

BRENDA Exercises I

Intro

Problems

Structure

Data

Search

SubStruc

Thesaurus

Ontologies

Outlook

● **Exercises**

Nomenclature



What's the EC number and recommended name of the enzyme 'enolase'?

Find the common reaction diagram catalized by this enzyme?



Find all synonyms of 'Glucose-6-phosphate isomerase'?



Find the CAS registry number of 'fructokinase'?

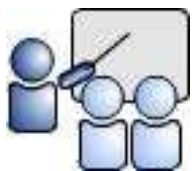


How many different hydrolyses exists which 'acting on ester bonds'?

Tip: EC tree -> EC in 'EC Number'



Find the catalytic activities of phosphofructokinases in an appropriated ontology?



BRENDA Exercises II

Intro

Problems

Structure

Data

Search

SubStruc

Thesaurus

Ontologies

Outlook

● Exercises

Reaction & specificity



Find all inhibitors of 'Helicobacter pylori' catalase



Find all pathways 'fructose-bisphosphate aldolase' is involved?



Find all substrates for enzymes derived by Human Immunodeficiency Virus



Find all reaction types catalysed by the enzyme 'enolase'



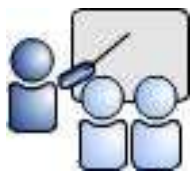
Find all common reactions (defined by the IUBMB) where glucose is a substrate and NADH is a product



Find all activating compounds which have phenol as a substructure (use substructure search)



Find all entries in db for oxidoreductases in bacteria where NAD⁺ is a Cofactor?



BRENDA Exercises III

Intro

Problems

Structure

Data

Search

SubStruc

Thesaurus

Ontologies

Outlook

● **Exercises**

Functional Parameters



Find enzymes with the highest and lowest pH optimum / pH range stored in db



Find enzymes with an pH Optimum higher 100°C



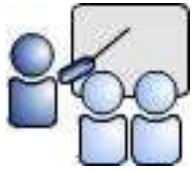
Find all KM values below 1e-04 [mM] of oxidoreductases (EC begins with 1.) where substrate name contains aldehyde in all mammalia (e.g. rat,human,...) with any commentary (contains '_' = at least one character)



Find all enzymes with Mg²⁺ dependence



You plan an experiment where you want to inhibit the alcohol dehydrogenase activity in cell extract without inhibiting any other enzyme. What's your strategy? Try to find a specific inhibitor



BRENDA Exercises IV

Intro

Problems

Organism related Information/ Enzyme Structure

Structure

Data

Search

SubStruc

Thesaurus

Ontologies

Outlook

● Exercises



How many enzymes expressed in 'glia'. Try **Quick Search** and different options in **Brenda Tissue Ontology**



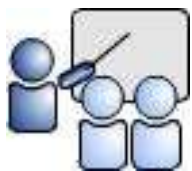
a.) Search for all sequences contains amino acid sequence with the following simple pattern RxxENFxl in Quick Search

b.) Search for the enolase PROSITE pattern
[LIVT]-[LIVP]-[LIV]-[KQ]-x-[ND]-Q-[INV]-G-[ST]-[LIV]- [STL]-[DERKAQG]-[STA]
(First translate to regular expression and use Sequence Search)

What kind of enzymes classes you find beside Acyltransferases (EC 2.3.*)?
(see Distribution of hits in the Enzyme Classification (EC) System)



Try to identify by a short DNA sequence a possible enzyme function:
TACTGTGCGAAAGCTTATTCTG



BRENDA Exercises V

Intro

Problems

Advanced Search / Fulltext / EC Tree

Structure

Data

Search

SubStruc

Thesaurus

Ontologies

Outlook

● **Exercises**



Search for all enzymes in chordata with:

