Metabolomics Workbench REST URL-based API Specification

Version: 1.0 Date: 5/7/2019

This document describes the Metabolomics Workbench REST API specifications, a web interface for accessing a variety of data such as metabolite structures, study metadata, experimental results etc. It details the syntax of the HTTP requests including both the names of the available requests and parameters. These HTTP requests may be carried out using a web browser or may be embedded in 3rd party applications or scripts to enable programmatic access. Most modern programming languages including PHP, Perl, Python, Java and Javascript have the capability to create HTTP requests and interact with datasets through the REST API.

The URL Path

The REST URL consists of three main parts, separated by forward slashes, after the common prefix specifying the invariant base URL (https://www.metabolomicsworkbench.org/rest/):

https://www.metabolomicsworkbench.org/rest/<context>/<input specification>/<output specification>

Part 1: The **context** determines the type of data to be accessed from the Metabolomics Workbench, such as metadata or results related to the submitted studies, data from metabolites, genes/proteins and analytical chemistry databases as well as other services related to mass spectrometry and metabolite identification:

<context> = study | compound | refmet | gene | protein | moverz | exactmass

Part 2: The **input specification** consists of two required parameters describing the REST request:

<input specification> = <input item>/<input value>

Part 3: The **output specification** consists of two parameters describing the output generated by the REST request:

<output specification> = <output item>/[<output format>]

The first parameter is required in most cases. The second parameter is optional. The input and output specifications are context sensitive. The context determines the values allowed for the remaining parameters in the input and output specifications as detailed in the sections below.

The “study” context

The "**study**" context refers to the studies available in the Metabolomics Workbench (www.metabolomicsworkbench.org), a public repository for metabolomics metadata and experimental data spanning various species and experimental platforms, metabolite standards, metabolite structures, protocols, tutorials and training material, and other educational resources. It provides a computational platform to integrate, analyze, track, deposit, and disseminate large volumes of heterogeneous data from a wide variety of metabolomics studies including Mass Spectrometry (MS) and Nuclear Magnetic Resonance (NMR) spectrometry data spanning a variety of species covering all the major taxonomic categories including humans and other mammals, plants, insects, invertebrates, and microorganisms. This context provides access to a variety of data associated with studies such as study summary, experimental factors for study design, analysis information, metabolites and results data, sample source and species etc.

<context> = study

<input item> = study_id | study_title | institute | last_name | analysis_id | metabolite_id

<input value> = <input item value>

<output item> = summary | factors | analysis | metabolites | mwtab | source | species | disease | number_of_metabolites | data | datatable | untarg_studies | untarg_factors | untarg_data

<output format> = txt | json [Default: json]

The “**summary**” output item retrieves the following information for a specified input name and value: study_id, study_title, study_type, institute, department, last_name, first_name, email, phone, submit_date, study_summary, subject_species.

https://www.metabolomicsworkbench.org/rest/study/study_id/ST000001/summary

A substring may be specified for the study_id input value. For example, using “ST” will retrieve the summary information for all studies, and using “ST0004” will retrieve summary information for studies ST000400 to ST000499.

https://www.metabolomicsworkbench.org/rest/study/study_id/ST/summary

https://www.metabolomicsworkbench.org/rest/study/study_id/ST0004/summary

The “**factors**” output item refers to the experimental conditions for each sample in the study and retrieves the following information for a specified input name and value: study_id, local_sample_id, subject_type, factors. For example:

https://www.metabolomicsworkbench.org/rest/study/study_id/ST000001/factors

The “**analysis**” output item accesses a number of key instrumentation parameters and retrieves the following information for a specified input name and value: study_id, analysis_id, analysis_summary, analysis_type, MS instrument_name, MS instrument_type, MS type, MS ion_mode, NMR instrument_type, NMR experiment_type, NMR spectrometer_frequency, NMR solvent. For example:

https://www.metabolomicsworkbench.org/rest/study/last_name/Kind/analysis

The “**metabolites**” output item exposes details for each named metabolite in a particular study or analysis. The input item must be either a study_id or analysis_id. The REST request retrieves the following information for the specified study/analysis ID: study_id, analysis_id, analysis summary, metabolite_name, refmet_name, pubchem_id, other_id, other_id_type. For example:

https://www.metabolomicsworkbench.org/rest/study/study_id/ST000001/metabolites

https://www.metabolomicsworkbench.org/rest/study/analysis_id/AN000001/metabolites

