ₄)(C ₃ H ₄)
1-Methylethylbenzene	Cumene	CUM	C ₉ H ₁₂	(C ₆ H ₅)CH(CH ₃) ₂
Anthracene	Anthracene	A	C ₁₄ H ₁₀	(C ₆ H ₄)(C ₂ H ₂)(C ₆ H ₄)
Benzene	Benzene	B	C ₆ H ₆	
Benzo(a)phenanthrene	Chrysene	CR	C ₁₈ H ₁₂	
Benzo(def)phenanthrene	Pyrene	PYR	C ₁₆ H ₁₀	
Bicyclo(2.2.1)hept-2-ene	2-Norbornene	2NOR	C ₇ H ₁₀	
Bicyclo(5.3.0)deca-2,4,6,8,10-	Azulene	AZE	C ₁₀ H ₈	
Dibenz(de,kl)anthracene	Perylene	PER	C ₂₀ H ₁₂	
Ethylbenzene	Styrene	STY	C ₈ H ₈	C ₆ H ₅ CHCH ₂
Methylbenzene	Toluene	TLU	C ₇ H ₈	C ₆ H ₅ CH ₃
Naphthalene	Naphthalene	N	C ₁₀ H ₈	(C ₆ H ₄)(C ₄ H ₄)
Phenanthrene	Phenanthrene	PA	C ₁₄ H ₁₀	

Paivi Riikonen
ORC-J

August 10, 2006

09006-

Phenylbenzene	Biphenyl	BP	C12H10	(C6H5) ₂
---------------	----------	----	--------	---------------------

29.2 Halogen compounds

All the halogen compounds containing the carbon atom are named hydrocarbons. If there is more than one halogen, halogens follow the alphabetical order. Halogens are also marked with single letters derived from the name.

Chemical Name	Common Name	Suffix	Formula	Structural formula
Bromotriiodomethane	Bromotriiodomethane	BTIM	CBrI ₃	
Chloromethane	Methyl chloride	CM	CH ₃ Cl	

29.3 Hydrocarbons containing nitrogen

29.3.1 Amines, R-NH₂, R1-NH-R2, R1,R2-N-R3

Amines are marked with **A**. For example, the hexanamine is marked **HA**. In an amine, the hydrogen atoms of nitrogen can be substituted by different appendage groups. If there is more than one substituent the place of the substituent is informed by the **N** character. The name of the compound is determined by the most complicated substituent in the amine. Many cyclic amines have specific names.

Chemical Name	Common Name	Suffix	Formula	Structural formula
1H-Indole	Indole	IND	C ₈ H ₇ N	
1H-Pyrrole	Azole	PYR	C ₄ H ₅ N	CHCHNCHCH
4-Methylbenzenamine	p-Toluidine	PTO	C ₇ H ₉ N	(C ₆ H ₄)CH ₃ NH ₂
Benzenamine	Aniline	ANI	C ₆ H ₇ N	C ₆ H ₅ NH ₂
Benzo(b)pyridine	Quinoline	QUI	C ₉ H ₇ N	
Isoquinoline	Isoquinoline	IQL	C ₉ H ₇ N	
Pyridine	Azine	PYR	C ₅ H ₅ N	NCHCHCHCHCH

29.3.2 Amino acids

Amino acids have specific names.

Chemical Name	Common Name	Suffix	Formula	Structural formula
2-Amino-3-hydroxy- butanoic acid	Threonine	THR	C ₄ H ₉ NO ₃	
2-Amino-3-indolepropanoic acid	Tryptophan	TRP	C ₁₁ H ₁₂ N ₂ O ₂	
2-Amino-3-methylbutanoic acid	Valine	VAL	C ₅ H ₁₁ NO ₂	
2-Amino-3-phenyl- propanoic acid	Phenylalanine	PHE	C ₉ H ₁₁ NO ₂	
2-Aminopentanoic acid	Glutamic acid	GLU	C ₅ H ₉ NO ₄	
2-Aminopropanoic acid	Alanine	ALA	C ₃ H ₇ NO ₂	CH ₃ CH(NH ₂)COOH
2,6-Diaminohexanoic acid	Lysine	LYS	C ₆ H ₁₄ N ₂ O ₂	H ₂ N(CH ₂) ₄ CH(NH ₂)CO ₂ H
2-Aninosuccinamic acid	Asparagine	ASN	C ₄ H ₈ N ₂ O ₃	H ₂ NCOCH ₂ CH(NH ₂)COOH
2-Aminobutanedioic acid	Aspartic acid	ASP	C ₄ H ₇ NO ₄	
3-(4-Hydroxyphenyl)alanine	Tyrosine	TYR	C ₉ H ₁₁ NO ₃	

