APPLICATION NOTE

ACD/LABS [ADVANCED CHEMISTRY DEVELOPMENT, INC.]

Implementing Modern IUPAC Nomenclature

Andrei Erin
Jesse Harris

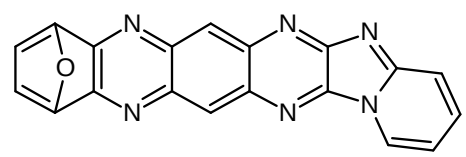
Keeping Up with Chemical Complexity

The growing complexity of chemical substances and corresponding advances in chemical nomenclature makes it practically impossible and potentially unsafe to rely on humans to create chemical names. Algorithmic name generation by computer programs long ago excelled human expertise in accuracy, productivity, and reliability of results. ACD/Name generates systematic names for diverse classes of chemical substances per IUPAC and IUBMB nomenclature recommendations. This application note explores the range of chemicals where Name can generate accurate names.

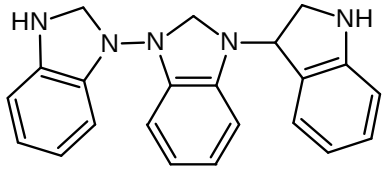
Organic Nomenclature

Name software can name structures with up to 1024 non-hydrogen atoms and 20 rings in polycyclic fragments. It treats most classes of organic compounds following IUPAC recommendations on Nomenclature of Organic Chemistry.[1-3]

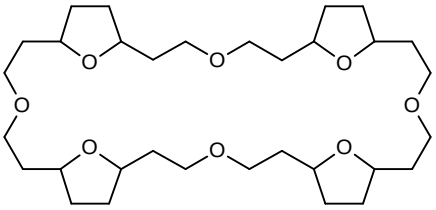
Name supports all nomenclature operations and treats practically all classes of organic structures according to IUPAC recommendations.[1] Most principal parent structures and their derivatives are recognized; except cyclic and linear phanes, fullerenes, and complex bridged and spiro ring systems.



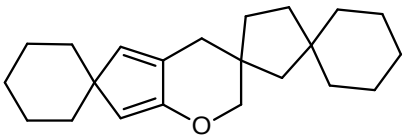
9,12-epoxypyrido[1'',2'':1',2']imidazo [4',5':5,6]pyrazino[2,3-*b*]phenazine



$1^{2},1^{3},3^{2},3^{3}$-tetrahydro-$2^{2}H$-$1^{1},2^{1}$:$2^{3},3^{1}$-terbenzimidazole



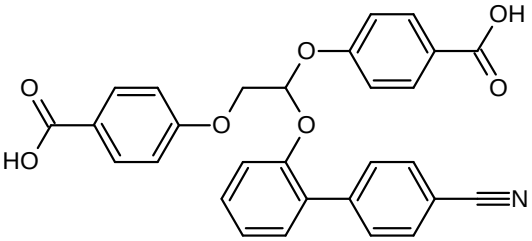
4,13,22,31,37,38,39,40-octaoxapentacyclo [32.2.1.$1^{7,10}$.$1^{16,19}$.$1^{25,28}$]tetracontane



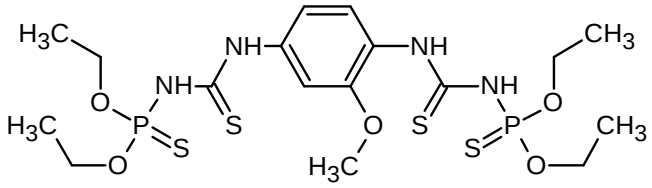
2''*H*,4''*H*-trispiro[cyclohexane-1,1'-cyclopentane-3',3''-cyclopenta[*b*]pyran-6'',1'''-cyclohexane]

Figure 1: Name is able to quickly and accurately number complex multi-ring systems.