The “**study_id**” or “analysis_id” input item is required for the “**mwtab**” output item. The mwTab file may be downloaded in JSON or txt format. For example:

https://www.metabolomicsworkbench.org/rest/study/study_id/ST000001/mwtab

The “**source**” output item retrieves the following information for the specified study: Study ID, Sample source (e.g. blood, urine, liver, etc.). For example:

https://www.metabolomicsworkbench.org/rest/study/study_id/ST000001/source

retrieves the sample source information for all study ST000001.

https://www.metabolomicsworkbench.org/rest/study/study_id/ST/source

retrieves the sample source information for all studies using “ST” as a wildcard.

The “**species**” output item retrieves the following information for the specified study: Study ID, Latin name, Common name. For example:

https://www.metabolomicsworkbench.org/rest/study/study_id/ST000001/species

The “**disease**” output item retrieves the following information for the specified study: Study ID, Disease. For example:

https://www.metabolomicsworkbench.org/rest/study/study_id/ST000007/disease

The “**number_of_metabolites**” output item retrieves the following information for a specified study, study title, last name or institute: study_id, analysis_id, study title, number of metabolites, analysis summary. For example:

https://www.metabolomicsworkbench.org/rest/study/study_id/ST000001/number_of_metabolites

https://www.metabolomicsworkbench.org/rest/study/study_title/Diabetes/number_of_metabolites

https://www.metabolomicsworkbench.org/rest/study/institute/Michigan/number_of_metabolites

The “**study_id**” or “**analysis_id**” input item is required for the “**data**” output item. The REST request retrieves the following information for the specified study ID : study_id, analysis_id, analysis_summary, metabolite_name, metabolite_id, refmet_name, units. In addition, the following results information is retrieved for each metabolite: local sample_ID and measured values. For example:

https://www.metabolomicsworkbench.org/rest/study/study_id/ST000001/data

The “**analysis_id**” input item is required for the “**datatable**” output item. The value of the output format is ignored. A tab delimited data table is generated containing available results for the metabolites. The header line includes sample IDs, class, and metabolite names. For example:

https://www.metabolomicsworkbench.org/rest/study/analysis_id/AN000001/datatable

The “**untarg_studies**” output item retrieves the following information for all untargeted MS studies in the data repository: study_id, analysis_id, analysis_summary, study_title, subject_species, institute. The input item and value parameter are ignored, but a study_id “placeholder” must specified to create a valid REST request. For example:

https://www.metabolomicsworkbench.org/rest/study/study_id/X/untarg_studies

The “**analysis_id**” input item is required for the “**untarg_factors**” output item. The “untarg_factors” output item retrieves a listing of the experimental conditions (factors) for untargeted MS studies. If more than one factor is present the factors are separated by “|” symbols. The integer at the end of each factor grouping indicates the number of sample replicates for that group. The output is displayed in JSON format. For example:

https://www.metabolomicsworkbench.org/rest/study/analysis_id/AN000113/untarg_factors

The “**analysis_id**” input item is required for the “**untarg_data**” output item. The “**untarg_data**” output item retrieves the table of measurements for the selected analysis. The untargeted data is downloaded as a tab delimited text file. For example:

https://www.metabolomicsworkbench.org/rest/study/analysis_id/AN000113/untarg_data

The “**metabolite_id**” input item retrieves the following information for a Metabolomics Workbench metabolite ID in a study: metabolite_id, metabolite_name, refmet_name, pubchem_id, other_id, other_id_type, kegg_id, ri, ri_type, moverz_quant. The output item is ignored, but an output item “placeholder” must be specified to create a valid REST request. For example:

https://www.metabolomicsworkbench.org/rest/study/metabolite_id/ME000096/summary

Description of input and output items

A brief description of input and output items in “**study**” context, along with representative examples, is provided in the following table:

Input item: study_id
Input value: Metabolomics Workbench (MW) study ID
Input value type: ST<6-digit integer>
Input example: ST000001
Output item: summary | factors | analysis | metabolites |mwtab | source |species | disease | number_of_metabolites | data
Output format: json | txt