Paivi Riikonen
ORC-J

August 10, 2006

09006-

Aminoacetic acid	Glycine	GLY	C2H5NO2	H2NCH2COOH
S-2-amino-3-hydroxy-	Serine	SER	C3H7NO3	
S-2-amino-4-methyl-	Leucine	LEU	C6H13NO2	
S-2,5-diamino-5-oxo-pentanoic acid	Glutamine	GLN	C5H10N2O3	

29.3.3 Hydrazines, *R-NH-NH₂*

Hydrazines are marked with **H** character.

Chemical Name	Common Name	Suffix	Formula	Structural formula
1,1-Dimethylhydrazine	1,1-Dimethyl-hydrazine	11DMH	C2H8N	(CH ₃) ₂ NNH ₂
Methylhydrazine	Methylhydrazine	MH	CN ₂ H ₆	H ₃ CNHNH ₂

Paivi Riikonen
ORC-J

August 10, 2006

09006-

29.3.4 Amides, $R-C=O - NH_2$

Amides are marked with **A**. In an amide the hydrogen atoms of nitrogen can be substituted by different appendage groups. If there is more than one substituent the place of the substituent is informed by the N character. The name of the compound is determined by the most complicated substituent in the amide.

Table 3.4.				
Chemical Name	Common Name	Suffix	Formula	Structural formula
Hexanamide	Hexanamide	HA	C ₆ H ₁₃ NO	CH ₃ (CH ₂) ₄ CONH ₂
Methanamide	Methanamide	MA	CH ₃ NO	HCONH ₂

29.3.5 Nitriles, $R\equiv N$

Nitriles are marked with **N**. Sometimes nitriles are called cyano-compounds, but in the HSC database cyano- is not used. Pyridine, which is a cyclic nitrile compound, is marked with **P**.

Table 3.5.				
Chemical Name	Common Name	Suffix	Formula	Structural formula
2,2-Dimethylpropanenitrile	tert-Butyl cyanide	22DMPN	C ₅ H ₉ N	(CH ₃) ₃ CCN
Hexanenitrile	Pentyl cyanide	HN	C ₆ H ₁₁ N	CH ₃ (CH ₂) ₄ CN
Propanenitrile	Ethyl cyanide	PN	C ₃ H ₅ N	CH ₃ CH ₂ CN

29.3.6 Nitro-compounds, nitrates, $R-NO_2$

Nitro-compounds are marked with **N**.

Table 3.6.				
Chemical Name	Common Name	Suffix	Formula	Structural formula
1-Nitrobutane	1-Nitrobutane	1NB	C ₄ H ₉ NO ₂	CH ₃ CH ₂ CH ₂ CH ₂ NO ₂
1-Nitropropane	1-Nitropropane	1NP	C ₃ H ₇ NO ₂	CH ₃ CH ₂ CH ₂ NO ₂

29.4 Hydrocarbons containing oxygen

29.4.1 Ethers, $R1-O-R2$

Ethers are marked with **E**. For example, the ethyl methyl ether is marked **EME**. If there is more than one ether-oxygen in the compound the compound is given an **oxy**-prefix. Some ethers have specific names.