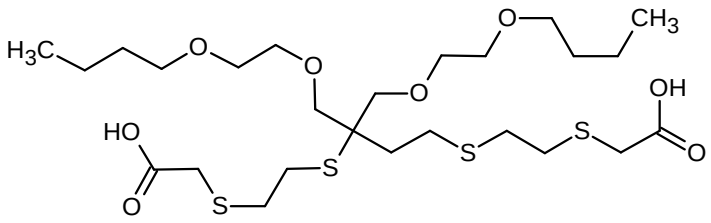
Name can also apply replacement and multiplicative nomenclature operations according to nomenclature recommendations, as shown in Figure 2.



4,4'-[{1-[(4'-cyano[1,1'-biphenyl]-2-yl)oxy]ethane-1,2-diyl}bis(oxy)]dibenzoic acid



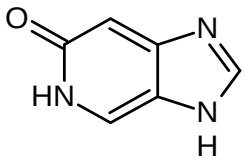
O,*O*,*O*',*O*'-tetraethyl *N*,*N*'-[(2-methoxy-1,4-phenylene)dicarbamothioyl]bis(phosphoramidothioate)



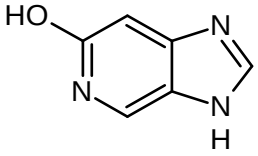
7,7-bis[(2-butoxyethoxy)methyl]-3,6,10,13-tetrathiapentadecane-1,15-dioic acid

Figure 2: Replacement and multiplicative nomenclature is used when naming chemicals with repeated complex side chains.

Name supports several other special operations like recognition of tautomers, specification of isotopic substitution, and naming of regular homopolymers. This is especially useful when naming complex aromatic structures, deuterated structures, or polymers, as seen in Figure 3.



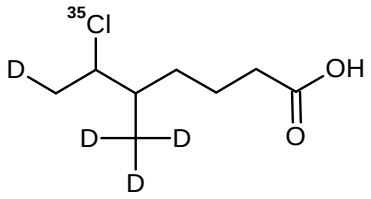
3,5-dihydro-6*H*-imidazo[4,5-*c*]pyridin-6-one



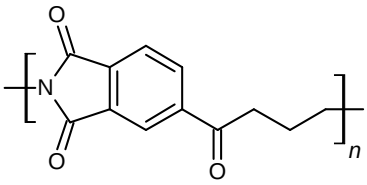
3*H*-imidazo[4,5-*c*]pyridin-6-ol

If tautomer recognition option is on, both are named:

1,5-dihydro-6*H*-imidazo[4,5-*c*]pyridin-6-one



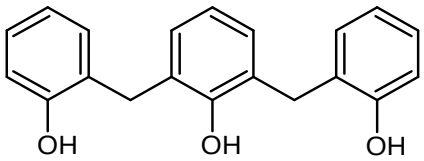
6-(^{35}Cl)chloro-5-($^{2}H_3$)methyl(7-$^{2}H_1$)heptanoic acid

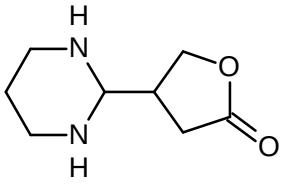


poly[(1,3-dioxo-1,3-dihydro-2*H*-isoindole-2,5-diyl)(1-oxobutane-1,4-diyl)]

Figure 3: Name can accurately name tautomeric, deuterated, and polymeric structures.

Name supports both old and new IUPAC organic nomenclature, allowing the generation of new and old names. This functionality helps when updating chemical names, including common synonyms for legal documentation, or generating structures based on outdated texts.





New name	2,2'-[(2-hydroxy-1,3-phenylene)bis(methylene)]diphenol	4-(1,3-diazinan-2-yl)oxolan-2-one
Old name	2,6-bis(2-hydroxybenzyl)phenol	4-(hexahydropyrimidin-2-yl)dihydrofuran-2(3*H*)-one

Figure 4: Chemical nomenclature changes based on updated recommendations from IUPAC. Name can generate new or old chemical names, depending on your preference.

