

Stefan Schulz

Medical Informatics
Research Group
University
Medical Center

Freiburg, Germany

2nd CHEBI User Group Workshop 2010
23-24 June 2010
EMBL-EBI, Hinxton, Cambridge, CB10 1SD, UK



Representation of Chemicals in Biomedical Terminologies



**UNIVERSITÄTS
FREIBURG KLINIKUM**



Purpose of this talk

- To give an overview of sources of chemicals in biomedical terminologies based on the UMLS
- To estimate their coverage related to ChEBI
- To analyze the ontological representation in the sources
- To discuss cross mapping with ChEBI

Overview of UMLS

Unified Medical Language System (UMLS)

- Metathesaurus
 - Very large, multi-purpose and multi-lingual vocabulary database (158 sources)
 - information about biomedical concepts (2M), their various names (8M), and relationships among them (41M)
 - IP restrictions apply
- Semantic Network
 - Semantic Types, that provide a consistent categorization of all concepts represented in the UMLS Metathesaurus

UMLS terms and concepts

C0000287	CZE	P	L6770587	PF	S7862131	2-hydroxy-5-nitrobenzylbromid	3	
C0000287	ENG	P	L0000287	PF	S0008061	2-Hydroxy-5-nitrobenzyl Bromide	0	
C0000287	ENG	P	L0000287	VO	S0007885	2 Hydroxy 5 nitrobenzyl Bromide	0	
C0000287	ENG	S	L0022780	PF	S0055692	Koshland's Reagent I	0	
C0000287	ENG	S	L0022780	VO	S0055691	Koshland Reagent I	0	
C0000287	ENG	S	L0022780	VO	S0055694	Koshlands Reagent I	0	
C0000287	ENG	S	L0022780	VW	S0080181	Reagent I, Koshland's	0	
C0000287	ENG	S	L0309506	PF	S0055693	Koshlands Reagent	0	
C0000287	ENG	S	L0309506	VO	S0055690	Koshland Reagent	0	
C0000287	ENG	S	L0309506	VO	S0080187	Reagent, Koshland	0	
C0000287	ENG	S	L0309506	VW	S0080188	Reagent, Koshlands	0	
C0000287	ENG	S	L0359802	PF	S0504134	Phenol, 2-(bromomethyl)-4-nitro-	0	
C0000287	ENG	S	L7671184	PF	S8865410	2-Hydroxy-5-nitrobenzyl Bromide [Chemical/Ingredient]	1	
C0000287	ENG	s	L6520804	PF	S7598104	KOSHLANDS REAGENT 01	0	
C0000287	ENG	s	L6524599	PF	S7596787	HYDROXYNITROBENZYL BROMIDE 02 05	0	
C0000287	FIN	P	L1507134	PF	S1803043	2-hydroksi-5-nitrobentsylibromidi	3	
C0000287	FRE	P	L3249939	PF	S3777562	Bromure 2-hydroxy-5-nitrobenzyl	3	
C0000287	FRE	S	L3245113	PF	S3772614	2-hydroxy-5-nitrobenzyl, bromure	3	
C0000287	GER	P	L1226332	PF	S1468278	2-Hydroxy-5-Nitrobenzylbromid	3	
C0000287	GER	S	L1787712	PF	S2084853	Koshland-Reagens I	3	
C0000287	GER	s	L8590862	PF	S10687072	HYDROXYNITROBENZYL BROMID 02 05	3	
C0000287	GER	s	L8590903	PF	S10687407	KOSHLAND REAGENS 01	3	
C0000287	ITA	P	L2136502	PF	S2474724	2-Idrossi-5-nitrobenzil bromuro	3	
C0000287	POR	P	L3290666	PF	S3818171	2-Hidroxi-5-nitrobenzil Brometo	3	
C0000287	POR	S	L3324426	PF	S3852791	Reagente de Koshland I	3	
C0000287	SPA	P	L3379007	PF	S3906512	2-Hidroxi-5-nitrobencil Bromuro	3	
C0000287	SPA	S	L3410013	PF	S3937780	Reactivo de Koshland I	3	
C0000287	SWE	P	L3419091	PF	S3946592	2-hydroxi-5-nitrobensylbromid	3	

