

# Metabolomics Software Tools

Craig Knox & Roman Eisner  
January 12, 2009

# DrugBank

- What is DrugBank?
  - Database of small molecule and biotech drugs and their protein enzymes/targets
  - Freely available structures/data
  - Linked to from KEGG, UniProt, Wikipedia, etc.

# DrugBank

[Home](#)[Browse](#)[Search](#)[About](#)[Downloads](#)[Contact Us](#)

Search:



The DrugBank database is a unique bioinformatics and cheminformatics resource that combines detailed drug (i.e. chemical, pharmacological and pharmaceutical) data with comprehensive drug target (i.e. sequence, structure, and pathway) information. The database contains nearly 4800 drug entries including >1,350 FDA-approved small molecule drugs, 123 FDA-approved biotech (protein/peptide) drugs, 71 nutraceuticals and >3,243 experimental drugs. Additionally, more than 2,500 non-redundant protein (i.e. drug target) sequences are linked to these FDA approved drug entries. Each DrugCard entry contains more than 100 data fields with half of the information being devoted to drug/chemical data and the other half devoted to drug target or protein data.

DrugBank is supported by [David Wishart](#), Departments of [Computing Science](#) & [Biological Sciences](#), [University of Alberta](#).

[More about DrugBank](#)

## What's New?

- 2 new fields have been added to the drug cards:
  - The predicted logS (via [ALOGPS](#)) is essentially the log units of molar solubility, and is one less field you will need to calculate.
  - The [InChIKey](#) has been added as a unique structural identifier. We will soon be adding the ability to link to a drug via the InChIKey.
- We have added a basic search interface for doing complex queries on specific fields. You can read more about it on the [Text Query](#). Or for a quick example, try [this query](#) (find all anti-depressive agents which are also approved)!
- The [downloads](#) page has been updated with files from release 2.5. You can now access files from the previous DrugBank releases as well, although keep in mind that the drug card format has changed.
- We have added a new SNP summary tool called SNPJam. This application is linked to by all targets and metabolizing enzymes. It provides a comprehensive view of the SNPs for a specific gene without being overwhelming.

[News archive](#)

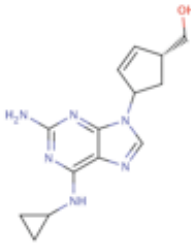
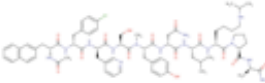
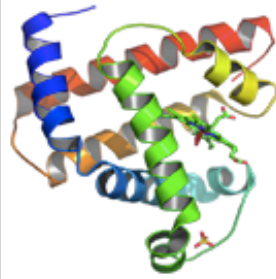
## Browsing drugs

All **Approved** Biotech Small Molecule Nutraceutical Experimental Withdrawn Illicit

Per Page: 10 | 25 | 50 | 100

Showing 1-10 out of 1492 drugs

previous **1** 2 3 4 5 6 7 8 9 10 11 ... 149 150 next

DrugBank ID	Name <span>▲</span>	Formula	Structure	Categories	Therapeutic Indication
	CAS Number	Weight			
DB01048 DRUGCARD	<b>Abacavir</b> 136470-78-5	$C_{14}H_{18}N_6O$ 286.3323		Anti-HIV Agents Nucleoside and Nucleotide Reverse Transcriptase Inhibitors Reverse Transcriptase Inhibitors	For the treatment of HIV-1 infection, in combination with other antiretroviral agents...
DB00106 DRUGCARD	<b>Abarelix</b> 183552-38-7	$C_{72}H_{95}ClN_{14}O_{14}$ 1416.0630		Anti-Testosterone Agents Antineoplastic Agents	For palliative treatment of advanced prostate cancer.
DB01281 DRUGCARD	<b>Abatacept</b> 332348-12-6	— 92000.0000		Antirheumatic Agents	For the second line reduction of the signs and symptoms of moderate-to-severe active...

