CSDB scheme v.4.54

This sample database contains data on two compounds:

1. b-D-Galf*N*Ac3NA-(1
   \6
   -4)-{{{a-D-Glc*p*-(1-3)}}}/n=3-5/-a-D-Glc*p*-(1-

2. [L-Ala OR Gly]-(2 ?%EtN-50%P
   \6 \7
   L-gro-?-D-manHep*p*(1-?)-a-D-Glc*p*A-(1-4)-a-Kdo

Compound **1** is described in three publications (**1, 2, 3**).

Within book chapter **1** it was extracted from the cell wall of the bacterium *Providencia stuartii* O32 (organism **2** no host) and its structure was erroneously elucidated as **4** (later revised to **1**), also the structure of compound **2** was elucidated;

Within article **2** this known structure was found also in the hybrid bacteria *Providencia stuartii* x … (organism **3** hybrid with **7**) and *Proteus penneri* (organism **4**) later reclassified to *Proteus valgaris* (organism **6**), polysaccharide was extracted from infected human (organism **5**) kidney;

Within article **3** no taxonomic data were available but there are revised ^{13}C NMR data and conformation studies.

These example data are NOT a complete example database!

basenames

id	basename
1	Glc
2	Gal
3	LDmanHep
4	Ala
5	Gly
6	Ac
7	Kdo
8	P
9	Et

residue_types

basename_id	qualifier_id	cycle size	res_type	has absolute	atoms	protonation	description	enanthiomer	default_out	wurcs
1	1	hex	ald	YES	xoooooo-oooodo-ooodoo	111112	glucose		1	?2122h
2	3	hex	ald	YES	xnnooa-onnoda-onndoa	111110	galacturonic acid 2,3-diamine		1	?2112A 2*N_3*N
3	1	hep	ald	NO	xoooooo-oooodoo-ooodooo	1111112	L-gro-D-manHep	DL-manHep	1	?11221h
4	1	nsu	pep	YES	and--	013	alanine		1,2	A2m_2*N
5	1	nsu	pep	NO	an--	02	glycine		1,2	Ah_2*N
1	2	hex	ald	YES	xoooaa-ooooda-oooodoa	111110	glucuronic acid		1	?2122A
6	1	mva	alk	NO	ad--	03	acetic acid		1	*OCC/3=O
7	1	oct	ket	NO	axdoooooo-aodoodoo-aododooo	00211112	Kdo	L-3dmanOct-ulosonic	2	A?d1122h
8	1	nsu		NO	a--	0	phosphoric acid		0	*OPO*/3O/3=O
9	4	nsu		NO	on--	22	ethanolamine		1	*1OCCN*2

qualifiers

id	name
1	
2	A
3	N3NA
4	N

keyword_link

article_id	keyword_id
1	1
1	2
2	1
2	2
3	1
3	2

keywords

id	keyword
1	o-antigen
2	elucidation
…	

residues

id	anomeric	absolute	basename_id	ringsize	qualifier_id	atoms	stereocode	smiles	wide spread	
1	a	D	1	p	1	oooodo	121220		YES	aDGlc*p*
2	b	D	2	f	3	onnoda	221120		NO	bDGalf*N*3NA
3	?	X	3	p	1	oooodoo	2112210		YES	L-gro-D-manHep
4	x	L	4	?	1	and	010		YES	L-Ala
5	x	X	5	?	1	an	00		NO	Gly
6	a	D	1	p	2	ooooda	121220		YES	aDGlc*p*A
7	x	X	6	?	1	ad	0		YES	Ac
8	a	X	7	?	1	aodo??oo	02011220		YES	a-Kdo
9	x	X	8	?	1	a	0		YES	P
10	x	X	9	?	4	on	00		YES	EtN

method_link

article_id	method_id
1	1
1	2
2	2
3	2

methods

id	method
1	Smith degradation
2	NMR spectroscopy
…	

subrepeats

residue_instance_id	subrepeat_id	subrepeat_start	subrepeat_end	repeat_data
1	1	YES	YES	3-5

institution_link

article_id	institution_id
1	1
1	2
2	2
3	3

insitutions

id	institution
1	Zelinsky Institute of Organic Chemistry
2	Hirszfeld Institute
3	Grupo de Carbohidratos, Departamento de Quimica Organica Biologica, Instituto de Quimica Organica, Madrid, Spain

residue_instances

id	compound_id	residue_id	root	donors	mva donors	has aglycon
1	1	1	NO	1	0	NO
2	1	2	NO	1	1	NO
3	1	1	YES	2	0	NO
4	2	3	NO	0	0	NO
5	2	4	NO	0	0	NO
6	2	5	NO	0	0	NO
7	2	6	NO	2	0	NO
8	1	7	NO	0	0	NO
9	2	8	YES	2	0	NO
10	2	9	NO	1	0	NO
11	2	10	NO	0	0	NO

connection_table

id	donor_id	goes from	goes_to	acceptor_id	link type_id	varset_id	stoichio metry	external
1	1	1	3	1	1	NULL	100	NO
2	1	1	3	3	1	NULL	100	YES
3	2	1	6	3	1	NULL	100	NO
4	3	1	4	1	1	NULL	100	YES
5	4	1	-1 (=?)	7	1	NULL	100	NO
6	8	1	2	2	2	NULL	100	NO
7	5	2	6	7	2	1	100	NO
8	6	2	6	7	2	1	100	NO
9	7	1	4	9	1	NULL	100	NO
10	10	0	7	9	3	NULL	50	NO
11	11	1	0	10	3	NULL	0 (=?)	NO

author_link

author_pos	article_id	author_id
1	1	1
2	1	2
3	1	3
4	1	4
1	2	4
2	2	5
1	3	6
2	3	7

authors

id	author
1	Toukach PV
2	Kocharova NA
3	Shashkov AS
4	Twarzhewska A
5	Lugowski C
6	Gonzalez-Jimenez I
7	Bernabe M