UMLS relations

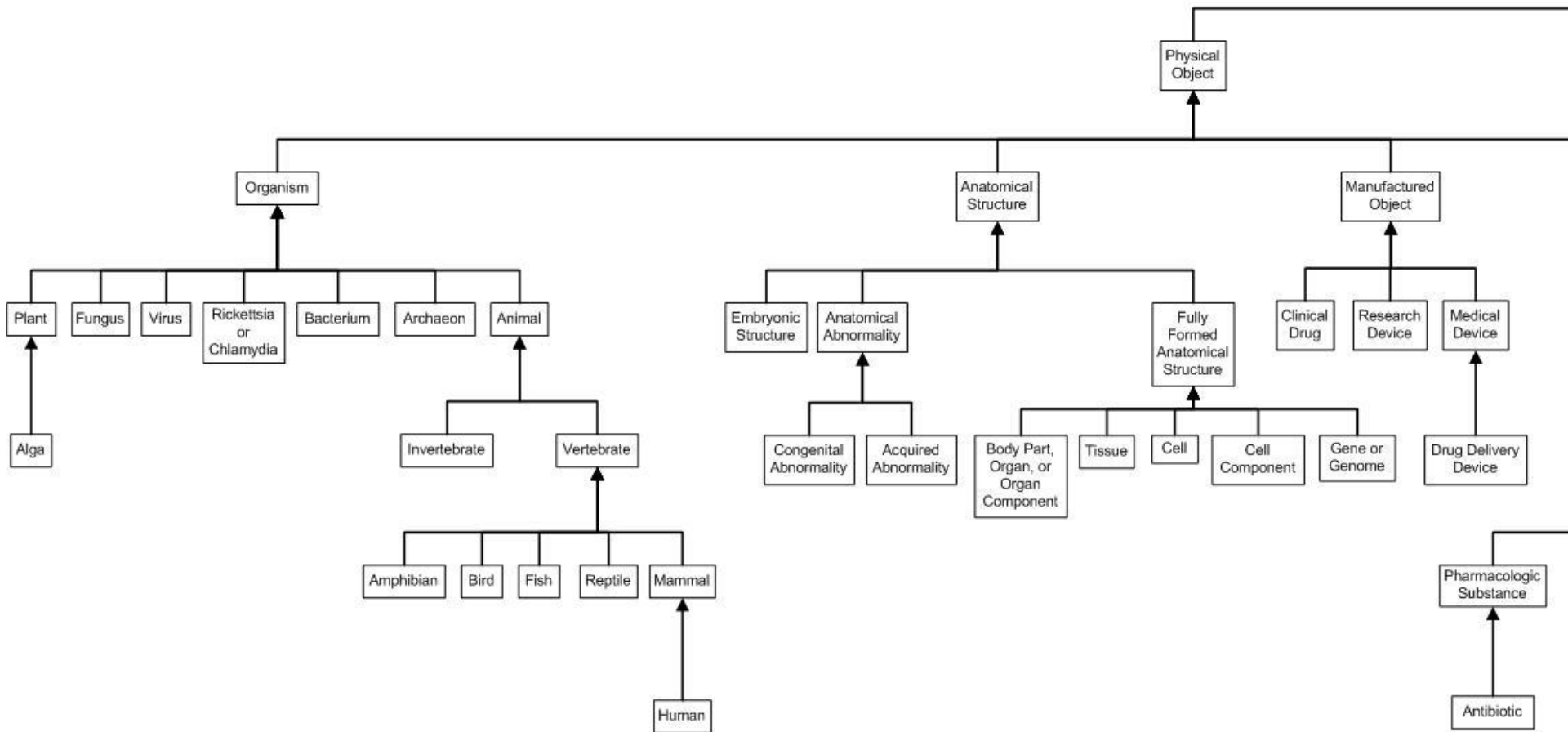
```
C0000726|CHD|C0041638||MSHSPA|MSHSPA||
C0000726|CHD|C0041638||MSHSWE|MSHSWE||
C0000726|CHD|C0041638||MSH|MSH||
C0000726|CHD|C0041638||SNMI|SNMI||
C0000726|CHD|C0151653||CST|CST||
C0000726|CHD|C0151705||CST|CST||
C0000726|CHD|C0225222|isa|SCTSPA|SCTSPA||
C0000726|CHD|C0225222|isa|SNOMEDCT|SNOMEDCT||
C0000726|CHD|C0225222||RCD|RCD||
C0000726|CHD|C0226727|isa|SCTSPA|SCTSPA||
C0000726|CHD|C0226727|isa|SNOMEDCT|SNOMEDCT||
C0000726|CHD|C0227345|isa|SCTSPA|SCTSPA||
C0000726|CHD|C0227345|isa|SNOMEDCT|SNOMEDCT||
C0000726|CHD|C0227613|part_of|UWDA|UWDA||
C0000726|CHD|C0227614|part_of|UWDA|UWDA||
C0000726|CHD|C0227667|part_of|UWDA|UWDA||
C0000726|CHD|C0227668|part_of|UWDA|UWDA||
C0000726|CHD|C0228904|isa|SCTSPA|SCTSPA||
C0000726|CHD|C0228904|isa|SNOMEDCT|SNOMEDCT||
C0000726|CHD|C0228905|isa|SCTSPA|SCTSPA||
C0000726|CHD|C0228905|isa|SNOMEDCT|SNOMEDCT||
C0000726|CHD|C0230165||SNMI|SNMI||
C0000726|CHD|C0230166|isa|SCTSPA|SCTSPA||
C0000726|CHD|C0230166|isa|SNOMEDCT|SNOMEDCT||
C0000726|CHD|C0230166||SNMI|SNMI||
C0000726|CHD|C0230167||SNMI|SNMI||
C0000726|CHD|C0230168|isa|SCTSPA|SCTSPA||
C0000726|CHD|C0230168|isa|SNOMEDCT|SNOMEDCT||
C0000726|CHD|C0230168|part_of|UWDA|UWDA||
C0000726|CHD|C0230168||RCD|RCD||
```

Relations

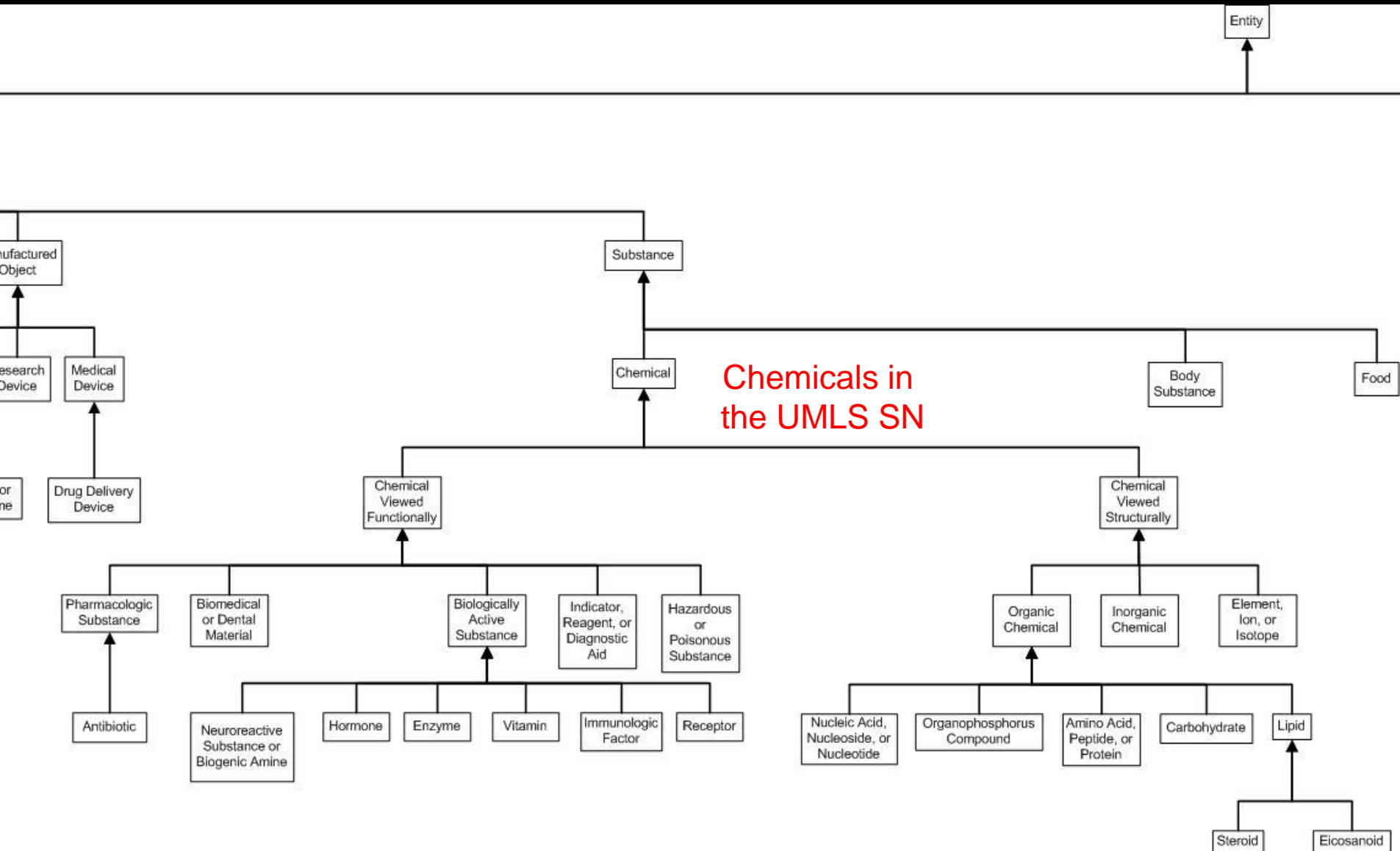
- preserved from their sources
- Thesaurus style relations (CHD / PAR)
- More precise relations (relationship attribute) i.e. part-of, is-a