Substrate: ADP

Cofactor: ATP

Inhibitor: L-phenylalanine

Activating Compound: D-fructose 1,6-diphosphate

KM Value < 0.3 [mM]

pH Optimum between 6.5-7.3

Molecular Weight: 100000 kDa

expressed in brain forming a tetramer



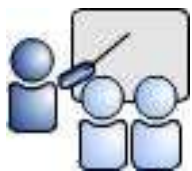
Find all entries in db where

a.) 'Glucose' is mentioned

b.) 'ATP' and 'brain' is mentioned



Download all amino acid sequences specific for alcohol dehydrogenase



BRENDA Exercises VI

Intro

Problems

Substructure Search

Structure



Data

Search

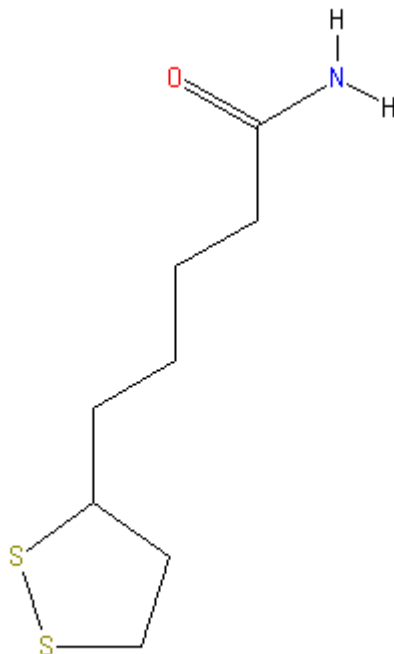
SubStruc

Thesaurus

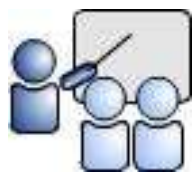
Ontologies

Outlook

● **Exercises**



Find out the name (and synonyms) of the displayed compound, the function of the compound in metabolism and other compound involved in some reactions



BRENDA Exercises VII

Intro

Problems

Structure

Data

Search

SubStruc

Thesaurus

Ontologies

Outlook

● **Exercises**

Taxonomy Tree



What are the scientific names for cat, rat, dog and mouse?



E. coli belongs to which family?



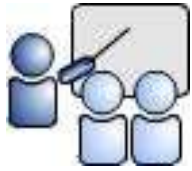
Find all species with a link to BRENDA?



You found no information in BRENDA and literature to the organism *Drosophila saltans*. What's your strategy to develop hypotheses for your experiments



<http://www.brenda.uni-koeln.de/temp/Memory.html>



ExPaSy

Intro

Problems

Structure

Data

Search

SubStruc

Thesaurus

Ontologies

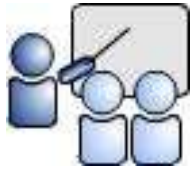
Outlook

● Exercises

Access to ENZYME				
● by EC number:	<input type="text" value="1"/>	<input type="text" value=".1"/>	<input type="text" value=".1"/>	<input type="text" value=".1"/>
	<input type="button" value="Search"/>			
● by enzyme class				
● by description (official name) or alternative name(s):	<input type="text"/>			<input type="button" value="Search"/>
● by chemical compound				
● by cofactor				
● by search in comments lines				
● SRS - Sequence Retrieval System				

Official Name
Alcohol dehydrogenase.
Alternative Name(s)
Aldehyde reductase.
Reaction catalysed
An alcohol + NAD(+) <=> an aldehyde or ketone + NADH
Cofactor(s)
Zinc or Iron.
Comments
<ul style="list-style-type: none">● Acts on primary or secondary alcohols or hemiacetals.● The animal, but not the yeast, enzyme acts also on cyclic secondary alcohols.

+SwissProtIDs



ExPaSy- ENZYME

<http://www.expasy.org/enzyme/>

Intro

Problems

Structure

Data

Search

SubStruc

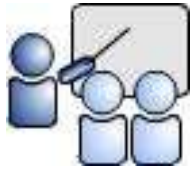
Thesaurus

Ontologies

Outlook

● **Exercises**

1. Find all informations concerning EC number 1.1.1.1
2. Browse (by enzyme class) to EC no. 1.7.1.10
What's the recommended name of EC no. 1.7.1.10 ?
3. How many distinct cofactors you find in ENZYME.
(BRENDA : 307)
Find all enzyme with flavin as a cofactor.
Compare with BRENDA information.



KEGG API

Intro

Problems

API definition :=

Structure

An application programming interface (API) is a set of definitions of the ways in which one piece of computer software communicates with another. It is a method of achieving abstraction, usually (but not necessarily) between lower-level and higher-level software.

Data

Search

SubStruc

KEGG API is a web service to use the KEGG system from your program via SOAP/WSDL.

Thesaurus

Ontologies

searching and retrieving data from the KEGG database.