# DrugBank 2.5

- Complete overhaul of code backend
  - Moved from flatfile to strictly database
  - Incorporated better structure validation
  - Native property prediction/conversion

# DrugBank 2.5

- Complete rewrite of user interface
  - Brand new data extractor (prototype, export to excel)
  - New search engine (KinoSearch) with built in advanced search interface

# Data Extractor

Legend: drug field target field enzyme field

Field Name	Value	Description
Absorption	rapid	Full or partial text (i.e. readily)
DrugBank ID		Full or partial text (i.e. DB00001 or DB05)
Generic Name		Full or partial text (i.e. acetaminophen or aceta)
Gene Name	CYP3A	Full or partial text (i.e. HIS1 or COX)
Chromosome Location	11	Full or partial text (i.e. 1 or X)
Name		Full or partial text (i.e. Histamine H1 Receptor or Hista)

**Drug Fields:**

- Absorption
- Organisms Affected
- AHFS Code
- ATC Code
- Biotransformation
- Brands
- Caco2 Permeability (experimental)
- CAS Number
- Category
- ChEBI ID
- Chemical Formula
- Drug Classification
- Description
- Dosage Form
- DPD ID (DIN)
- DrugBank ID
- Genbank ID
- Half Life
- HET ID
- InChI Identifier
- InChI Key
- Indication
- Drug-Drug Interaction
- Drug-Food Interaction
- Chemical IUPAC Name
- KEGG Compound ID
- KEGG Drug ID

Drug type: Approved Query type:  AND  OR Output: HTML Reset Submit

# Data Extractor: Results

Your search returned 14 results

Legend:   drug field   target field   enzyme field

DrugBank ID	Generic Name	Absorption	Gene Name	Chromosome Location	Name
DB00185 <span style="border: 1px solid black; border-radius: 50%; padding: 2px;">DRUGCARD</span>	Cevimeline	Rapidly absorbed with peak concentration after 1.5 to 2 hours	CYP3A4	11	Muscarinic acetylcholine receptor M1
DB00332 <span style="border: 1px solid black; border-radius: 50%; padding: 2px;">DRUGCARD</span>	Ipratropium	Inhalation (local)-minimal; Nasal-rapid and minimal	CYP3A4	11	Muscarinic acetylcholine receptor M1
DB00370 <span style="border: 1px solid black; border-radius: 50%; padding: 2px;">DRUGCARD</span>	Mirtazapine	Rapid and complete, but, due to first-pass metabolism, absolute bioavailability is 50%.	CYP3A4	11	5-hydroxytryptamine 3 receptor
DB00433 <span style="border: 1px solid black; border-radius: 50%; padding: 2px;">DRUGCARD</span>	Prochlorperazine	Rapidly absorbed following oral administration	CYP3A4	11	D(2) dopamine receptor
DB00490 <span style="border: 1px solid black; border-radius: 50%; padding: 2px;">DRUGCARD</span>	Buspirone	Rapidly absorbed in man. Bioavailability is low and variable (approximately 5%) due to extensive first pass metabolism.	CYP3A4	11	D(2) dopamine receptor
DB00603 <span style="border: 1px solid black; border-radius: 50%; padding: 2px;">DRUGCARD</span>	Medroxyprogesterone	Rapidly absorbed from GI tract	CYP3A4	11	Progesterone receptor
DB00604 <span style="border: 1px solid black; border-radius: 50%; padding: 2px;">DRUGCARD</span>	Cisapride	Cisapride is rapidly absorbed after oral administration, with an absolute bioavailability of 35-40%.	CYP3A4	11	5-hydroxytryptamine 3 receptor
DB00731 <span style="border: 1px solid black; border-radius: 50%; padding: 2px;">DRUGCARD</span>	Nateglinide	Rapidly absorbed following oral administration prior to a meal, absolute bioavailability is estimated to be approximately 73%.	CYP3A4	11	ATP-binding cassette transporter sub-family C member 8
DB00731	Nateglinide	Rapidly absorbed following oral	CYP3A4	11	ATP-sensitive

# HMDB

- Converted HMDB to a fork of the DrugBank codebase for 2.0 release
- Harder as back end database (MetaboLIMS) is not normalized + no format validation

# Structure Search

- Rails/jRuby based structure search for DrugBank and HMDB
- Uses ChemAxon software to implement the backend structure database

# DrugBank Future

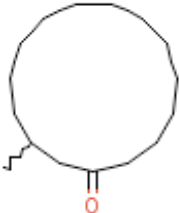
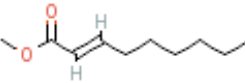
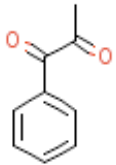
- DrugBank SDMS - Ruby on Rails (ROR)
- DrugBank 3.0 - convert website to ROR
- RESTful API (common request) for free
- DrugBank becomes a true web application
- Let users provide feedback/corrections