Natural Parents and their Derivatives

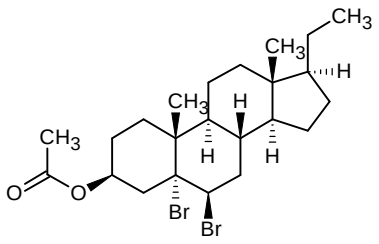
Name recognizes about 130 principle natural product parent structures following naming conventions based on the published IUPAC and IUBMB recommendations.

The following operations in the natural product parent structures are supported:

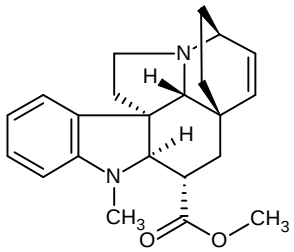
→ Use of retained numbering for each natural parent

- → Substitution, saturation/unsaturation and functional derivatives
- → Stereochemical modifications, (α/β, R/S, cis/trans, E/Z)
- → Substituent names for non-functional parent structures
- → Cyclo- and epoxy-modifications

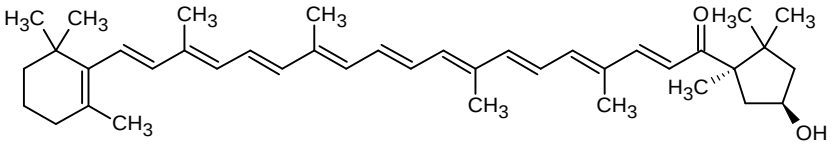
For example:



5,6β-dibromo-5α-pregnan-3β-yl acetate



methyl 1-methyl-8α-6,7-didehydro-8,21-cycloaspidospermidine-3α-carboxylate



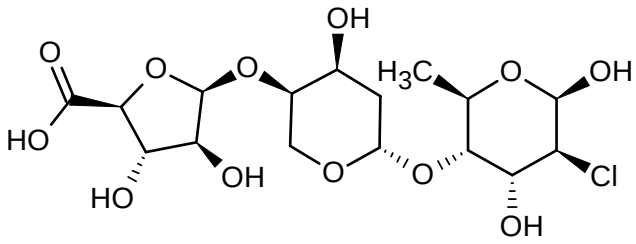
(3'*R*,5'*R*)-3'-hydroxy-β,κ-caroten-6'-one

Figure 5: Examples of natural products that can be named by Name.

Biochemical Nomenclature

Name recognizes names for several classes of biochemicals following IUPAC and IUBMB nomenclature recommendations.[4-6] The degree of development varies by class.

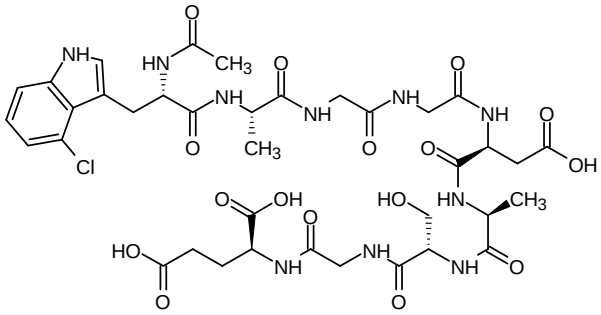
Carbohydrates: Most of the nomenclature procedures are recognized. Derivatives that are not supported include hydrazones, oximes, some cyclic acetals and carbohydrate acids.



Biochemical names enabled	β-D-arabinofuranuronosyl-(1→4)-2-deoxy-β-D-*erythro*-pentopyranosyl-(1→4)-2-chloro-2,6-dideoxy-β-D-altropyranose
Biochemical names disabled	(2S,3S,4S,5R)-5-{[(3R,4S,6S)-6-{[(2R,3S,4R,5S,6R)-5-chloro-4,6-dihydroxy-2-methyloxan-3-yl]oxy}-4-hydroxyoxan-3-yl]oxy}-3,4-dihydroxyoxolane-2-carboxylic acid

Figure 6: Example of carbohydrates included in a name. The biochemical naming setting can be turned off if you prefer unsimplified organic chemistry-based naming.