book_editors

editor_pos	book_id	editor_id
1	1	1
2	1	7

genera

id	genus
1	*unspecified*
2	Providencia
3	Proteus
4	Homo

parenthesized "genera" mean higher ranks where no exact taxonomy is known, e.g. (plant)

books

id	book_name	publisher	symposium
1	Glycans; Ed.: Que H; Berlin	Springer	NO

nmr_signal_instances

residue_instance_id	spectrum_id	atom number	chem_shift
1	1	1	100.6
1	1	2	73.2
1	1	3	74.5
1	1	4	77.7
1	1	5	72.1
1	1	6	61.7
…			

link_types

id	link_type	description
1	glycosic	bla-bla-bla
2	amidic	bla-bla-bla
3	diester	bla-bla-bla
4	amine	bla-bla-bla
5	carbon-carbon	bla-bla-bla
6	other	bla-bla-bla

fuzzy_sets

id	varset_id	logic
1	NULL (if not a variant in another fuzzy-set)	XOR

journals

id	full_name	abbrev	alt_abbrev	pubisher	NLM_id
1	Carbohydrate Research	Carb. res.	NULL	Elsevier	0043535

phyla

id	phylum
1	
2	Proteobacteria
3	Chordata

species

id	species
1	sp.
2	stuartii
3	penneri
4	sapiens
5	vulgaris

strains

id	strain
1	
2	O32
3	ATCC 31

publications

id	book_id	journal_id	title	abstract	year	volume	sub volume	start page	end page	www	PMID	DOI	NLMID	email
1	1	NULL	<title1>	<abs1>	2003	202	NULL	2001	2011	NULL	12345	10.12/20	1234567	NULL
2	NULL	1	<title2>	<abs2>	2004	203	1	123	150	NULL	NULL	NULL	0043535	NULL
3	NULL	1	<title3>	<abs3>	2004	203	3	130	140	http://sci.ru?id=23	NULL	NULL	0043535	tou@glyc.ru

nmr_solvent_link

solvent_id	spectrum_id
1	1
2	1

nmr_spectra

id	article_id	compound_id	nucleus	spectrum_lincode	temperature	pH
1	1	1	C	<spectrum1>	298	7.5
2	1	2	H	<spectrum2>	303	NULL
3	3	1	C	<spectrum3>	303	NULL

nmr_solvents

id	solvent	amount	unit
1	D2O	90	vol %
2	NaOD	25	mM

publication_specific

article_id	compound_id	orig compound_id	structure location	enzymes	notes	has_biogen	synthesis	has 3D	bio activity	elucidated
1	1	4	fig.2	LgtA	O-deacylated	genetic data available	enzymatic	NULL	NULL	YES
1	2	NULL	fig.3	glucosidase	NULL	serological data	chemical	NULL	cross-reactivity with Mabs ER1	YES
2	1	NULL	p. 125	NULL	NULL	NULL	NULL	NULL	NULL	NO
3	1	NULL	p. 135	NULL	NULL	NULL	chemical fragmentary	conformational data	NULL	NO

organisms

id	tax_group	phylum_id	genus_id	species_id	strain_id	NCBI_id	exact_taxid	hybrid_with
1	NULL	NONE (1)	NONE (1)	NONE (1)	NONE (1)	NULL	NULL	NULL
2	bacteria	2	2	2	2	588	NO	NULL
3	bacteria	2	2	2	NONE (1)	588	NO	7
4	bacteria	2	3	3	3	102862	YES	NULL
5	mammal	3	4	4	NONE (1)	9606	YES	NULL
6	bacteria	2	3	5	3	102862	YES	NULL

organism_remap

article_id	published organism_id	organism_id	relation
2	4	6	NEW

compounds

id	lincode	st_comment	unit_type	mw	formula	aglycon	glycoct	_normRU	virtual
1	<struct 1>		chem	n=30	NULL	lipid A	…	…	NO
2	<struct 2>	EtN = ….	oligo	NULL	NULL	NULL	…	…	NO

diseases

id	disease
1	NONE
2	diarrhea
3	pneumonia

disease_attribs

disease_id	attr_name	attr_value
2	ICD11	ME05.1
3	ICD11	CA40

compound_refs

compound_id	resource_id	reference
1	1	123456
1	2	USA #12

external_resources

id	resource
1	CA-RN
2	patent

trivial_name_link

compound_id	name_id
1	1
1	2

trivial_names

id	name
1	ofigeose
2	khrenobiose

organ_tissue

id	organ_tissue
1	NONE
2	cell wall
3	kidney
4	brain melanoma

main_link

article_id	tax_group	organism_id	host_id	oti_id	disease_id	compound_id	bigfile id
1	bacteria	2	NONE(1)	2	2	1	60001
1	bacteria	2	NONE(1)	NONE(1)	2	2	60002
2	bacteria	3	5	NONE(1)	2	1	60003
2	bacteria	4	5	3	2	1	60003
3	bacteria	NONE(1)	NONE(1)	NONE(1)	NONE(1)	1	60004

class_link

compound_id	class_id
1	1
2	2

compound_classes

id	class
1	O-antigen
2	exooligisaccharide

dump_rrs

bigfile id1	bigfile id2
60001	60002
60002	60001
…	

dump_file

order_num	bigfile id	submitter_id	last_update	refman_id1	refman_id2
1	60001	1	01.02.2009	4534	NULL
2	60002	1	02.02.2009	5512	NULL
3	60003	1	04.09.2009	5520	NULL
4	60004	2	10.01.2009	5560	6667

submitters

id	name
1	Schwarz
2	Toukach