Input item: study_title
Input value: Title of a study
Input value type: string
Input example: Diabetes
Output item: summary | factors | analysis | number_of_metabolites | source |species | disease
Output format: json | txt

Input item: institute
Input value: Name of an institute for a study
Input value type: string
Input example: Michigan
Output item: summary | factors | analysis | number_of_metabolites | source |species | disease
Output format: json | txt

Input item: last_name
Input value: Last name of an investigator for a study
Input value type: string
Input example: Kind
Output item: summary | factors | analysis | number_of_metabolites | source |species | disease
Output format: json | txt

Input item: analysis_id
Input value: Metabolomics Workbench analysis ID for a study
Input value type: AN<6-digit integer>
Input example: AN000001

Output item: mwtab | metabolites | datatable | untarg_factors | untarg_data
Output format: txt file for datatable and untarg_data; json for untargeted_factors; json | txt for others

Input item: metabolite_id
Input value: Metabolomics Workbench metabolite ID for a study
Input value type: ME<6-digit integer>
Input example: ME000096
Output item: summary (ignored but needed as a placeholder)
Output format: json | txt

Examples

A representative set of RSET URLs examples for "study" context along with the description of the request are shown below:

Request: Fetch summary information for a study

URL: https://www.metabolomicsworkbench.org/rest/study/study_id/ST000001/summary

Request: Fetch samples and experimental variables (factors) for a study

URL: https://www.metabolomicsworkbench.org/rest/study/study_id/ST000001/factors

Request: Fetch information for a metabolite_id in a study

URL: https://www.metabolomicsworkbench.org/rest/study/metabolite_id/ME000096/summary

Request: Fetch summary information for all studies

URL: https://www.metabolomicsworkbench.org/rest/study/study_id/ST/summary

Request: Fetch analysis information for a study

URL: https://www.metabolomicsworkbench.org/rest/study/study_id/ST000001/analysis

Request: Fetch metabolites and annotations detected in a study (one study at a time)

URL: https://www.metabolomicsworkbench.org/rest/study/study_id/ST000009/metabolites

Request: Fetch metabolites measurements for a study (one study at a time)

URL: https://www.metabolomicsworkbench.org/rest/study/study_id/ST000001/data

Request: Fetch mwTab output for a study (one study at a time)

URL: https://www.metabolomicsworkbench.org/rest/study/study_id/ST000001/mwtab

Request: Fetch species information as text for all studies

URL: https://www.metabolomicsworkbench.org/rest/study/study_id/ST/species/txt

Request: Fetch sample source information as text for all studies

URL: https://www.metabolomicsworkbench.org/rest/study/study_id/ST/source/txt

Request: Fetch disease association (where applicable) for all studies

URL: https://www.metabolomicsworkbench.org/rest/study/study_id/ST/disease

study context: Table of input name, output name and output field permutations

Match	Input_name	Output_name	Output fields
exact match	metabolite_id	all (ignored)	metabolite_id,metabolite_name,refmet_name, pubchem_id,other_id,other_id_type,kegg_id, ri,ri_type,moverz_quant
substring match	study_id|study_title|institute|last_name	number_of_metabolites	study_id,analysis_id,num_metabolites, analysis_display,study_title
substring match	study_id|study_title|institute|last_name	summary	study_id,study_title,study_type,institute, department,last_name,first_name,email,phone, submit_date,study_summary,subject_species
substring match	study_id|study_title|institute|last_name	factors	study_id,local_sample_id, subject_type,factors
substring match	study_id|study_title|institute|last_name	analysis	study_id,analysis_id,analysis_summary,analysis_type, ms_instrument_name,ms_instrument_type,ms_type, ion_mode,nmr_instrument_type,nmr_experiment_type, nmr_spectrometer_frequency,nmr_solvent
exact match	study_id|analysis_id	metabolites	study_id,analysis_id, analysis_summary,metabolite_name,refmet_name ,pubchem_id,other_id,other_id_type
exact match	study_id|analysis_id	data	study_id,analysis_id,analysis_summary, metabolite_name,metabolite_id,refmet_name, units,local_sample_id,measured_value
exact match	study_id|analysis_id	mwtab	mwtab
substring match	study_id|study_title|institute|last_name	species	study_id,subject_species ,species_common
substring match	study_id|study_title|institute|last_name	source	study_id,source
substring match	study_id|study_title|institute|last_name	disease	study_id,disease
NONE	study_id (ignored)	untarg_studies	study_id,analysis_id,analysis_display, study_title, subject_species, institute
exact match	analysis_id	untarg_factors	factors
exact match	analysis_id	untarg_data	data table
exact match	analysis_id	datatable	Tab-delimited table of metabolite measurements

