29. HSC Formula Syntax for Hydrocarbon Species

The 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 issues 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 issue is the wide number of synonyms for many organic compounds. Therefore **cross formulae with a specific suffix** have been used in the HSC database for the most of the organic compounds, see Chapter 28 (section 28.3). This appendix will provide more information on the syntax for organic species and give instructions on how to find a specific compound from the database.

Chemical and common names, as well as CAS numbers, are given for 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 formula and suffix. For example, 4-Ethyl-1,2-dimethylbenzene C10H14(4E12DMB) is before Butylbenzene C10H14(BB) in the species list.

The chemical names of the species are usually based on IUPAC[1] 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 locations.

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, there are aminoacids containing 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 alkane homologous 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 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 an isobutane. In HSC, **n** is not used in 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 differently, which 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.

Table 1.

Chemical Name	Suffix	Formula
1-Butyl-2-methylbenzene	1B2MB	C11H16
3-Ethyl-2,2-dimethylpentane	3E22DMP	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(CH3)3
1-Methylethyl-	Isopropyl-	IP-		-CH(CH3)2
2-Chlorobutane	sec-Butyl chloride	SBC	C4H9Cl	CH3CH2CHClCH3

29.1.2. Aromatic Compounds

Benzenes are marked **B**. In 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 orto-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	C16H10	
1,2-Dimethylbenzene	o-Xylene	OXY	C8H10	C(CH3)C(CH3)CHCHCHCH
1,3-Dimethylbenzene	m-Xylene	MXY	C8H10	C(CH3)CHC(CH3)CHCHCH
1,4-Dimethylbenzene	p-Xylene	PXY	C8H10	H3C(C6H4)CH3
1H-Indene	Indene	IN	C9H8	(C6H4)(C3H4)
1-Methylethylbenzene	Cumene	CUM	C9H12	(C6H5)CH(CH3)2
Anthracene	Anthracene	A	C14H10	(C6H4)(C2H2)(C6H4)
Benzene	Benzene	B	C6H6	
Benzo(a)phenathrene	Chrysene	CR	C18H12	
Benzo(def)phenanthrene	Pyrene	PYR	C16H10	
Bicyclo(2.2.1)hept-2-ene	2-Norbornene	2NOR	C7H10	
Bicyclo(5.3.0)deca-2,4,6,8,10-pentaene	Azulene	AZE	C10H8	
Dibenz(de,kl)anthracene	Perylene	PER	C20H12	
Ethenylbenzene	Styrene	STY	C8H8	C6H5CHCH2
Methylbenzene	Toluene	TLU	C7H8	C6H5CH3
Naphthalene	Naphthalene	N	C10H8	(C6H4)(C4H4)
Phenanthrene	Phenanthrene	PA	C14H10	
Phenylbenzene	Biphenyl	BP	C12H10	(C6H5)2

29.2. Halogen Compounds

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

Table 2.

Chemical Name	Common Name	Suffix	Formula	Structural formula
Bromotriiodomethane	Bromotriiodomethane	BTIM	CBrI3	
Chloromethane	Methyl chloride	CM	CH3Cl	

29.3. Hydrocarbons Containing Nitrogen

29.3.1. Amines, R-NH_2, R1-NH-R2, R1,R2-N-R3

Amines are marked with **A**. For example, 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, its position is indicated 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.

Table 3.1.

Chemical Name	Common Name	Suffix	Formula	Structural formula
1H-Indole	Indole	IND	C8H7N	
1H-Pyrrole	Azole	PYR	C4H5N	CHCHNCHCH
4-Methylbenzenamine	p-Toluidine	PTO	C7H9N	(C6H4)CH3NH2
Benzenamine	Aniline	ANI	C6H7N	C6H5NH2
Benzo(b)pyridine	Quinoline	QUI	C9H7N	
Isoquinoline	Isoquinoline	IQL	C9H7N	
Pyridine	Azine	PYR	C5H5N	NCHCHCHCHCH

29.3.2. Amino Acids

Amino acids have specific names.

Table 3.2.

Chemical Name	Common Name	Suffix	Formula	Structural formula
2-Amino-3-hydroxy- butanoic acid	Threonine	THR	C4H9NO3	
2-Amino-3-indolepropanoic acid	Tryptophan	TRP	C11H12N2O2	
2-Amino-3-methylbutanoic acid	Valine	VAL	C5H11NO2	
2-Amino-3-phenyl- propanoic acid	Phenylalanine	PHE	C9H11NO2	
2-Aminopentanoic acid	Glutamic acid	GLU	C5H9NO4	
2-Aminopropanoic acid	Alanine	ALA	C3H7NO2	CH3CH(NH2)COOH
2,6-Diaminohexanoic acid	Lysine	LYS	C6H14N2O2	H2N(CH2)4CH(NH2)CO2H
2-Aninosuccinamic acid	Asparagine	ASN	C4H8N2O3	H2NCOCH2CH(NH2)COOH
2-Aminobutanedioic acid	Aspartic acid	ASP	C4H7NO4	
3-(4-Hydroxyphenyl)alanine	Tyrosine	TYR	C9H11NO3	
Aminoacetic acid	Glysine	GLY	C2H5NO2	H2NCH2COOH
S-2-amino-3-hydroxy-propanoic acid	Serine	SER	C3H7NO3	
S-2-amino-4-methyl-pentanoic acid	Leucine	LEU	C6H13NO2	
S-2,5-diamino-5-oxo-pentanoic acid	Glutamine	GLN	C5H10N2O3	

29.3.3. Hydrazines, R-NH-NH_2

Hydrazines are marked with **H**.