Table 4.1.				
Chemical Name	Common Name	Suffix	Formula	Structural formula
Ethoxybenzene	Phenetole	PLE	C ₈ H ₁₀ O	C ₆ H ₅ OCH ₂ CH ₃
Ethyl methyl ether	Methoxyethane	EME	C ₃ H ₈ O	CH ₃ OCH ₂ CH ₃
Furan	Furan	F	C ₄ H ₄ O	CHOCHCHCH (cyclic)
Methyl phenyl ether	Anisole,	ANS	C ₇ H ₈ O	C ₆ H ₅ OCH ₃
Oxirane	Ethylene oxide	OXI	C ₂ H ₄ O	OCH ₂ CH ₂ (cyclic)
Oxetane	Trimethylene oxide	OXE	C ₃ H ₆ O	OCH ₂ CH ₂ CH ₂ (cyclic)
Tetrahydrofuran	Oxolane	THF	C ₄ H ₈ O	

Paivi Riikonen
ORC-J

August 10, 2006

09006-

29.4.2 Aldehydes, R-C=O -H

The end of the aldehyde name is the suffix-nal, **AL** stands for an aldehyde.

Table 4.2.				
Chemical Name	Common Name	Suffix	Formula	Structural formula
Acetaldehyde	Ethanal	ACE	C ₂ H ₄ O	CH ₃ CHO
Formaldehyde	Methanal		CH ₂ O	HCHO, H ₂ CO
Hexanal	Caproaldehyde	HAL	C ₆ H ₁₂ O	CH ₃ (CH ₂) ₄ CHO
Propanal	Propionaldehyde	PAL	C ₃ H ₆ O	CH ₃ CH ₂ CHO

29.4.3 Ketones, R1-C=O -R2

The suffix **one** is used at the end of ketone names. **N** stands for a ketone in the formula suffix. Ketones are named as straight chain alkanes, not like ethers or by the oxo-prefix.

Table 4.3.				
Chemical Name	Common Name	Suffix	Formula	Structural formula
3-Pentanone	Diethyl ketone	3PN	C ₅ H ₁₀ O	CH ₃ CH ₂ COCH ₂ CH ₃
Butanone	Ethyl methyl ketone	BN	C ₄ H ₈ O	CH ₃ CH ₂ COCH ₃
Propanone	Acetone	PN	C ₃ H ₆ O	CH ₃ COCH ₃

29.4.4 Esters

Esters are marked by taking one letter from the alcohol-derived name and two letters from the acid-derived name. In the HSC database methanoates and ethanoates are formates and acetates as they are commonly named.

Table 4.4.				
Chemical Name	Common Name	Suffix	Formula	Structural formula
Butyl acetate	Butyl acetate	BAC	C ₆ H ₁₂ O ₂	CH ₃ COOCH ₂ CH ₂ CH ₂ CH ₃
Methyl 2-methyl-2-propenoate	Methyl methacrylate	M2M2PR	C ₅ H ₈ O ₂	CH ₂ C(CH ₃)COOCH ₃
Octyl formate	Octyl formate	OFO	C ₉ H ₁₈ O ₂	HCOO(CH ₂) ₇ CH ₃
Propyl propanoate	Propyl propionate	PPR	C ₆ H ₁₂ O ₂	CH ₃ CH ₂ COOCH ₂ CH ₂ CH ₃

29.4.5 Alcohols and carbohydrates

Alcohols are marked with **OL**. Diols and triols are marked respectively **DOL** and **TOL**, if they do not have a specific common name. Many carbohydrates have specific names like glucose and mannose

Table 4.5.				
Chemical Name	Common Name	Suffix	Formula	Structural formula
1,2,3-Propanetriol	Glycerol	GLY	C ₃ H ₈ O ₃	CH ₂ OHCHOHCH ₂ OH
1,2-Ethandiol	Ethleneglycol	EGL	C ₂ H ₆ O ₂	CH ₂ OHCH ₂ OH
Ethanol	Ethanol	EOL	C ₂ H ₆ O	CH ₃ CH ₂ OH
D-(+)-glucose	D-(+)-glucose	DGLU	C ₆ H ₁₂ O ₆	

Paivi Riikonen
ORC-J

August 10, 2006

09006-

29.4.6 Phenols

Many phenol-derived compounds have specific common names.