UMLS Semantic Network

The Semantic Network of the UMLS



UMLS Semantic Network



Semantic Labeling of UMLS concepts

4-Hydroxyphenylpyruvate Dioxygenase	Amino Acid, Peptide, or Protein
4-Hydroxyphenylpyruvate Dioxygenase	Enzyme
4-Nitroquinoline-1-oxide	Organic Chemical
4-Nitroquinoline-1-oxide	Hazardous or Poisonous Substance
5 beta-Dihydrotestosterone	Steroid
5 beta-Dihydrotestosterone	Pharmacologic Substance
5'-NUCLEOTIDASE	Amino Acid, Peptide, or Protein
5'-NUCLEOTIDASE	Enzyme
5'-NUCLEOTIDASE	Immunologic Factor
5,12-diHETE	Eicosanoid
5,6-Dihydroxytryptamine	Organic Chemical
5,6-Dihydroxytryptamine	Pharmacologic Substance
5,7-Dihydroxytryptamine	Organic Chemical
5,7-Dihydroxytryptamine	Pharmacologic Substance
Eicosapentaenoic Acid	Lipid
Eicosapentaenoic Acid	Pharmacologic Substance
Eicosapentaenoic Acid	Biologically Active Substance
5,8,11,14-Eicosatetraynoic Acid	Eicosanoid
5,8,11,14-Eicosatetraynoic Acid	Pharmacologic Substance
Androstane-3,17-diol	Steroid
Androstane-3,17-diol	Hormone
5-Fluoro-2'-deoxyuridine Phosphorylase	Amino Acid, Peptide, or Protein
5-Fluoro-2'-deoxyuridine Phosphorylase	Enzyme
5-Hydroxytryptophan	Amino Acid, Peptide, or Protein
5-Hydroxytryptophan	Pharmacologic Substance

Semantic labeling

- Done by the NLM
- Each UMLS concept is assigned to one or more semantic types

Chemicals in UMLS and its sources

Semantic Network types for chemicals:

- T103|Chemical
- T104|Chemical Viewed Structurally
- T109|Organic Chemical
- T110|Steroid
- T111|Eicosanoid
- T114|Nucleic Acid, Nucleoside, or Nucleotide
- T115|Organophosphorus Compound
- T116|Amino Acid, Peptide, or Protein
- T118|Carbohydrate
- T119|Lipid
- T120|Chemical Viewed Functionally
- T121|Pharmacologic Substance
- T122|Biomedical or Dental Material
- T123|Biologically Active Substance
- T124|Neuroreactive Substance or Biogenic Amine
- T125|Hormone
- T126|Enzyme
- T195|Antibiotic
- T192|Receptor
- T127|Vitamin
- T129|Immunologic Factor
- T130|Indicator, Reagent, or Diagnostic Aid
- T131|Hazardous or Poisonous Substance
- T196|Element, Ion, or Isotope
- T197|Inorganic Chemical
- T200|Clinical Drug

Chemicals in UMLS source vocabularies

Source	Size	Chemicals (Broad)	Chemicals (Narrow)	Overlap with MeSH (%)	Purpose
Medical Subject Headings (MeSH)	296,338	266,927	220,228	100.0%	Biomedical Literature
LOINC	114,351	24,859	22,900	11.4%	Laboratory Medicine
SNOMED CT	317,177	26,659	22,727	35.1%	Health Records
Clinical Terms Version 3	181,192	20,685	20,673	23.1%	Health Records
Multum MediSource Lexicon	52,851	20,152	20,146	24.1%	Health Records
NCI Thesaurus	67,803	15,917	14,886	58.0%	Research
SNOMED International	112,712	15,878	13,178	41.9%	Health Records
National Drug File - Reference Terminology	39,163	13,786	11,969	66.8%	Health Records
UMLS Metathesaurus	120,458	15,285	11,240	61.3%	Biomedical Literature
National Drug Data File Plus Source Vocabulary	31,141	10,959	10,903	46.1%	Health Records
RXNORM	186,066	10,795	10,430	62.2%	Health Records
Veterans Health Administration National Drug File	24,913	6,991	6,984	44.5%	Health Records
MEDCIN	269,443	6,325	6,323	38.7%	Health Records
Physician Data Query	10,642	4,864	4,839	55.0%	Health Records
CRISP Thesaurus	16,682	5,045	4,064	75.5%	Research
SNOMED 2	35,207	4,423	4,043	66.1%	Health Records
UMDNS: product category thesaurus	12,857	3,055	3,054	1.6%	Pharma
Alcohol and Other Drug Thesaurus	15,888	2,563	2,264	77.9%	Library
Master Drug Data Base	11,860	2,198	2,198	2.5%	Manufacturing
USP Model Guidelines	1,768	1,768	1,768	82.0%	Health Records
Alternative Billing Concepts	4,619	1,440	1,439	4.1%	Hospital Administration
Library of Congress Subject Headings	6,585	1,521	1,428	90.5%	Library
Metathesaurus FDA Structured Product Labels	6,824	1,344	1,344	88.6%	Regulation
Standard Product Nomenclature	4,809	930	927	12.4%	Pharma
Metathesaurus FDA National Drug Code Directory	17,580	549	549	33.9%	Regulation
Thesaurus of Psychological Index Terms	6,742	572	544	90.4%	Library
Medical Entities Dictionary	3,078	537	491	70.1%	Health Records
All	2,311,194	522,095	301,646		

Medical Subject Headings (MeSH)