The “compound” context

The “**compound**” context provides services for the Metabolomics Workbench Metabolite Database which contains structures and annotations of biologically relevant metabolites. The database contains over 64,000 entries, collected from public repositories such as LIPID MAPS, ChEBI, HMDB, BMRB, PubChem, and KEGG, as well as from literature sources. This context provides access to many structural features including molfile, SMILES, InChIKey, exact mass, formula common and systematic names, chemical classification and cross-references to other databases.

<context> = compound

<input item> = regno | formula | inchi_key | lm_id | pubchem_cid | hmdb_id | kegg_id | chebi_id | metacyc_id | abbrev

<input value> = <input item value>

<output item> = all | regno | formula | exactmass | inchi_key | name | sys_name | smiles | lm_id | pubchem_cid | hmdb_id | kegg_id | chebi_id | metacyc_id | classification | molfile | png | regno,formula,exactmass,...

<output format> = txt | json [Default: json]

Only one input item and input value may be specified in the REST URL.

The “**regno**” input item refers to the Metabolomics Workbench Metabolite database internal identifier. It specifies a unique metabolite structure. For example:

https://www.metabolomicsworkbench.org/rest/compound/regno/34361/name

Multiple items from the following list may be specified as output by placing commas between the items: regno, formula, exactmass, inchi_key, name, sys_name, lm_id, pubchem_cid, hmdb_id, kegg_id, chebi_id, metacyc_id, smiles. For example an output item **name,formula,exactmass** will display those 3 fields in the REST output. For example:

https://www.metabolomicsworkbench.org/rest/compound/regno/34361/name,formula,exactmass

The “**all**” output item is automatically expanded to include the following items: regno, formula, exactmass, inchi_key, name, sys_name, lm_id, pubchem_cid, hmdb_id, kegg_id, chebi_id, metacyc_id, smiles. These output items should not be individually specified with the “all” output item. For example:

https://www.metabolomicsworkbench.org/rest/compound/regno/34361/all

The “**classification**” output item is automatically expanded to include the following items: regno, name, sys_name, cf_superclass, cf_class, cf_subclass, cf_direct_parent, cf_alternative_parents, lm_category, lm_main_class, lm_sub_class, lm_class_level4. These output items should not be individually specified with the “classification” output item. The “cf” and “lm” correspond to ClassyFire and LIPID MAPS classification systems respectively. For example:

https://www.metabolomicsworkbench.org/rest/compound/pubchem_cid/5997/classification

The “**regno**” input item is required for the “**png**” output item. No other comma delimited output items are allowed alongside “**png**”. The PNG image is displayed in the browser. For example:

https://www.metabolomicsworkbench.org/rest/compound/regno/34361/png

The “**regno**” input item is required for the “**molfile**” output item. No other comma delimited output items are allowed with alongside “**molfile**”. The user is given the option to download the molfile as a text file. For example:

https://www.metabolomicsworkbench.org/rest/compound/regno/34361/molfile

The “**abbrev**” input item specifies a “bulk” lipid abbreviation such as Cer(d32:1) or PC(38:4) The “**classification**” output item must be specified and will display the following fields: lm_category, lm_main_class, lm_sub_class. Any other output item is not allowed. For example:

https://www.metabolomicsworkbench.org/rest/compound/abbrev/PC(34:1)/classification

Description of input and output items

A brief description of input and output items in “**compound**” context, along with representative examples, is provided in the following table:

Input item: regno
Input value: Metabolomics Workbench Metabolite database ID
Input value type: integer
Input example: 11
Output item: all | any, some or all of: regno, formula, exactmass, inchi_key, name, sys_name, smiles, lm_id, pubchem_cid, hmdb_id, kegg_id, chebi_id, metacyc_id, classification
Output format: json | txt