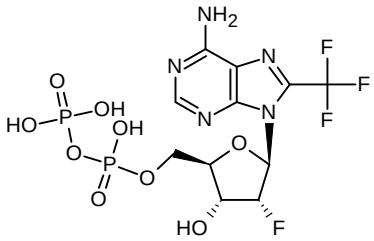
Amino acids: names for natural amino acids and their derivatives including peptides are available. All amino acids and short peptide sequences, including simple cyclic peptides, are named in accordance with the established principles. Amino acid derivatives that are not supported by Name will generate names in accordance with organic nomenclature and will include a notice that it is a non-preferred name.



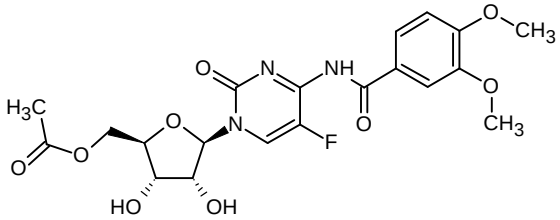
Biochemical names enabled	N-acetyl-4-chloro-L-tryptophyl-L-alanylglycylglycyl-L-α-aspartyl-L-alanyl-L-serylglycyl-L-glutamic acid
Biochemical names disabled	(3*S*,9*S*,12*S*,15*S*)-15-{(8*S*,11*S*)-11-[(4-chloro-1*H*-indol-3-yl)methyl]-8-methyl-4,7,10,13-tetraoxo-3,6,9,12-tetraazatetradecanan-1-amido}-9-(hydroxymethyl)-12-methyl-5,8,11,14-tetraoxo-4,7,10,13-tetraazahexadecane-1,3,16-tricarboxylic acid

Figure 7: Example of a peptide named with and without biochemical naming.

Standard nucleosides and their basic modifications are also recognized by Name.



2'-deoxy-2'-fluoro-8-(trifluoromethyl)adenosine 5'-(trihydrogen diphosphate)

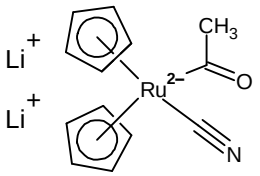


5'-*O*-acetyl-*N*-(3,4-dimethoxybenzoyl)-5-fluorocytidine

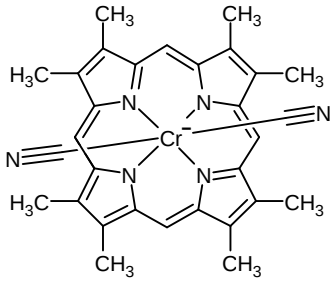
Figure 8: Examples of modified nucleosides that can be named by Name.

Inorganic Nomenclature

Name supports naming of various coordination and organometallic compounds according to recommendations from IUPAC.[7] Coordination sites are specified according to κ- and η-conventions. This includes the naming of complex coordination and organometallic structures, including π-complexes.



dilithium acetyl(cyano-κ*C*)bis(η^5-cyclopentadienyl)ruthenate(2-)



bis(cyano-κ*C*)[2,3,7,8,12,13,17,18-octamethylporphyrinato(2-)-κ*N*^{21},κ*N*^{22},κ*N*^{23},κ*N*^{24}] chromate(1-)

Figure 9: Examples of complex organometallic structures that can be named by Name.

Note that such structures still lack a universally agreed convention for digital representation. Name currently offers three methods to indicate coordination—common single bonds, coordination bonds, or a combination of single bonds with charges assuming standard valence states (Figure 10). Due to the complex nature of coordination and organometallic structures, the generated names should be carefully reviewed by the user.