MSH	2,4,6-TIP
MSH	hurin protein, Hura crepitans
MSH	norfentanyl monohydrochloride
MSH	Phenylglyoxal
MSH	gas vesicle structural protein A, Bacteria
MSH	yttrium silicate
MSH	2-aminoethanethiosulfuric acid, 35S-labeled
MSH	acethropan-S, acetate
MSH	N-methyl-alpha-tocopheramine nitroxide
MSH	antimony pentachloride
MSH	ISG20 protein, human
MSH	medosulepine, (Z)-isomer
MSH	MUC1 protein, human
MSH	valacyclovir, x-hydrochloride, (D)-isomer
MSH	Fmn1 protein, mouse
MSH	cytochrome c, N-epsilon-acetimidate
MSH	14,16-dianhydrogi-toxigenin-3-O-xylopyranosyl-1-2-O-galactopyranoside
MSH	4-iodoclonidine
MSH	poly(ethylenimine sulfide)
MSH	2-methoxy-6-tridecyl-1,4-benzoquinone
MSH	Lisuride
MSH	Man-(1-3)-(Man-(1-6))-Man
MSH	Sre1 protein, S pombe
MSH	slou protein, Drosophila
MSH	Ac-odv-e56 protein, Autographa californica nucleopolyhedrovirus
MSH	YoYo-3
MSH	cripowellin B
MSH	3-quinuclidinyl atrolactate, (S-(R*,S*))-isomer
MSH	4-vinyl-N-carboxymethylpyridinium
MSH	2,5-dihydroxybenzylidene aminoguanidine
MSH	

SNOMED Clinical Terms

SNOMEDCT bisalbumins
SNOMEDCT Spiramycin Adipate
SNOMEDCT 2,2,2-trichloroethanol
SNOMEDCT Promethazine Hydrochloride
SNOMEDCT thenium closylate
SNOMEDCT CD67 Antigen
SNOMEDCT Ethyleneimine antineoplastic
SNOMEDCT hydrocortisone acetate and neomycin sulfate
SNOMEDCT Hexan-2,5-dione
SNOMEDCT Steroidal neuromuscular blocker
SNOMEDCT trospium chloride
SNOMEDCT Neostigmine Methylsulfate
SNOMEDCT Oligotriacrylate 480
SNOMEDCT Lemon specific immunoglobulin E
SNOMEDCT Monocarboxylate
SNOMEDCT Ethosuximide
SNOMEDCT Phthalic acid ester
SNOMEDCT Combination ulcer healing drugs
SNOMEDCT darunavir
SNOMEDCT Ophthalmic form clotrimazole
SNOMEDCT Blood group antigen Horn
SNOMEDCT Mycoplasma synoviae bacterin
SNOMEDCT Dimethoxanate
SNOMEDCT Demeton
SNOMEDCT Silicon Dioxide
SNOMEDCT ¹³³Iodine
SNOMEDCT Amylose
SNOMEDCT glymidine
SNOMEDCT Parsley specific immunoglobulin E
SNOMEDCT Antihistaminic

LOINC

LNC	Manihot esculenta crantz Antibody.immunoglobulin E
LNC	Glutamine ; urine
LNC	Chlamydia trachomatis D+E+F+G+H+I+J+K IgA ; Bld-Ser-Plas
LNC	HLA-D w16 ; bld-ser-plas
LNC	Colorado tick fever virus Ab ; bld-ser-plas
LNC	Fluconazole ; isolate & serum
LNC	African horse sickness virus Antigen
LNC	Zea mays Ab
LNC	Le ^a b Antibody
LNC	Triglyceride ; Semen
LNC	Leishmania tropica Antibody.immunoglobulin G
LNC	Cystine ; White blood cells
LNC	annexin A5
LNC	Artemisia douglasiana Antibody.immunoglobulin E
LNC	2,4,5-trichlorophenoxyacetate ; bld-ser-plas
LNC	Streptococcus pneumoniae 9 IgG ; bld-ser-plas
LNC	2-hydroxyglutarate ; urine
LNC	Dodecenoylcarnitine (C12:1) ; Cerebral spinal fluid
LNC	Toxocara canis Ab ; cerebral spinal fluid
LNC	Globulin ; bld-ser-plas
LNC	Streptococcus species antibody
LNC	BSA (Bovine serum albumin) ; White blood cells
LNC	Threonine/Creatinine
LNC	Amylases
LNC	Streptococcus pneumoniae 9n Antibody.immunoglobulin G
LNC	Mycoplasma pneumoniae Ab ; body fluid
LNC	Amobarbital ; gastric fluid
LNC	Mycoplasma pneumoniae Antibody.immunoglobulin G
LNC	Haemophilus influenzae B
LNC	Insulin-Like Growth-Factor-Binding Proteins

Product Category Thesaurus

UMD	Reagents, Serology, Virus, Retrovirus, HIV-1, Antibody
UMD	Reagents, Molecular Assay, Tumor Marker, Chromosome, Translocation, t(12;15)
UMD	Reagents, Microbiology, Bacteria, Identification, Listeria monocytogenes
UMD	Reagents, Molecular Assay, Infection, Bacteria, Bordetella Species
UMD	Reagents, Molecular Assay, Infection, Virus, Epstein-Barr, DNA
UMD	B Loci Human Leukocyte Antigen Determination Reagents
UMD	Cell Culture Media, Serum
UMD	Trench Fever Diagnostic Reagents
UMD	Reagents, Serology, Virus, Retrovirus, Human T-Cell Lymphotropic Virus-I/II
UMD	Clostridium botulinum Identification/Detection Reagents
UMD	Reagents, Molecular Assay, Infection, Virus, Eastern Equine Encephalitis, RNA
UMD	Reagents, Hematology, Standard, Coagulation, Plasma
UMD	Reagents, Immunohematology, Antibody Detection/Identification, Enhancement Media, Polyethylene Glycol
UMD	Listeria monocytogenes Detection/Identification Reagents
UMD	Reagents, Molecular Assay, Infection, Virus, Hepatitis G
UMD	Reagents, Immunoassay, Toxicology, Salicylate
UMD	Reagents, Immunoassay, Control, Bone Metabolism
UMD	Alpha2-Antiplasmin Determination Reagents
UMD	Pyridoline Crosslink Determination Reagents
UMD	Activated Partial Thromboplastin Time (APTT) Determination Reagents
UMD	Reagents, Hematology, Fibrinolysis, Plasminogen Activator, Urokinase
UMD	Tuberculosis Diagnostic Reagents
UMD	Reagents, Immunoassay, Tumor Marker, Enzyme, Neuron Specific Enolase
UMD	Reagents, Immunoassay, Tumor Marker, Fecal Occult Blood
UMD	CLEANSER
UMD	Reagents, Molecular Assay, Infection, Bacteria, Ehrlichia Species
UMD	Central Nervous System Drug Level Determination Reagents, Anticonvulsant Agent
UMD	Sinusitis Diagnostic Reagents
UMD	Anti-Filaggrin Antibody Determination Reagents
UMD	Kappa Reagents, Light Chain Monoclonal Immunoglobulin
UMD	Reagents, Molecular Assay, Infection, Virus, Eastern Equine Encephalitis