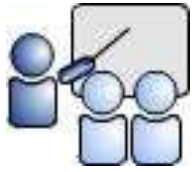
Outlook

Languages: Perl, Ruby, Python, Java

● **Exercises**

Manual:

http://www.genome.jp/kegg/docs/keggapi_manual.html



KEGG API

Intro

Problems

Structure

Data

Search

SubStruc

Thesaurus

Ontologies

Outlook

● Exercises

requirements for python- KEGG API:

fpconst
pyXML
SOAPpy

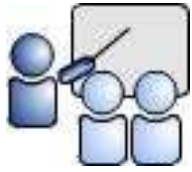
simple example:

```
#!/usr/bin/env python
```

```
from SOAPpy import WSDL
```

```
wsdl = 'http://soap.genome.jp/KEGG.wsdl'  
serv = WSDL.Proxy(wsdl)
```

```
results = serv.get_genes_by_pathway('path:eco00020')  
print results
```



KEGG API

Intro

Problems

Structure

Data

Search

SubStruc

Thesaurus

Ontologies

Outlook

METHODS

Meta information

`list_databases()`

List of database names and its definitions available

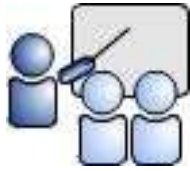
`list_organisms()`

List up the organisms in the KEGG/GENES database. 'org' code and the organism's full name

● Exercises

`list_pathways(org)`

List up the pathway maps of the given organism in the KEGG/PATHWAY database



KEGG API

Intro

Problems

Structure

Data

Search

SubStruc

Thesaurus

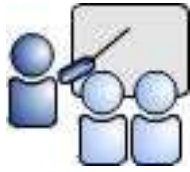
Ontologies

Outlook

● **Exercises**

Databases I:

aa	Non-redundant protein sequence database
gb	GenBank nucleic acid sequence database
gbu	Cumulative daily updates of GenBank since the latest release
rs	NCBI Reference Sequences
emb	EMBL nucleic acid sequence database
embu	Cumulative daily updates of EMBL since the latest release
est	dbEST database of Expressed Sequence Tags
gss	dbGSS database of Genome Survey Sequences
sts	dbSTS database of Sequence Taged Sites
htgs	HTGs database of High Throughput Genomic Sequences
sp	SWISS-PROT protein sequence database
sprel	SWISS-PROT protein sequence database
tr	TrEMBL protein sequence database
trnew	TrEMBL_new protein sequence database
pir	PIR protein sequence database
prf	PRF protein sequence database
gp	Translated protein sequences from genbank
gpu	Translated protein sequences from genbank-upd
rt	Translated protein sequences from refseq
pdb	RCSB Protein Data Bank
str	Re-organized Protein Data Bank
epd	Eukaryotic Promoter Database
ps	Dictionary of Protein Sites and Patterns
pdoc	PROSITE documentation file



KEGG API

Intro

Problems

Structure

Data

Search

SubStruc

Thesaurus

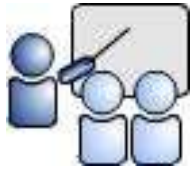
Ontologies

Outlook

● Exercises

Databases I:

bl	Blocks Database
pr	Protein Motif Fingerprint Database
pd	ProDom protein domain database
pf	Protein families database of alignments and HMMs
pmd	Protein Mutant Database
aax	Amino Acid Index Database, Combination of AAindex1 and AAindex2
lit	Literature Database compiled by PRF
mim	Online Mendelian Inheritance in Man
genes	KEGG Genes Database
gn	KEGG Genome Database
dgenes	KEGG Draft/Partial Genomes Genes Database
vg	KEGG Virus Genes Database
og	KEGG Organelle Genes Database
eg	KEGG EST Contigs Genes Database
ko	KEGG Orthology
path	KEGG Pathway Database
cpd	KEGG Ligand Chemical Database for Chemical Compound
gl	KEGG Ligand Chemical Database for Carbohydrate Structure
rn	KEGG Ligand Chemical Database for Chemical Reaction
rp	KEGG Ligand Chemical Database for Reactant Pairs
ec	KEGG Ligand Chemical Database for Enzyme Nomenclature
exp	KEGG Expression Database
ld	Database of Link Information