Input item: formula
Input value: Molecular Formula
Input value type: string
Input example: C20H34O11
Output item: all | any, some or all of: regno, formula, exactmass, inchi_key, name, sys_name, smiles, lm_id, pubchem_cid, hmdb_id, kegg_id, chebi_id, metacyc_id, classification
Output format: json | txt

Input item: inchi_key
Input value: valid InChIKey
Input value type: 27-character string
Input example: JTWQQJDENGGSBJ-UHFFFAOYSA-N
Output item: all | any, some or all of: regno, formula, exactmass, inchi_key, name, sys_name, smiles, lm_id, pubchem_cid, hmdb_id, kegg_id, chebi_id, metacyc_id, classification
Output format: json | txt

Input item: lm_id
Input value: LIPID MAPS ID
Input value type: LM<2-character LIPID MAPS category><8-10 character string>
Input example: LMFA03010001

Output item: all | any, some or all of: regno, formula, exactmass, inchi_key, name, sys_name, smiles, lm_id, pubchem_cid, hmdb_id, kegg_id, chebi_id, metacyc_id, classification
Output format: json | txt

Input item: pubchem_cid
Input value: PubChem Compound ID
Input value type: integer
Input example: 52921723
Output item: all | any, some or all of: regno, formula, exactmass, inchi_key, name, sys_name, smiles, lm_id, pubchem_cid, hmdb_id, kegg_id, chebi_id, metacyc_id, classification
Output format: json | txt

Input item: hmdb_id
Input value: Human Metabolome Database ID
Input value type: HMDB<integer>
Input example: HMDB0002886
Output item: all | any, some or all of: regno, formula, exactmass, inchi_key, name, sys_name, smiles, lm_id, pubchem_cid, hmdb_id, kegg_id, chebi_id, metacyc_id, classification
Output format: json | txt

Input item: kegg_id
Input value: KEGG compound ID
Input value type: CO<integer>
Input example: C05961
Output item: all | any, some or all of: regno, formula, exactmass, inchi_key, name, sys_name, smiles, lm_id, pubchem_cid, hmdb_id, kegg_id, chebi_id, metacyc_id, classification
Output format: json | txt

Input item: chebi_id
Input value: ChEBI compound id
Input value type: integer
Input example: 30805
Output item: all | any, some or all of: regno, formula, exactmass, inchi_key, name, sys_name, smiles, lm_id, pubchem_cid, hmdb_id, kegg_id, chebi_id, metacyc_id, classification
Output format: json | txt

Input item: metacyc_id
Input value: METACYC compound ID
Input value type: string
Input example: CPD-7836
Output item: all | any, some or all of: regno, formula, exactmass, inchi_key, name, sys_name, smiles, lm_id, pubchem_cid, hmdb_id, kegg_id, chebi_id, metacyc_id, classification
Output format: json | txt

Input item: abbrev
Input value: Lipid bulk abbreviation
Input value type: string
Input example: LPC(18:0)
Output item: classification
Output format: json | txt

Input item: regno
Input value: Metabolomics Workbench Metabolite database ID (integer)

Input value type: integer
Input example: 11
Output item: png
Output format: png image (By default; No specification allowed.)

Input item: regno
Input value: Metabolomics Workbench Metabolite database ID (integer)
Input value type: integer
Input example: 11
Output item: molfile
Output format: downloadable text file (By default; No specification allowed.)

Examples

A representative set of REST URLs examples for the "**compound**" context along with the description of the request are shown below:

Request: Fetch compound common name from regno

URL: https://www.metabolomicsworkbench.org/rest/compound/regno/11/name

Request: Fetch all compound fields from regno

URL: https://www.metabolomicsworkbench.org/rest/compound/regno/11/all

Request: Fetch all compound fields as text from regno

URL: https://www.metabolomicsworkbench.org/rest/compound/regno/11/all/txt

Request: Fetch compound smiles from PubChem CID

URL: https://www.metabolomicsworkbench.org/rest/compound/pubchem_cid/52921723/smiles

Request: Fetch compound common name and systematic name from inchi key

URL: https://www.metabolomicsworkbench.org/rest/compound/inchi_key/JTWQQJDENGGSBJ-UHFFFAOYSA-N/name,sys_name

Request: Fetch all compound fields with a given molecular formula (multiple records)