Master Drug Database

MDDB	Deoxyribose (Bulk) Powder
MDDB	Lithium Chloride (Bulk) Powder
MDDB	Mercurous Chloride (Bulk) Powder
MDDB	Ferric Chloride (Bulk) Powder
MDDB	Ginger Oil (Bulk)
MDDB	Levodopa Powder
MDDB	Antimony Trichloride (Bulk) Powder
MDDB	Calcium Lactate Powder
MDDB	Danthron Powder
MDDB	Vitamin E Acetate (Bulk) Liquid
MDDB	Niacin Powder
MDDB	Betamethasone Acetate (Bulk) Powder
MDDB	Dill Seed Oil
MDDB	Emollient Cream
MDDB	Dye FDC Blue 1 (Brilliant Blue FCF) - Powder
MDDB	Corticotropin (Bulk) Powder
MDDB	Xylometazoline HCl (Bulk) Powder
MDDB	Dentifrices - Solution
MDDB	Orphenadrine Citrate Powder
MDDB	Blood Glucose Calibration - Liquid - Low
MDDB	Xanthan Gum Powder
MDDB	L-Alpha Pinene (Bulk) Powder
MDDB	lavender oil
MDDB	juniper tar
MDDB	Rice Bran (Bulk) Oil
MDDB	Bay Oil (Myrcia Oil)
MDDB	Tamoxifen Citrate (Bulk) Powder
MDDB	Eucalyptol (Bulk) Liquid
MDDB	Hyoscyamine Sulfate Powder
MDDB	Triclosan (Bulk) Powder
MDDB	lavender Oil (Bulk) Oil

RxNORM

RXNORM	Phenazopyridine HCl Powder
RXNORM	Pri-Methylate
RXNORM	Perdiem
RXNORM	Chewing Gum, dose form
RXNORM	Xylocaine-MPF-Epinephrine
RXNORM	Robitussin-DM
RXNORM	proxymetacaine
RXNORM	Petroleum distillate
RXNORM	Miradon
RXNORM	Ciproflaxacin
RXNORM	PerioChip
RXNORM	Oral Strip
RXNORM	Rectal Ointment
RXNORM	Octocaine with Epinephrine
RXNORM	Hydro Pro D
RXNORM	Dynacirc
RXNORM	Hydrophene DH
RXNORM	Paloxin
RXNORM	Levsin/SL Tablets
RXNORM	Glutarol
RXNORM	Benzaclin
RXNORM	Diethylstilbestrol
RXNORM	AMINOSALICYLATE
RXNORM	CEFACLOR MISCELL POWDER (GM)
RXNORM	Chlorpromazine HCl Powder
RXNORM	L-All 12
RXNORM	Therapy Bayer
RXNORM	Bellamine S
RXNORM	Lavacol
RXNORM	

Multum MediSource Lexicon

MMSL	Vitamins
MMSL	Gen-Cyclobenzaprine
MMSL	Uni-Tussin DM
MMSL	Afrin Pump Mist
MMSL	Meperidine+promethazine
MMSL	Lipidil Supra
MMSL	Icy Hot PM
MMSL	Rocuronium
MMSL	Gormel
MMSL	DHT brand of dihydrotachyesterol
MMSL	Schuessler's Acne Remedy
MMSL	AlphaBath
MMSL	Dymadon P
MMSL	Rynesa 12S
MMSL	Mersol
MMSL	doxycycline topical
MMSL	Calcium Sulfate, Anhydrous
MMSL	PSE Allergy
MMSL	Z-Cof DM
MMSL	Ramses Personal
MMSL	Tri-Hist Pediatric
MMSL	Cortisone Acetate Micronized, compounding powder
MMSL	Micrainin
MMSL	Ceron Drops
MMSL	Vasotec
MMSL	epinephrine compounding powder
MMSL	Benoquin
MMSL	Spastrin
MMSL	Tramal SR
MMSL	Antimicrobial Chlorine

Alternative Billing Concepts

ALT	Matthiola graeca / giliflower
ALT	Adoxa moschatellina / common moschatel
ALT	Cinnamomum camphora, camphor, Homeopathic preparation
ALT	Croton eleuteria / cascarilla / amber kabug / sweet bark
ALT	Pediculus capitis, Homeopathic preparation
ALT	Zea italica, corn silk, Homeopathic preparations
ALT	Hippozaeninum / glanders nosode
ALT	Cobalt
ALT	Salvia officinalis, homeopathic preparation
ALT	Arbutus andrachne preparation
ALT	Andira araroba / chrysarobinum / chrysophan / goa powder
ALT	Sedum acre / small houseleek
ALT	Urinum humanum / human urine
ALT	Aurum muriaticum natronatum / double chloride of gold and sodium / sodium chloroaurate
ALT	Cistus canadensis preparation
ALT	Xanthorrhoea arborea preparation
ALT	Cornus florida preparation
ALT	Aquilegia vulgaris preparation
ALT	Ergotinum, homeopathic preparation
ALT	Mimulus lewisii / rose colored musk
ALT	Lac vaccinum coagulatum / milk curds
ALT	Robinia pseudoacacia / yellow locust
ALT	Solidago virgaurea, homeopathic preparation
ALT	Cholesterinum / cholesterine
ALT	Benzinum dinitricum, benzinum, benzol, coal naphtha, Homeopathic preparation
ALT	Derris pinnata / pongram
ALT	Calcarea renalis, Homeopathic preparations
ALT	Centella asiatica, homeopathic preparation
ALT	Culex musca, Homeopathic preparation
ALT	Python regia (homeopathic remedy)
ALT	Psidium cattleianum / Calif. guava