# FooDB

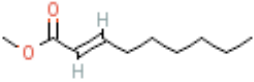
- What is the FooDB?
  - Food component/additive database of 2000 compounds
  - Automatically annotated via BioSpider
  - Ready for manual annotation



## Listing food components

ID ▼	Name	CAS	Formula Weight	Structure	Show
1	3-Methylcyclopentadecanone	541-91-3	$C_{16}H_{30}O$ 238.409		<a href="#">Show</a>
2	2-Nonenoic acid, methyl ester	111-79-5	$C_{10}H_{18}O_2$ 170.249		<a href="#">Show</a>
3	1-phenyl-1,2-propanedione	579-07-7	$C_9H_8O_2$ 148.159		<a href="#">Show</a>

[Show Similar Compounds](#)**Summary**

<b>FooDB ID</b>	2
<b>Common Name</b>	<b>2-Nonenoic acid, methyl ester</b>
<b>Structure</b>	 <p>The image shows the chemical structure of methyl (2E)-2-nonenoate. It consists of a methyl ester group (CH<sub>3</sub>-O-C(=O)-) attached to a double bond at the 2-position of a nine-carbon chain. The double bond has a methyl group and a hydrogen atom on one carbon, and a hydrogen atom and a seven-carbon alkyl chain on the other carbon. The methyl group and the seven-carbon chain are on opposite sides of the double bond, indicating an E configuration.</p>
<b>Created</b>	Mon Jan 22 22:20:24 -0700 2007
<b>Updated</b>	Mon Jan 22 22:20:25 -0700 2007
<b>Description</b>	Not Available
<b>Synonyms</b>	<ol style="list-style-type: none"><li>1. 2-Nonenoic acid, methyl ester, (E)-</li><li>2. FEMA No. 2725</li><li>3. Methyl (2E)-2-nonenoate</li><li>4. Methyl 2-nonenoate</li><li>5. methyl 2-nonenoate (ACD/Name 4.0)</li><li>6. Methyl non-2-enoate</li><li>7. Methyl nonenoate</li><li>8. Methyl nonylenate</li><li>9. Neofolione</li><li>10. Nonenoic acid, methyl ester</li></ol>
<b>CAS Number</b>	111-79-5
<b>Average Molecular Weight</b>	170.249

Molecular Weight  
Average

170.249

# FooDB

- Ruby on Rails backend
- Full text search via ferret (Ruby lucene)
- Built in web-API (REST)

# FooDB Future

- Lots of manual work to do
- Addition of structure, sequence, extraction tools
- Lots of interest and potential

# HMDBB Statistics

- Statistics for
  - issues
  - reports
- Dynamically generated from metaboLIMS
- Some interactivity



human metabolome project

# MetaboLIMS

WebTools for the Human Metabolome Project



GenomeCanada



GenomeAlberta

[Home](#)

[Administration](#)

[Minutes](#)

[Planning](#)

[SOPs](#)

[Bulletin](#)

[Summary](#)

[Search](#)

[Logout](#)

## HMP Document Browser

[Forms/Posters](#)

[Contacts](#)

[Searches](#)

[Mass file](#)

[Stats](#)

[Safety Check](#)

[BioSpider](#)

[HMDB](#)

## HMP Data Browser

[-Metabolite Database-](#)

[Metabolite Library](#)

[BioSamples](#)

[Identified Metabolites](#)

[Users](#)

[Audit Trail](#)

### [Enter New Metabolite](#)

Navigation: First, Prev, Next, Last, 150 Rows Displayed, Export XLS

8,624 results found, displaying 1 to 150

[Search](#) [Clear](#)

HMDB ID	Names	Chemical Formula	Molecular Weight (Da)	CAS	Export to HMDB
<a href="#">HMDB00001</a>	1 Methylhistidine; 1 methylhistidine; 1 methyl histidine; 1 MHIs; 1 Methyl Histidine; 1 Methyl L histidine; 1 N Methyl L histidine; L 1 Methylhistidine; N1 Methyl L histidine; pl methylhistidine	C7H11N3O2	169.08513	332-80-9	Yes
<a href="#">HMDB00002</a>	1,3 Diaminopropane; 1,3 Diamino n propane; 1,3 Propylenediamine; 1,3 Trimethylenediamine; 1,3 propanediamine; 3 Aminoproplamine; Propane 1,3 diamine;	C3H10N2	74.08440	109-76-2	Yes