URL: https://www.metabolomicsworkbench.org/rest/compound/formula/C20H34O/all

Request: Fetch compound classification hierarchy for a PubChem CID

URL: https://www.metabolomicsworkbench.org/rest/compound/pubchem_cid/5281365/classification

Request: Download compound molfile for a regno

URL: https://www.metabolomicsworkbench.org/rest/compound/regno/28606/molfile

Request: Fetch png image of structure for a regno

URL: https://www.metabolomicsworkbench.org/rest/compound/regno/11/png

compound context: Table of input name, output name and output field permutations

Match	Input_name	Output_name	Output fields
exact match	regno|formula|inchi_key|lm_id|pubchem_cid|hmdb_id|kegg_id|chebi_id|metacyc_id|smiles|abbrev	all	regno,formula,exactmass,inchi_key,name,sys_name,smiles,lm_id,pubchem_cid,hmdb_id,kegg_id,chebi_id,metacyc_id
exact match	regno|formula|inchi_key|lm_id|pubchem_cid|hmdb_id|kegg_id|chebi_id|metacyc_id|smiles|abbrev	*Any, some or all of: regno,formula,exactmass,inchi_key,name,sys_name,smiles,lm_id,pubchem_cid,hmdb_id,kegg_id,chebi_id,metacyc_id	Any, some or all of: regno,formula,exactmass,inchi_key,name,sys_name,smiles,lm_id,pubchem_cid,hmdb_id,kegg_id,chebi_id,metacyc_id
exact match	regno|formula|inchi_key|lm_id|pubchem_cid|hmdb_id|kegg_id|chebi_id|metacyc_id|smiles|abbrev	classification	regno,name,sys_name,cf_superclass,cf_class,cf_subclass,cf_direct_parent,cf_alternative_parents,lm_category,lm_main_class,lm_sub_class,lm_class_level4
exact match	regno	molfile	Downloadable molfile (text)
exact match	regno	png	PNG image of compound

*multiple output names must be separated by commas, e.g. name,formula,exactmass

The “refmet” context

The “refmet” context refers to a standardized reference nomenclature for both discrete metabolite structures and metabolite species identified by spectroscopic techniques in metabolomics experiments. This is an essential prerequisite for the ability to compare and contrast metabolite data across different experiments and studies. The use of identifiers such as PubChem compound IDs and InChiKeys offers only a partial solution because these identifiers will vary depending on parameters such as the salt form and degree of stereochemical detail. In addition, many metabolite species, especially lipids, are not reported by MS methods as discrete structures but rather as isobaric mixtures (such as PC(34:1) and TG(54:2)). To this end, a list of over 160,000 names from a set of over 800 MS and NMR studies on the Metabolomics Workbench has been used as a starting point to generate a highly curated analytical chemistry-centric list of common names for metabolite structures and isobaric species. Additionally, the vast majority of these names have been linked to a metabolite classification system using a combination of LIPID MAPS and ClassyFire classification methods. A name-conversion user interface is provided where users can submit a list of metabolite names and map them to the corresponding Refmet names. This is a work-in-progress with the caveat that many metabolite names generated by metabolomics experiments will not currently map to RefMet identifiers. Nevertheless, RefMet has the ability to greatly increase the data-sharing potential of metabolomics experiments and facilitate "meta-analysis" and systems biology objectives for the majority of commonly encountered metabolite species.

This context provides access to many structural features including InChIKey, exact mass, formula common and systematic names, chemical classification and cross-references to other database

<context> = refmet

<input item> = all | match | name | inchi_key | regno | pubchem_cid | formula | main_class | sub_class

<input value> = <input item value>

<output item> = all | name | inchi_key | regno | pubchem_cid | exactmass | formula | synonyms | sys_name | main_class | sub_class | name,inchi_key,regno,...