Chemicals in UMLS source vocabularies

Source	Size	Chemicals (Broad)	Chemicals (Narrow)	Overlap with MeSH (%)	Purpose
Medical Subject Headings (MeSH)	296,338	266,927	220,228	100.0%	Biomedical Literature
LOINC	114,351	24,859	22,900	11.4%	Laboratory Medicine
SNOMED CT	317,177	26,659	22,727	35.1%	Health Records
Clinical Terms Version 3	181,192	20,685	20,673	23.1%	Health Records
Multum MediSource Lexicon	52,851	20,152	20,146	24.1%	Health Records
NCI Thesaurus	67,803	15,917	14,886	58.0%	Research
SNOMED International	112,712	15,878	13,178	41.9%	Health Records
National Drug File - Reference Terminology	39,163	13,786	11,969	66.8%	Health Records
UMLS Metathesaurus	120,458	15,285	11,240	61.3%	Biomedical Literature
National Drug Data File Plus Source Vocabulary	31,141	10,959	10,903	46.1%	Health Records
RXNORM	186,066	10,795	10,430	62.2%	Health Records
Veterans Health Administration National Drug File	24,913	6,991	6,984	44.5%	Health Records
MEDCIN	269,443	6,325	6,323	38.7%	Health Records
Physician Data Query	10,642	4,864	4,839	55.0%	Health Records
CRISP Thesaurus	16,682	5,045	4,064	75.5%	Research
SNOMED 2	35,207	4,423	4,043	66.1%	Health Records
UMDNS: product category thesaurus	12,857	3,055	3,054	1.6%	Pharma
Alcohol and Other Drug Thesaurus	15,888	2,563	2,264	77.9%	Library
Master Drug Data Base	11,860	2,198	2,198	2.5%	Manufacturing
USP Model Guidelines	1,768	1,768	1,768	82.0%	Health Records
Alternative Billing Concepts	4,619	1,440	1,439	4.1%	Hospital Administration
Library of Congress Subject Headings	6,585	1,521	1,428	90.5%	Library
Metathesaurus FDA Structured Product Labels	6,824	1,344	1,344	88.6%	Regulation
Standard Product Nomenclature	4,809	930	927	12.4%	Pharma
Metathesaurus FDA National Drug Code Directory	17,580	549	549	33.9%	Regulation
Thesaurus of Psychological Index Terms	6,742	572	544	90.4%	Library
Medical Entities Dictionary	3,078	537	491	70.1%	Health Records
All	2,311,194	522,095	301,646		

Hidden references to chemicals

Example: ICD9-CM

Accidental poisoning by other opiates NOS
Accidental poisoning by codeine
Accidental poisoning by pethidine
Accidental poisoning by morphine
Accidental poisoning by opium
Accidental poisoning by aromatic analgesics NOS
Accidental poisoning by aromatic analgesics NEC
Accidental poisoning by acetanilide
Accidental poisoning by phenacetin
Accidental poisoning by aminophenazone
Accidental poisoning by antirheumatics NOS
Accidental poisoning by pentazocine
Accidental poisoning by pentobarbitone
Accidental poisoning by quinalbarbitone
Accidental poisoning by bromides
Accidental poisoning by cabromal derivatives
Accidental poisoning by carbamic esters
Accidental poisoning by chlorpromazine
Accidental poisoning by fluphenazine
Accidental poisoning by prochlorperazine
Accidental poisoning by promazine
Accidental poisoning by spiperone
Accidental poisoning by chlordiazepoxide

Example:

intox or “*poison* or

allerg returns 10800

non-chemical concepts

roughly half of them refer to
chemicals

Explicit reference to chemicals

VTSL SNOMED CT Core Browser

Concepts

Descriptions

Codes

Roots

Main Menu

Parent(s):

(Select a parent to make it the "Current Concept".)

[Toxic effect of cadmium AND/OR its compounds \(disorder\)](#)

Current Concept:

[Cadmium poisoning \(disorder\)](#)

Child(ren):

(N=1) (Select a child to make it the "Current Concept".)

[Cadmium pneumonitis \(disorder\)](#)

Current Concept:

Fully Specified Name: Cadmium poisoning (disorder)

ConceptId: 3398004

Defining Relationships:

Is a Toxic effect of cadmium AND/OR its compounds (disorder)

Causative agent [Cadmium \(substance\)](#)

This concept is primitive.

Qualifiers:

[View Qualifying Characteristics and Facts](#)

Descriptions (Synonyms):

Fully Specified Name: Cadmium poisoning (disorder)

Preferred: Cadmium poisoning

Chemicals in UMLS: Summary

- MeSH (85% Substance terms) is the most important source for chemicals
- Health care related sources include also natural products, drugs, lab procedures
- Pharmacy related sources include pharmaceutical preparations and products
- Many sources are rather heterogeneous (UMLS typing not always consistent)
- (implicit) reference to chemicals in most clinical terminologies

Ontology aspects of UMLS chemistry sources

Ontology aspects of UMLS chemistry sources

- UMLS only includes Concept – Relation – Concept triplets
- Only very few UMLS sources are “ontology-like”, i.e. they have some formal semantics, e.g. SNOMED CT or NDF-RT
- UMLS distinguishes thesaurus-style broader/narrower hierarchy-building relations from more precise ones (“relation attributes”)
- Only part of the latter describe the entities to be represented themselves (e.g. part-of, has-active ingredient), other ones describe the representational units and the attached terms (“mapped-to”, “has-translation”)

Ontological relations involving chemicals (608,315)

has_ingredient	207469
has_dose_form	106962
has_component	87491
isa	35621
measures	28368
has_va_product_component	27105
has_causative_agent	20580
has_active_ingredient	19290
chemotherapy_regimen_has_component	10713
may_be_treated_by	8894
contraindicated_drug	7412
physiologic_effect_of	5008
has_direct_substance	4948
has_mechanism_of_action	4040
biological_process_involves_gene_product	3835
uses_substance	3265
associated_with	2686
gene_encodes_gene_product	1942
mechanism_of_action_of	1673
has_gene_product_element	1399
may_be_prevented_by	1258
has_divisor	1247
has_contraindicating_class	1198
entry_combination_of	1139
is_physiologic_effect_of_chemical_or_drug	1060
has_challenge	1036

Chemical – Rel – Non-Chemical

Ontological relations between chemicals (173,502)

isa	131073
has_active_ingredient	13345
has_ingredient	8158
has_dose_form	4026
has_precise_ingredient	2500
used_for	1970
contains	1859
has_mechanism_of_action	1603
has_form	1507
associated_with	1335
is_biochemical_function_of_gene_product	1266
may_be_a	1123
see	714
has_va_product_component	633
has_free_acid_or_base_form	614
has_contraindicating_class	482
has_target	158
subtype_of	150
has_contraindicating_mechanism_of_action	139
co-occurs_with	118
reformulation_of	115
is_chemical_classification_of_gene_product	108
complex_has_physical_part	101
biomarker_type_includes_gene_product	91
chemical_or_drug_affects_gene_product	71
has_chemical_structure	68

Chemical – Rel – Chemical

Analysis of relations in UMLS

- Broad spectrum and high number of relations between chemicals and non-chemicals. Of interest for relating chemical with other concepts of biomedical interest.
- Rather poor in terms of inter-chemical relations, often due to Semantic type misassignments

SNOMED CT:	quinupristin-dalfopristin	has_active_ingredient	dalfopristin
NDFRT:	Raloxifene Hydrochloride	has_mechanism_of_action	Selective Estrogen Receptor Modulators
CRISP:	Reserpine	used_for	reserpate derivative
NCI:	Rimantadine Hydrochloride	has_free_acid_or_base_form	Rimantadine

MeSH in PubChem

Caffeine

Pharmacological Action

- Central Nervous System Stimulants - A loosely defined group of drugs that tend to increase behavioral alertness, agitation,
- Phosphodiesterase Inhibitors - Compounds which inhibit or antagonize the biosynthesis or actions of phosphodiesterases.