Stat #		Statistic Description	Statistic Value
1	<a href="#">View Data</a>	Total Compounds in database (Including "Deleted Compounds" in MetaboLIMS)	10315
2	<a href="#">View Data</a>	Total Compounds not deleted in MetaboLIMS	8624
3	<a href="#">View Data</a>	<b>HMDB Compounds</b>	7231
4	<a href="#">View Data</a>	HMDB Compounds that have been purchased and have no NMR and no HSQC2D assigned spectrum	203
5	<a href="#">View Data</a>	HMDB Compounds that have been purchased and have no NMR and no MS and no HSQC2D assigned spectrum	70
6	<a href="#">View Data</a>	HMDB Compounds that have not been purchased	6266
7	<a href="#">View Data</a>	HMDB Compounds that have not been purchased and have an MSDS	81
8	<a href="#">View Data</a>	Compounds in the HML	965
9	<a href="#">View Data</a>	HML Compounds that have no HNMR assigned spectrum	211
10	<a href="#">View Data</a>	HML Compounds that have no CNMR assigned spectrum	856
11	<a href="#">View Data</a>	HML Compounds that have no HSQC2D assigned spectrum	220
12	<a href="#">View Data</a>	HML Compounds that have no MS Spectrum	190
13	<a href="#">View Data</a>	Synthesized Compounds	57
14	<a href="#">View Data</a>	Purchased Compounds	911
15	<a href="#">View Data</a>	CNMR Assigned Spectra	120
16	<a href="#">View Data</a>	HNMR-Sykes Assigned Spectra	136

# HMDB Sanity Checks

- Program for flagging issues in the HMDB
- Run nightly
- Looks at HMDB not metaboLIMS



human metabolome project

# MetaboLIMS

WebTools for the Human Metabolome Project



GenomeCanada



GenomeAlberta

[Home](#)

[Administration](#)

[Minutes](#)

[Planning](#)

[SOPs](#)

[Bulletin](#)

[Summary](#)

[Search](#)

[Logout](#)

## HMP Document Browser

[Forms/Posters](#)

[Contacts](#)

[Searches](#)

[Mass Calc](#)

[Stats](#)

[Sanity Check](#)

[BioSpider](#)

[HMDB](#)

## HMP Data Browser

[-Metabolite Database-](#)

[Metabolite Library](#)

[BioSamples](#)

[Identified Metabolites](#)

[Users](#)

[Audit Trail](#)

### [Enter New Metabolite](#)

Navigation: First, Prev, Next, Last, 150 Rows Displayed, Export XLS

8,624 results found, displaying 1 to 150

[Search](#) [Clear](#)

HMDB ID	Names	Chemical Formula	Molecular Weight (Da)	CAS	Export to HMDB
<a href="#">HMDB00001</a>	1 Methylhistidine; 1 methylhistidine; 1 methyl histidine; 1 MHIs; 1 Methyl Histidine; 1 Methyl L histidine; 1 N Methyl L histidine; L 1 Methylhistidine; N1 Methyl L histidine; pl methylhistidine	C7H11N3O2	169.08513	332-80-9	Yes
<a href="#">HMDB00002</a>	1,3 Diaminopropane; 1,3 Diamino n propane; 1,3 Propylenediamine; 1,3 Trimethylenediamine; 1,3 propanediamine; 3 Aminoproplamine; Propane 1,3 diamine;	C3H10N2	74.08440	109-76-2	Yes

No reference found for the following experimental water solubilities. (References should be written in square brackets [])

-----  
HMDB02111

Value but no unit for the following concentrations (Number of Cases: 2)

-----  
HMDB00098: Concentration: 2867.1

HMDB00098: Concentration: 483.0

The following Synonyms have invalid formatting (Number of Cases: 3)

-----  
HMDB03156 (Synonym: " 2,3-Butanediol, isomer 1")

HMDB03156 (Synonym: " 2,3-Butanediol, isomer 2")

HMDB03156 (Synonym: " 2.3-Butanediol, isomer 3")

The following groups of metabolites have the same IUPAC Name (Number of Cases: 22)