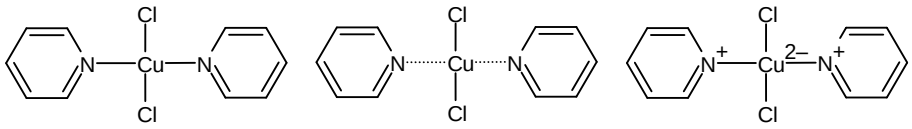
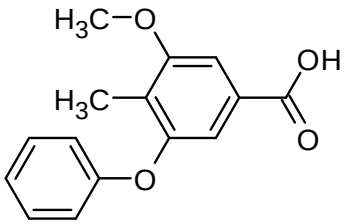


Figure 10: Users can display organometallic structures using single bonds, coordination bonds, or combinations of bonds and charges. Each representation will generate the name dichlorido[bis(pyridine)]copper.

Multilingual Naming

Name generates IUPAC names in over 20 languages, including English, French, Spanish, German, and Russian. For a complete, up-to-date list of supported languages, please refer to the Name product page on our website.



English	**EN**	3-methoxy-4-methyl-5-phenoxybenzoic acid
French	**FR**	acide 3-méthoxy-4-méthyl-5-phénoxybenzoïque
Spanish	**ES**	ácido 3-fenoxi-4-metil-5-metoxibenzoico
German	**DE**	3-Methoxy-4-methyl-5-phenoxybenzoesäure
Russian	**RU**	4-метил-3-метокси-5-феноксибензойная кислота
Swedish	**SV**	3-fenoxi-5-metoxi-4-metylbensoesyra

Figure 11: Name is able to generate names in over 20 languages. These names can be exported to word processing software.

Two letter designations given in the table are used in Name to indicate the language in which the specific name is generated. Name considers all main aspects of naming in the specific language:

- Different name grammar, e.g., compare French and Spanish to other names
- Different alphabetical order and locants for substituents
- Capitalization of the first letter for German names

Due to the lack of published recommendations in certain languages, Name uses the most recent English IUPAC recommendations and adapts them for other languages. As a result, in some instances the generated name may deviate somewhat from the local language recommendations.

Maintaining Excellence in Chemical Nomenclature

Chemical nomenclature is complex and always evolving. The IUPAC Blue Book 2013 increased the complexity of chemical nomenclature by introducing many additional criteria and operations. Name offers best-in-class chemical naming for a wide range of chemical structures and will continue to improve to meet the needs of chemists and expand the coverage of chemical space.

References

1 Favre, H.A., Powell, W.H. (2013). *Nomenclature of Organic Chemistry: IUPAC Recommendations and Preferred Names 2013*. The Royal Society of Chemistry, Cambridge.

2 International Union of Pure and Applied Chemistry, Klesney, S.P., Rigaudy, J. (1979). *Nomenclature of Organic Chemistry: Sections A, B, C, D, E, F and H*, Pergamon Press, Oxford.

3 International Union of Pure and Applied Chemistry, Panico, R. (1993). *A Guide to IUPAC Nomenclature of Organic Compounds: Recommendations 1993*. Blackwell Scientific Publications, Oxford.

4 IUPAC and IUBMB, IUPAC-IUB Joint Commission on Biochemical Nomenclature. (1989). The Nomenclature of Steroids (Recommendations 1989). *Eur. J. Biochem., 186*, 429–458.

5 IUPAC and IUBMB Joint Commission on Biochemical Nomenclature, McNaught, A.D. (1996). Nomenclature of Carbohydrates (IUPAC Recommendations 1996). *Pure and Applied Chemistry, 68*(10), 1919–2008.

6 IUPAC and IUB Joint Commission on Biochemical Nomenclature. (1984). Nomenclature and Symbolism for Amino Acids and Peptides (Recommendations 1983). *Pure Appl. Chem., 56*(5), 595-624.

7 IUPAC. (2005). *Nomenclature of Inorganic Chemistry: IUPAC Recommendations 2005*. (N.G. Connelly, T. Damhus, R.M. Hartshorn, A.T. Hutton, Eds.). Royal Society of Chemistry, Cambridge.