Pharmacological Classification

- Chemical Actions and Uses
 - Pharmacologic Actions
 - Molecular Mechanisms of Pharmacological Action
 - Enzyme Inhibitors
 - Phosphodiesterase Inhibitors**
 - Physiological Effects of Drugs
 - Central Nervous System Stimulants**
- Therapeutic Uses
 - Central Nervous System Agents
 - Central Nervous System Stimulants**

Chemical Classification

- Heterocyclic Compounds
 - Alkaloids
 - Xanthines
 - Caffeine
 - Heterocyclic Compounds, 2-Ring
 - Purines
 - Purinones
 - Xanthines
 - Caffeine

Properties as parents in informal hierarchy

Mapping / Tagging

UMLS MetaMap / Medical Text Indexer

MetaMap Version Used: metamap09

MetaMap Options: -A+

Lexicon Used: 2009

Knowledge Source Used: 09

Input Text:

Accidental poisoning by codeine

Accidental poisoning by pethidine

Accidental poisoning by morphine

Accidental poisoning by opium

Accidental poisoning by aromatic analgesics NOS

Accidental poisoning by aromatic analgesics NEC

Accidental poisoning by acetanilide

Accidental poisoning by phenacetin

Accidental poisoning by aminophenazone

Accidental poisoning by antirheumatics NOS

Accidental poisoning by pentazocine

Accidental poisoning by pentobarbitone

Accidental poisoning by quinalbarbitone

Accidental poisoning by bromides

Accidental poisoning by cabromal derivatives

Accidental poisoning by carbamic esters

Accidental poisoning by chlorpromazine

Accidental poisoning by fluphenazine

567 Morphinans [Organic Chemical]

577 Seconal [Organic Chemical,Pharmacologic Substance]

604 Talwin [Organic Chemical,Pharmacologic Substance]

627 Acetanilides [Organic Chemical,Pharmacologic Substance]

637 Aromatic (AROMATICS) [Organic Chemical,Pharmacologic Substance]

645 Esters [Organic Chemical]

645 derivatives [Chemical Viewed Structurally]

660 Acetanilid (acetanilide) [Organic Chemical,Pharmacologic Substance]

660 Amidophenazon (Aminopyrine) [Organic Chemical,Pharmacologic Substa

660 Bromides [Inorganic Chemical]

660 Chlorpromazine [Organic Chemical,Pharmacologic Substance]

660 Codeine [Organic Chemical,Pharmacologic Substance]

660 Morphine [Organic Chemical,Pharmacologic Substance]

660 Opium [Organic Chemical,Pharmacologic Substance]

660 Pentazocine [Organic Chemical,Pharmacologic Substance]

660 Pentobarbitone (Pentobarbital) [Organic Chemical,Pharmacologic Substa


660 Pethidine (Meperidine) [Organic Chemical,Pharmacologic Substance]


660 Phenacetin [Organic Chemical,Pharmacologic Substance]

660 Quinalbarbitone (Secobarbital) [Organic Chemical,Pharmacologic Substa

1000 Fluphenazine [Organic Chemical,Pharmacologic Substance]

Whatizit

EMBL-EBI  EB-eye Search Enter Text Here Advanced Search

Databases Tools EBI Groups Training Industry About Us Help Site Index 



1 Place your text/query here:

Accidental poisoning by codeine
Accidental poisoning by pethidine
Accidental poisoning by morphine
Accidental poisoning by opium
Accidental poisoning by aromatic analgesics NOS
Accidental poisoning by aromatic analgesics NEC
Accidental poisoning by acetanilide
Accidental poisoning by phenacetin
Accidental poisoning by aminophenazone


2 The text is:

3 Select a pipeline:

[Pipelines Info/Update](#)

What is it?

Whatizit is a text processing system that allows you to do textmining tasks on text. The tasks come defu

EMBL-EBI  EB-eye Search Enter Text Here

Databases Tools EBI Groups Training Industry About Us

Resulting tagged text

Accidental poisoning by **codeine** Accidental poisoning by pethidine
Accidental poisoning by **morphine** Accidental poisoning by opium
Accidental poisoning by aromatic **analgesics** NOS Accidental
poisoning by aromatic **analgesics** NEC Accidental poisoning by
acetanilide Accidental poisoning by **phenacetin** Accidental poisoning
by aminophenazone Accidental poisoning by antirheumatics NOS
Accidental poisoning by pentazocine Accidental poisoning by
pentobarbitone Accidental poisoning by quinalbarbitone
Accidental poisoning by **bromides** Accidental poisoning by
cabromal derivatives Accidental poisoning by carbamic **esters**
Accidental poisoning by **chlorpromazine** Accidental poisoning by
fluphenazine .



Conclusions

- Most Biomedical Terminologies contain chemical concepts, drugs or concepts referring to them
- MeSH has the highest coverage
- Fairly good coverage of semantic relations linking chemicals to non-chemicals
- No significant source for semantic relations between chemicals
- Mappings ChEBI – UMLS:
 - to MeSH via PubChem, but only higher level MeSH terms
 - NLP tools (MetaMap, Medical Text Indexer, WhatizIt) not yet optimized for Chemical names.