Table 4.6.				
Chemical Name	Common Name	Suffix	Formula	Structural formula
1,2-Benzenediol	Catechol	CAT	C ₆ H ₆ O ₂	HO(C ₆ H ₄)OH
1,3-Benzenediol	Resorcinol	RES	C ₆ H ₆ O ₂	HO(C ₆ H ₄)OH
1,4-Benzenediol	Hydroquinone	HQU	C ₆ H ₆ O ₂	HO(C ₆ H ₄)OH
2-Methoxyphenol	Guaiacol	GUA	C ₇ H ₈ O ₂	CH ₃ O(C ₆ H ₄)OH
2-Methylphenol	o-Cresol	OCR	C ₇ H ₈ O	C(OH)C(CH ₃)CHCHCHCH
3-Methylphenol	m-Cresol	MCR	C ₇ H ₈ O	C(OH)CHC(CH ₃)CHCHCH
4-Methylphenol	p-Cresol	PCR	C ₇ H ₈ O	H ₃ C(C ₆ H ₄)OH
Phenol	Phenol	PHE	C ₆ H ₆ O	C ₆ H ₅ OH

29.4.7 Acids

Acids are marked with **A** character and diacids with **DA**.

Table 4.7.				
Chemical Name	Common Name	Suffix	Formula	Structural formula
Butanedioic acid	Succinic acid	SUC	C ₄ H ₈ O ₂	CH ₃ CH ₂ CH ₂ COOH
Ethanoic acid	Acetic acid	ACE	C ₂ H ₄ O ₂	CH ₃ COOH
Methanoic acid	Formic acid	FOR	CH ₂ O ₂	CHOOH
Propanoic acid	Propionic acid	PA	C ₃ H ₆ O ₂	CH ₃ CH ₂ COOH

Paivi Riikonen
ORC-J

August 10, 2006

09006-

29.5 Hydrocarbons containing sulfide**29.5.1 Thiols, R-SH**Thiols are marked with **T**.**Table 5.1.**

Chemical Name	Common Name	Suffix	Formula	Structural formula
1,4-Butanedithiol	Tetramethylenedithiol	14BDT	C ₄ H ₁₀ S ₂	CH ₂ SHCH ₂ CH ₂ CH ₂ SH
Ethanethiol	Ethyl mercaptan	ET	C ₂ H ₆ S	CH ₃ CH ₂ SH

29.5.2 Sulfides, thia-compounds, R1-S-R2Thia-compounds are named like ethers. Thiophene, which is a cyclic sulfide compound, is marked with **TH**.**Table 5.2.**

Chemical Name	Common Name	Suffix	Formula	Structural formula
Dimethyl sulfide	2-Thiapropane	DMS	C ₂ H ₆ S	CH ₃ SCH ₃
Ethyl methyl sulfide	2-Thiabutane	EMS	C ₃ H ₈ S	CH ₃ SCH ₂ CH ₃

29.5.3 Disulfides, dithia-compounds, R1-S-S-R2Disulfides are named like ethers and marked with **DS**.**Table 5.3.**

Chemical Name	Common Name	Suffix	Formula	Structural formula
Ethyl methyl disulfide	2,3-Dithiapentane	EMDS	C ₃ H ₈ S ₂	CH ₃ SSCH ₂ CH ₃

29.5.4 SulfoxidesSulfoxides are named like ethers and marked with **SX**.**Table 5.4.**

Chemical Name	Common Name	Suffix	Formula	Structural formula
Diethyl sulfoxide	1,1'-Sulfinyl-bis(ethane)	DESX	C ₄ H ₁₀ SO	(CH ₃ CH ₂) ₂ SO

29.5.5 SulfonesSulfones are named like ethers and marked with **SN**.**Table 5.5.**

Chemical Name	Common Name	Suffix	Formula	Structural formula
Dimethyl sulfone	Sulfonylbismethane	DMSN	C ₂ H ₆ SO ₂	(CH ₃) ₂ SO ₂

Paivi Riikonen
ORC-J

August 10, 2006

09006-

29.6 *Reference*

Streitweiser, A., Heatcock, C. H., Introduction to Organic Chemistry, Macmillan Publishing Company, New York, 1989.