Table 3.3.

Chemical Name	Common Name	Suffix	Formula	Structural formula
1,1-Dimethylhydrazine	1,1-Dimethyl-hydrazine	11DMH	C2H8N	(CH3)2NNH2
Methylhydrazine	Methylhydrazine	MH	CN2H6	H3CNHNH2

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, its position is indicated by N. 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	C6H13NO	CH3(CH2)4CONH2
Methanamide	Methanamide	MA	CH3NO	HCONH2

29.3.5. Nitriles, R≡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	C5H9N	(CH3)3CCN
Hexanenitrile	Pentyl cyanide	HN	C6H11N	CH3(CH2)4CN
Propanenitrile	Ethyl cyanide	PN	C3H5N	CH3CH2CN

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	C4H9NO2	CH3CH2CH2CH2NO2
1-Nitropropane	1-Nitropropane	1NP	C3H7NO2	CH3CH2CH2NO2

29.4. Hydrocarbons Containing Oxygen

29.4.1. Ethers, R1-O-R2

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

Table 4.1.

Chemical Name	Common Name	Suffix	Formula	Structural formula
Ethoxybenzene	Phenetole	PLE	C8H10O	C6H5OCH2CH3
Ethyl methyl ether	Methoxyethane	EME	C3H8O	CH3OCH2CH3
Furan	Furan	F	C4H4O	CHOCHCHCH (cyclic)
Methyl phenyl ether	Anisole, Methoxybenzene	ANS	C7H8O	C6H5OCH3
Oxirane	Ethylene oxide	OXI	C2H4O	OCH2CH2 (cyclic)
Oxetane	Trimethylene oxide	OXE	C3H6O	OCH2CH2CH2 (cyclic)
Tetrahydrofuran	Oxolane	THF	C4H8O	

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	C2H4O	CH3CHO
Formaldehyde	Methanal		CH2O	HCHO, H2CO
Hexanal	Caproaldehyde	HAL	C6H12O	CH3(CH2)4CHO
Propanal	Propionaldehyde	PAL	C3H6O	CH3CH2CHO

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	C5H10O	CH3CH2COCH2CH3
Butanone	Ethyl methyl ketone	BN	C4H8O	CH3CH2COCH3
Propanone	Acetone	PN	C3H6O	CH3COCH3

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 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	C6H12O2	CH3COOCH2CH2CH2CH3
Methyl 2-methyl-2-propenoate	Methyl methacrylate	M2M2PR	C5H8O2	CH2C(CH3)COOCH3
Octyl formate	Octyl formate	OFO	C9H18O2	HCOO(CH2)7CH3
Propyl propanoate	Propyl propionate	PPR	C6H12O2	CH3CH2COOCH2CH2CH3

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	C3H8O3	CH2OHCHOHCH2OH
1,2-Ethanediol	Ethyleneglycol	EGL	C2H6O2	CH2OHCH2OH
Ethanol	Ethanol	EOL	C2H6O	CH3CH2OH
D-(+)-glucose	D-(+)-glucose	DGLU	C6H12O6	

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	C6H6O2	HO(C6H4)OH
1,3-Benzenediol	Resorcinol	RES	C6H6O2	HO(C6H4)OH
1,4-Benzenediol	Hydroquinone	HQU	C6H6O2	HO(C6H4)OH
2-Methoxyphenol	Guiacol	GUA	C7H8O2	CH3O(C6H4)OH
2-Methylphenol	o-Cresol	OCR	C7H8O	C(OH)C(CH3)CHCHCHCH
3-Methylphenol	m-Cresol	MCR	C7H8O	C(OH)CHC(CH3)CHCHCH
4-Methylphenol	p-Cresol	PCR	C7H8O	H3C(C6H4)OH
Phenol	Phenol	PHE	C6H6O	C6H5OH

29.4.7. Acids

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

Table 4.7.

Chemical Name	Common Name	Suffix	Formula	Structural formula
Butanedioic acid	Succinic acid	SUC	C4H8O2	CH3CH2CH2COOH
Ethanoic acid	Acetic acid	ACE	C2H4O2	CH3COOH
Methanoic acid	Formic acid	FOR	CH2O2	CHOOH
Propanoic acid	Propionic acid	PA	C3H6O2	CH3CH2COOH

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	C4H10S2	CH2SHCH2CH2CH2SH
Ethanethiol	Ethyl mercaptan	ET	C2H6S	CH3CH2SH

29.5.2. Sulfides, Thia-Compounds, R1-S-R2

Thia-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	C2H6S	CH3SCH3
Ethyl methyl sulfide	2-Thiabutane	EMS	C3H8S	CH3SCH2CH3

29.5.3. Disulfides, Dithia-Compounds, R1-S-S-R2

Disulfides 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	C3H8S2	CH3SSCH2CH3

29.5.4. Sulfoxides

Sulfoxides 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	C4H10SO	(CH3CH2)2SO

29.5.5. Sulfones

Sulfones are named like ethers and marked with **SN**.

Table 5.5.				
Chemical Name	Common Name	Suffix	Formula	Structural formula
Dimethyl sulfone	Sulfonylbismethane	DMSN	C2H6SO2	(CH3)2SO2

29.6. References

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