-----  
001.) (8z,11z,13e,15s)-15-hydroxyicosa-8,11,13-trienoic acid: ['HMDB05045', 'HMDB10207']  
002.) 1'-[1-(9z,12z-octadecadienoyl)-2-(9z-octadecenoyl)-sn-glycero-3-phospho],3'-[1,2-di(11z-octadecenoyl)-rac-glycero-3-phospho]-glycerol: ['HMDB01193', 'HMDB11135']  
003.) 1'-[1,2-di(9z-octadecenoyl)-rac-glycero-3-phospho],3'-[1,2-di(9z-octadecenoyl)-rac-glycero-3-phospho]-glycerol: ['HMDB01193', 'HMDB11135']  
004.) 1'-[1-(4z,7z,10z,13z,16z,19z-docosahexaenoyl)-2-(9z,12z-octadecadienoyl)-sn-glycero-3-phospho],3'-[1,2-di(11z-octadecenoyl)-rac-glycero-3-phospho]-glycerol: ['HMDB01193', 'HMDB11135']  
005.) 1'-[1-(4z,7z,10z,13z,16z,19z-docosahexaenoyl)-2-(9z,12z-octadecadienoyl)-sn-glycero-3-phospho],3'-[1,2-di(9z,12z-octadecadienoyl)-rac-glycero-3-phospho]-glycerol: ['HMDB01193', 'HMDB11135']  
006.) (z)-7-[(1r,2r,3s,5s)-3,5-dihydroxy-2-[(e,3s)-3-hydroxyoct-1-enyl]cyclopentyl]hept-5-enoic acid: ['HMDB10199', 'HMDB10207']  
007.) 1'-[1-(4z,7z,10z,13z,16z,19z-docosahexaenoyl)-2-(5z,8z,11z-eicosatrienoyl)-sn-glycero-3-phospho],3'-[1,2-di(9z,12z-octadecadienoyl)-rac-glycero-3-phospho]-glycerol: ['HMDB01193', 'HMDB11135']  
008.) 1'-[1-(4z,7z,10z,13z,16z,19z-docosahexaenoyl)-2-(9z,12z-octadecadienoyl)-sn-glycero-3-phospho],3'-[1,2-di(9z-octadecenoyl)-rac-glycero-3-phospho]-glycerol: ['HMDB01193', 'HMDB11135']  
009.) 1'-[1-(5z,8z,11z,14z-eicosatetraenoyl)-2-(5z,8z,11z-eicosatrienoyl)-sn-glycero-3-phospho],3'-[1,2-di(9z,12z-octadecadienoyl)-rac-glycero-3-phospho]-glycerol: ['HMDB01193', 'HMDB11135']  
010.) 1'-[1-(4z,7z,10z,13z,16z,19z-docosahexaenoyl)-2-(8z,11z,14z-eicosatrienoyl)-sn-glycero-3-phospho],3'-[1,2-di(9z,12z-octadecadienoyl)-rac-glycero-3-phospho]-glycerol: ['HMDB01193', 'HMDB11135']  
011.) 1'-[1-(4z,7z,10z,13z,16z-docosapentaenoyl)-2-(9z,12z-octadecadienoyl)-sn-glycero-3-phospho],3'-[1,2-di(9z,12z-octadecadienoyl)-rac-glycero-3-phospho]-glycerol: ['HMDB01193', 'HMDB11135']  
012.) 1'-[1,2-di(9z,12z-octadecadienoyl)-rac-glycero-3-phospho],3'-[1,2-di(11z-octadecenoyl)-rac-glycero-3-phospho]-glycerol: ['HMDB01193', 'HMDB11135']  
013.) 1'-[1,2-di(9z,12z-octadecadienoyl)-rac-glycero-3-phospho],3'-[1,2-di(9z,12z-octadecadienoyl)-rac-glycero-3-phospho]-glycerol: ['HMDB01193', 'HMDB11135']  
014.) 1'-[1,2-di(11z-octadecenoyl)-rac-glycero-3-phospho],3'-[1,2-di(11z-octadecenoyl)-rac-glycero-3-phospho]-glycerol: ['HMDB01193', 'HMDB11135']  
015.) 1'-[1-(11z,14z-eicosadienoyl)-2-(9z,12z-octadecadienoyl)-sn-glycero-3-phospho],3'-[1,2-di(9z,12z-octadecadienoyl)-rac-glycero-3-phospho]-glycerol: ['HMDB01193', 'HMDB11135']  
016.) 1'-[1-(9z,12z-octadecadienoyl)-2-(11z-octadecenoyl)-sn-glycero-3-phospho],3'-[1,2-di(9z-octadecenoyl)-rac-glycero-3-phospho]-glycerol: ['HMDB01193', 'HMDB11135']  
017.) 1'-[1-(11-eicosenoyl)-2-(9z,12z-octadecadienoyl)-sn-glycero-3-phospho],3'-[1,2-di(11z-octadecenoyl)-rac-glycero-3-phospho]-glycerol: ['HMDB01193', 'HMDB11135']  
018.) 1'-[1-(5z,8z,11z,14z-eicosatetraenoyl)-2-(8z,11z,14z-eicosatrienoyl)-sn-glycero-3-phospho],3'-[1,2-di(9z,12z-octadecadienoyl)-rac-glycero-3-phospho]-glycerol: ['HMDB01193', 'HMDB11135']  
019.) (5s,6e,8z,11z,14z)-5-hydroperoxyicosa-6,8,11,14-tetraenoic acid: ['HMDB01193', 'HMDB11135']