<output format> = txt | json

The “**all**” input item retrieves all available data in JSON format for all the entries in the RefMet. The rest of the input and output parameters are not required. For example:

https://www.metabolomicsworkbench.org/rest/refmet/all

The “**all**” output item is automatically expanded to include the following items: name, regno, pubchem_cid, inchi_key, exactmass, formula, sys_name, main_class, sub_class, synonyms. These output items should not be specified with “**all**” output item. The input value and output items are ignored. For example:

https://www.metabolomicsworkbench.org/rest/refmet/name/Cholesterol/all

The "**match**" input item performs a search against a customized synonym table in the database. The submitted synonyms are matched in a 'fuzzy' manner by dropping the following types of characters from the specified input value: <space>_+-/(){}[]*";@. In addition, some common ion adduct suffixes (e.g. [M+H]+) are removed. The output item is ignored and the following output information is retrieved: refmet_name, formula, exact mass, main class, sub class. For example:

https://www.metabolomicsworkbench.org/rest/refmet/match/LysoPC16:0

Description of input and output items

A brief description of input and output items in "**refmet**" context, along with representative examples, is provided in the following table:

Input item: all
Input value: none
Input example: none
Output item: none; automatically retrieves all (name, regno, pubchem_cid, inchi_key, exactmass, formula, sys_name, main_class, sub_class, synonyms)
Output format: json

Input item: match
Input value: Character string for a synonym match
Input value type: string
Input example: Lyso PC (16:0)
Output item: none; automatically retrieves refmet_name, formula, exactmass, main_class, sub_class
Output format: json

Input item: name
Input value: Compound name
Input value type: string
Input example: Cholesterol
Output item: all | any, some or all of: name, regno, pubchem_cid, inchi_key, exactmass, formula, sys_name, main_class, sub_class, synonyms
Output format: json | txt

Input item: inchi_key
Input value: valid InChIKey
Input value type: 27-character string
Input example: HVYWMOMLDIMFJA-DPAQBDIFSA-N
Output item: all | any, some or all of: name, regno, pubchem_cid, inchi_key, exactmass, formula, sys_name, main_class, sub_class, synonyms
Output format: json | txt

Input item: regno
Input value: Metabolomics Workbench Metabolite database ID (integer)

Input value type: integer
Input example: 11
Output item: all | any, some or all of: name, regno, pubchem_cid, inchi_key, exactmass, formula, sys_name, main_class, sub_class, synonyms
Output format: json | txt

Input item: pubchem_cid
Input value: PubChem Compound ID
Input value type: integer
Input example: 5997
Output item: all | any, some or all of: name, regno, pubchem_cid, inchi_key, exactmass, formula, sys_name, main_class, sub_class, synonyms
Output format: json | txt

Input item: formula
Input value: Molecular Formula
Input value type: string
Input example: C27H46O
Output item: all | any, some or all of: name, regno, pubchem_cid, inchi_key, exactmass, formula, sys_name, main_class, sub_class, synonyms
Output format: json | txt

Input item: main_class
Input value: Refmet main class
Input value type: string
Input example: Sterols
Output item: all | any, some or all of: name, regno, pubchem_cid, inchi_key, exactmass, formula, sys_name, main_class, sub_class, synonyms
Output format: json | txt

Input item: sub_class
Input value: Refmet sub class
Input value type: string
Input example: Cholesterol and derivatives
Output item: all | any, some or all of: name, regno, pubchem_cid, inchi_key, exactmass, formula, sys_name, main_class, sub_class, synonyms
Output format: json | txt

Examples

A representative set of RSET URLs examples for "**refmet**" context along with the description of the request are shown below:

Request: Fetch all REFMET fields from name

URL: https://www.metabolomicsworkbench.org/rest/refmet/name/Cholesterol/all

Request: Fetch all REFMET fields from formula

URL: https://www.metabolomicsworkbench.org/rest/refmet/formula/C12H24O2/all

Request: Fetch all REFMET fields for all Sterols

URL: https://www.metabolomicsworkbench.org/rest/refmet/main_class/Sterols/all

refmet context: Table of input name, output name and output field permutations

<table>
<tr><th>Match</th><th>Input_name</th><th>Output_name</th><th>Output fields</th></tr>
<tr><td><b>exact match</b></td><td>name|regno|pubchem_cid|inchi_key||formula|main_class|sub_class</td><td>all</td><td>name,sys_name,synonyms,regno,pubchem_cid,inchi_key,exactmass,formula,main_class,sub_class</td></tr>
<tr><td><b>exact match</b></td><td>name|regno|pubchem_cid|inchi_key||formula|main_class|sub_class</td><td>*Any, some or all of:
name|sys_name|synonyms|regno|pubchem_cid|inchi_key|exactmass|formula|main_class|sub_class</td><td>Any, some or all of:
name|sys_name|synonyms|regno|pubchem_cid|inchi_key|exactmass|formula|main_class|sub_class</td></tr>
<tr><td><b>exact match</b></td><td>match</td><td>IGNORED</td><td>name,formula,exactmass,main_class,sub_class</td></tr>
</table>

*multiple output names must be separated by commas, e.g. name,formula,exactmass

The “gene” context

The “**gene**” context refers to a Human Metabolome Gene/Protein Database (MGP) of metabolome-related genes and proteins contains data for over 7300 genes and over 15,500 proteins. It provides access to gene related information such as MGP ID, gene ID and, symbols, gene names and synonyms, alternate gene names, taxonomy ID, species, etc.