Veterans Health Administration National Drug File

VANDF	WHEAT DEXTRIN
VANDF	ALOE/BENZOCAINE/LANOLIN/MENTHOL
VANDF	benazepril
VANDF	COLOR,ARTIFICIAL
VANDF	Acacia Extract
VANDF	Secobarbital sodium
VANDF	Cilastatin Sodium
VANDF	POLYTHIAZIDE/PRAZOSIN
VANDF	CARDIOVASCULAR AGENTS,OTHER
VANDF	Doxazosin
VANDF	Phenylephrine + promethazine + codeine
VANDF	SODIUM XYLENESULFONATE
VANDF	Potassium
VANDF	ALLERGENIC EXTRACT, PENICILLIUM NOTATUM
VANDF	LOXILAN
VANDF	Fosfomycin
VANDF	CEPHALOSPORIN 2ND GENERATION
VANDF	ALLERGENIC EXTRACT, TREE, MAPLE MIX
VANDF	CALCIUM IODATE
VANDF	Antiemetics
VANDF	DYE EVANS BLUE
VANDF	ACETAMINOPHEN/DEXTROMETHORPHAN/GUAIFENESIN/PSEUDOEPHEDRINE
VANDF	ALLERGENIC EXTRACT, JUNE POLLEN
VANDF	DUODERM HYDROGEL C#1879-87
VANDF	Equine diphtheria antitoxin
VANDF	Loxipine Succinate
VANDF	WOOL WAX ALCOHOL

CRISP Thesaurus

SAB	CUIStr
CSP	Tetrabenazine
CSP	gamma-Aminobutyrate
CSP	methoxyindole
CSP	Yellow Fever Vaccine
CSP	Clomiphene
CSP	Chenodeoxycholate
CSP	Crack Cocaine
CSP	H antigen, bacterial
CSP	Salts
CSP	Methimazole
CSP	erythroidine
CSP	halobiphenyl/halotriphenyl compound
CSP	Selenoprotein P
CSP	cyclohexane carboxylate
CSP	Lomustine
CSP	Shiga Toxins
CSP	Prodrugs
CSP	Diuretics
CSP	Leukotrienes E
CSP	Proteolipids
CSP	Thymidine Monophosphate
CSP	aspidospermine
CSP	halocarbon compound
CSP	Mitomycin
CSP	Abortifacient Agents
CSP	Morning After Pill
CSP	Cyclophosphamide
CSP	Poisons
CSP	virus envelope

NCI Thesaurus

NCI	Zaditor
NCI	hydrocortisone acetate
NCI	KOS-953
NCI	Neoral
NCI	Isotretinoin
NCI	Spigelia Fluidextract
NCI	palmitoleic acid
NCI	Absorbine
NCI	Monoclonal Antibody N901-bR
NCI	Coreg Butoxamine HCl
NCI	Citofolin
NCI	Amoxil
NCI	Procyclidine hydrochloride
NCI	Methylene Chloride
NCI	SC 48334
NCI	Egtazic Acid
NCI	Valproate
NCI	CD3-Epsilon-Associated Protein
NCI	Differentiation Inducer
NCI	Clonoxifen
NCI	Myristic Acid
NCI	piroxicam
NCI	Methoxamine
NCI	Dynacirc
NCI	Hexa-Germ
NCI	Trihexyphenidyl Hydrochloride
NCI	Iodamide
NCI	CD11b Antigens
NCI	Abbokinase

Clinical Terms V3

RCD	Compound salicylic acid powder
RCD	Expulin
RCD	Chlorinated phenol disinfectant
RCD	amyl nitrate
RCD	Vioform Hydrocortisone
RCD	Alcuronium
RCD	methyl isocyanate
RCD	Lopid
RCD	Neupogen
RCD	Shannon stoma adhesive plaster
RCD	Eolarix vaccine
RCD	X-porphyrin
RCD	Deltastab
RCD	Geref 50
RCD	C-Peptide
RCD	Abidec
RCD	soldering flux
RCD	Buspirone
RCD	cabergoline
RCD	Dental etching agent
RCD	E104
RCD	Adenoscan
RCD	Fefol-Vit Spansule
RCD	Budesonide
RCD	deteclo
RCD	Rigid gas permeable contact lens preparations
RCD	Cannabis substance
RCD	lypressin
RCD	Progynova
RCD	Fraxiparin

Metathesaurus

SAB	CUIStr
MTH	Nile Blue
MTH	Manganum sulphuricum, homeopathic preparation
MTH	Dichlorodiphenyldichloroethane
MTH	WT2 protein
MTH	Neuraminic acid
MTH	Hemoglobin Parchman
MTH	Evi-1 protein
MTH	Aesculus hippocastanum, homeopathic preparation
MTH	sulfur oxide
MTH	Fiboran
MTH	Synaptotagmin XII
MTH	chondrocyte expressed protein-68
MTH	Tantalum
MTH	Prothrombin
MTH	Albumin ; dialysis fluid peritoneal
MTH	Tylos Preparation
MTH	Helicobacter pylori antibody
MTH	Coccus cacti, Homeopathic preparation
MTH	Equisetum hyemale, Homeopathic preparation
MTH	ovocleidin-116
MTH	Bupleurum preparation
MTH	Nux moschata, Homeopathic preparation
MTH	Ear Drops brand of carbamide peroxide
MTH	SLC5A5 protein, human
MTH	Keratin-1
MTH	PERILLA preparation
MTH	Prostaglandins I
MTH	Phalaris arundinacea antigen
MTH	Horse Chestnut Preparation

Alcohol and Other Drug Thesaurus

SAB	CUIStr
AOD	Thrombin
AOD	Progestins
AOD	Organic Chemicals
AOD	Sulfonylurea Compounds
AOD	Chlorpromazine
AOD	Phenacetin
AOD	Metallothionein
AOD	Anti-Infective Agents, Local
AOD	R-38486
AOD	Apolipoproteins A
AOD	Beta-glucuronidase
AOD	Cinchona Alkaloids
AOD	ethanol metabolite
AOD	Acyclovir
AOD	Phosphothreonine
AOD	Sulfanilamide
AOD	IgE
AOD	Captopril
AOD	compound with nitrogen-nitrogen bond
AOD	excitatory neurotransmitters
AOD	Asheim-Zondek hormone
AOD	Phenylthiohydantoin
AOD	Anthracycline
AOD	Polysaccharides, Bacterial
AOD	Turpentine
AOD	Aliphatic unsaturated hydrocarbon
AOD	Hydromorphone Hydrochloride
AOD	Neomycin
AOD	Vitamin K

Library of Congress Subject Headings

SAB	CUIStr
LCH	Collodion
LCH	Auxins
LCH	Mannose
LCH	Vitamin U
LCH	Ethylene
LCH	Nicergoline
LCH	Platinum
LCH	Guanidine
LCH	Indophenol
LCH	Spironolactone
LCH	Glycolipids
LCH	Oxides
LCH	Amoxicillin
LCH	Drug vehicle
LCH	Tetrachlorodibenzodioxin
LCH	Endosulfan
LCH	Cyclacillin
LCH	Etoposide
LCH	Amidines
LCH	Veratrine
LCH	Charcoal
LCH	Saralasin
LCH	Aluminum Silicates
LCH	Aminobutyric Acid
LCH	Glutamine
LCH	Amino Alcohols
LCH	acetamide
LCH	Theophylline
LCH	Aerosols