# Lipids Annotation

- Rule-based annotation
- Generates:
  - IUPAC Name
  - Synonyms
  - Descriptions
- Based on type (e.g. DG, TG, PE, ...) and fatty acids

Accession Number	HMDB05359
Common Name	<b>TG(16:0/16:0/16:1(9Z))[iso3]</b>
Description	<p>TG(16:0/16:0/16:1(9Z))[iso3] is a dipalmitic acid triglyceride. Triglycerides (TGs) are also known as triacylglycerols or triacylglycerides, meaning that they are glycerides in which the glycerol is esterified with three fatty acid groups (i.e. fatty acid trimesters of glycerol). TGs may be divided into three general types with respect to their acyl substituents. They are simple or monoacid if they contain only one type of fatty acid, diacid if they contain two types of fatty acids and triacid if three different acyl groups. Chain lengths of the fatty acids in naturally occurring triglycerides can be of varying lengths and saturations but 16, 18 and 20 carbons are the most common. TG(16:0/16:0/16:1(9Z))[iso3], in particular, consists of one chain of palmitic acid at the C-1 position, one chain of palmitic acid at the C-2 position and one chain of palmitoleic acid at the C-3 position. TGs are the main constituent of vegetable oil and animal fats. TGs are major components of very low density lipoprotein (VLDL) and chylomicrons, play an important role in metabolism as energy sources and transporters of dietary fat. They contain more than twice the energy (9 kcal/g) of carbohydrates and proteins. In the intestine, triglycerides are split into glycerol and fatty acids (this process is called lipolysis) with the help of lipases and bile secretions, which can then move into blood vessels. The triglycerides are rebuilt in the blood from their fragments and become constituents of lipoproteins, which deliver the fatty acids to and from fat cells among other functions. Various tissues can release the free fatty acids and take them up as a source of energy. Fat cells can synthesize and store triglycerides. When the body requires fatty acids as an energy source, the hormone glucagon signals the breakdown of the triglycerides by hormone-sensitive lipase to release free fatty acids. As the brain cannot utilize fatty acids as an energy source, the glycerol component of triglycerides can be converted into glucose for brain fuel when it is broken down. (<a href="http://www.cyberlipid.org">www.cyberlipid.org</a>, <a href="http://www.wikipedia.org">www.wikipedia.org</a>)</p>
Synonyms	<ol style="list-style-type: none"> <li>1. TG(16:0/16:0/16:1)</li> <li>2. TAG(16:0/16:0/16:1)</li> <li>3. Tracylglycerol(16:0/16:0/16:1)</li> <li>4. TG(48:1)</li> <li>5. TAG(48:1)</li> <li>6. Tracylglycerol(48:1)</li> <li>7. TG(16:0/16:0/16:1n7)</li> <li>8. TAG(16:0/16:0/16:1n7)</li> <li>9. Tracylglycerol(16:0/16:0/16:1n7)</li> <li>10. TG(16:0/16:0/16:1w7)</li> <li>11. TAG(16:0/16:0/16:1w7)</li> <li>12. Tracylglycerol(16:0/16:0/16:1w7)</li> <li>13. 1-palmitoyl-2-palmitoyl-3-palmitoleoyl-glycerol</li> <li>14. Triglyceride</li> <li>15. Triacylglycerol</li> </ol>
Chemical IUPAC Name	1-hexadecanoyl-2-hexadecanoyl-3-(9Z-hexadecenoyl)-glycerol

# Questions

- Thanks!