KEGG API

Intro

Problems

Structure

DBGET methods

Data

`binfo(database_id)`

Search

version information of the specified database

SubStruc

`bfind(database_id search_string)`

Thesaurus

searching entries by keywords

Ontologies

`bget(options dbname1:identifier1 [dbname2:identifier2...])`

Outlook

options:

-f fasta format

-n sequence number, 'a' for amino acid seq. or 'n' for nucleotide seq.

examples:

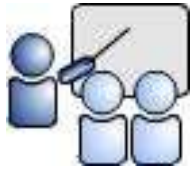
```
# retrieve nucleic acid sequences in a FASTA format
```

```
bget("-f -n n eco:b0002 hin:tRNA-Cys-1")
```

```
# retrieve amino acid sequence in a FASTA format
```

```
bget("-f -n a eco:b0002")
```

● Exercises



KEGG API

Intro

Problems

Pathway methods

Structure

get_genes_by_pathway(pathway_id)

all the genes on a specified pathway

example:

```
# Returns all the E. coli genes on the pathway map '00020'.
```

Search

```
get_genes_by_pathway('path:eco00020')
```

SubStruc

get_enzymes_by_pathway(pathway_id)

all enzymes on the specified pathway

example:

```
# Returns all the enzymes on the pathway map '00020'.
```

Ontologies

```
get_enzymes_by_pathway('path:eco00020')
```

Outlook

get_compounds_by_pathway(pathway_id)

compounds on the specified pathway.

example:

```
# Returns all the compounds on the pathway map '00020'.
```

```
get_compounds_by_pathway('path:eco00020')
```

get_reactions_by_pathway(pathway_id)

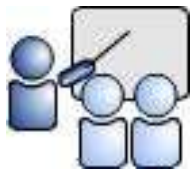
all reaction_ids on the specified pathway

example:

```
# Returns all the reactions on the pathway map '00260'
```

```
get_reactions_by_pathway('path:map00260')
```

● Exercises



KEGG API

Intro

Problems

Structure

Data

Search

SubStruc

Thesaurus

Ontologies

Outlook

● Exercises

Pathway methods II - Pathways by objects

get_pathways_by_genes(genes_id_list)

```
get_pathways_by_genes(['eco:b0077', 'eco:b0078'])
```

get_pathways_by_enzymes(enzyme_id_list)

```
get_pathways_by_enzymes(['ec:1.3.99.1'])
```

get_pathways_by_compounds(compound_id_list)

```
get_pathways_by_compounds(['cpd:C00033', 'cpd:C00158'])
```

get_pathways_by_reactions(reaction_id_list)

```
get_pathways_by_reactions(['rn:R00959', 'rn:R02740', 'rn:R00960'])
```

Pathway methods IV - Relation among enzymes, compounds and reactions

get_enzymes_by_compound(compound_id)

```
get_enzymes_by_compound('cpd:C00345')
```

get_enzymes_by_reaction(reaction_id)

```
get_enzymes_by_reaction('rn:R00100')
```

get_compounds_by_enzyme(enzyme_id)

```
get_compounds_by_enzyme('ec:2.7.1.12')
```

get_compounds_by_reaction(reaction_id)

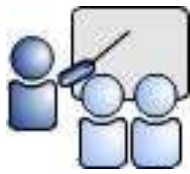
```
get_compounds_by_reaction('rn:R00100')
```

get_reactions_by_enzyme(enzyme_id)

```
get_reactions_by_enzyme('ec:2.7.1.12')
```

get_reactions_by_compound(compound_id)

```
get_reactions_by_compound('cpd:C00199')
```



KEGG API

Intro

Problems **Pathway methods III - Relation among pathways**

Structure `get_linked_pathways(pathway_id)`
`get_linked_pathways('path:eco00620')`

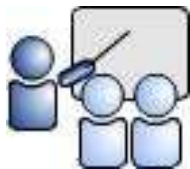
Data **Genes**
`get_genes_by_organism(org, start, max_results)`
Search `get_genes_by_organism('hin', 1, 100)`

SubStruc **Genomes**
`get_number_of_genes_by_organism(org)`
Thesaurus `get_number_of_genes_by_organism('eco')`

Ontologies

Outlook

● **Exercises**



python |

Intro

Data types

Commentary: `#commentary`

Problems

String: `a="dog"`

`b="cat"`

Structure

`#combine to strings`

`c=a+b`

Data

`d=a+" and "+b`

`a.startswith("do")`

Search

List: `list_a=["first","second"]`

SubStruc

Dictionary: `dict_dog={"english":"dog","german":"Hund"}`

Thesaurus

`print dict_dog["german"]`

Ontologies

if / elif / else statement:

`today="Sunday"`

`if today=="Saturday" or today=="Sunday":`

`print "weekend."`

`elif variable_a=="Monday":`

`print "beginning of the week"`

`else:`

`print "working"`

Outlook

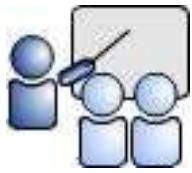
● Exercises

for statement:

`list_a=["first","second"]`

`for x in list_a:`

`print x`



python Exercises

Intro

Problems

Structure

Data

Search

SubStruc

Thesaurus

Ontologies

Outlook

● Exercises



preparation:

1. Open a shell (linux) or putty(windows)
2. ssh login with: `ssh euroschool@134.95.151.45`
3. enter password
4. open python shell with: `python`



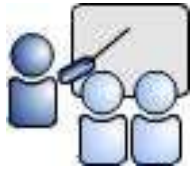
STRINGS

- a.) save the string „Hello“ and „world“ respectively in a string variable (a,b)
- b.) save the combination of both in a new variable (d) separated by a blank
- c.) print variable c



LISTS

- a.) Save the compounds ATP , ADP and SDS in a list
- b.) Print the list element by element (use 'for' statement)



KEGG

Intro

Problems

webinterface

Structure

1. open KEGG Protein Network

Data

2. Find in Glycolysis / Gluconeogenesis Fructose-bisphosphate aldolase (EC 4.1.2.13)

Search

3. How many genes for EC 4.1.2.13 in homo sapiens exists?

4. What are substrates and products?

SubStruc

KEGG-API

Thesaurus

5. Find all

Ontologies

a.) genes `01_genes_in_homo_sapiens_gylcolysis.py`

b.) enzymes `02_enzymes_in_homo_sapiens_gylcolysis.py`

Outlook

c.) compounds `03_compounds_in_homo_sapiens_gylcolysis.py`

d.) reactions `04_reactions_in_homo_sapiens_gylcolysis.py`

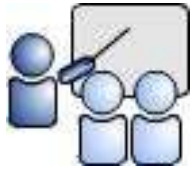
in homo sapiens Glycolysis / Gluconeogenesis

Tipp: Organism-ID = hsa

Pathway-ID = 00010

use KEGG-API

● Exercises



KEGG

Intro

Problems

Structure

Data

Search

SubStruc

Thesaurus

Ontologies

Outlook

● **Exercises**

<http://www.genome.jp/kegg/>

KEGG-API

6. Find all compounds IDs of oxidoreductases which involved in E.coli Glycolysis

Tip: Organism-ID=eco

Path-ID=00010

pathid=Organism-ID+Path-ID

ecid="ec:"+EC_Number

methods you need :

`string.startswith()`

`get_enzymes_by_pathway(path:pathid)`

`get_compounds_by_enzyme(ecid)`

`05_all_Oxidoreductases_of_Glycolysis_and_in_other_pathways.py`

7. Find all reactions IDs and informations of L-arabinitol 4-dehydrogenase (EC 1.1.1.12) and the corresponding compounds IDs and informations

`06_Compounds_of_EC_1_1_1_12.py`

acknowledgment

Helmholz Gesellschaft

Dietmar Schomburg

Ida Schomburg

Antje Chang

Gregor Huhn

Marion Gremse

Christian Held

Oliver Hoffmann

many Coworkers

references

***Nucleic Acids Res.* 2004 Jan 1;32**

BRENDA, the enzyme database: updates and major new developments

***Metabolic Engineering* 5 (2003); 71-73**

Review of BRENDA database

***Science* 2002 October 25; 298: 709 (in NetWatch)**

DATABASE: Whole Enzyme Catalog

***Trends Biochem Sci* 2002 Jan;27(1):54-6**

BRENDA: a resource for enzyme data and metabolic information.

Schomburg I, Chang A, Hofmann O, Ebeling C, Ehrentreich F, Schomburg D.

***Nucleic Acids Res* 2002 Jan 1;30(1):47-9**

BRENDA, enzyme data and metabolic information.

Schomburg I, Chang A, Schomburg D.