<context> = gene

<input item> = mgp_id | gene_id | gene_name | gene_symbol | taxid

<input value> = <input item value>

<output item> = all | lmp_id | mgp_id | gene_name | gene_symbol | gene_synonyms | alt_names | chromosome | map_location | summary | taxid | species | species_long | mgp_id,gene_id,gene_name,...

<output format> = txt | json

The “**all**” output item is automatically expanded to include the following items: mgp_id, gene_id, gene_name, gene_symbol, gene_synonyms, alt_names, chromosome, map_location, summary, taxid, species, species_long. These output items should not be specified with “**all**” output item. For example:

https://www.metabolomicsworkbench.org/rest/gene/gene_id/31/all

Description of input and output items

A brief description of input and output items in “**gene**” context, along with representative examples, is provided in the following table:

Input item: mgp_id
Input value: Human Metabolome Gene/Protein (MGP) database gene ID
Input value type: MGP<6-digit integer>
Input example: MGP000016
Output item: all | any, some or all of: mgp_id, gene_id, gene_name, gene_symbol, gene_synonyms, alt_names, chromosome, map_location, summary, taxid, species, species_long
Output format: json | txt

Input item: gene_id
Input value: Entrez gene ID
Input value type: integer
Input example: 31
Output item: all | any, some or all of: mgp_id, gene_id, gene_name, gene_symbol, gene_synonyms, alt_names, chromosome, map_location, summary, taxid, species, species_long
Output format: json | txt

Input item: gene_name
Input value: Gene name
Input value type: string

Input example: acetyl-CoA carboxylase
Output item: all | any, some or all of: mgp_id, gene_id, gene_name, gene_symbol, gene_synonyms, alt_names, chromosome, map_location, summary, taxid, species, species_long
Output format: json | txt

Input item: gene_symbol
Input value: Gene symbol
Input value type: string
Input example: ACACA
Output item: all | any, some or all of: mgp_id, gene_id, gene_name, gene_symbol, gene_synonyms, alt_names, chromosome, map_location, summary, taxid, species, species_long
Output format: json | txt

Input item: taxid
Input value: NCBI taxonomy ID
Input value type: integer
Input example: 9606
Output item: all | any, some or all of: mgp_id, gene_id, gene_name, gene_symbol, gene_synonyms, alt_names, chromosome, map_location, summary, taxid, species, species_long
Output format: json | txt

Examples

A representative set of RSET URLs examples for "**gene**" context along with the description of the request are shown below:

Request: Fetch all gene fields from gene symbol

URL: https://www.metabolomicsworkbench.org/rest/gene/gene_symbol/acaca/all

Request: Fetch gene name from Entrez gene id

URL: https://www.metabolomicsworkbench.org/rest/gene/gene_id/31/gene_name

gene context: Table of input name, output name and output field permutations

Match	Input_name	Output_name	Output fields
exact match **partial match for gene_name**	mgp_id|gene_id|gene_name|gene_symbol|taxid	all	mgp_id,gene_id,gene_name,gene_symbol, gene_synonyms,alt_names,chromosome, map_location,summary,taxid, common species name,latin species name
exact match **partial match for gene_name**	mgp_id|gene_id|gene_name|gene_symbol|taxid	*Any, some or all of: mgp_id,gene_id,gene_name,gene_symbol, gene_synonyms,alt_names,chromosome, map_location,summary,taxid, species,species_long	Any, some or all of: mgp_id,gene_id,gene_name,gene_symbol, gene_synonyms,alt_names,chromosome, map_location,summary,taxid, common species name,latin species name

*multiple output names must be separated by commas, e.g. gene_id,gene_name,gene_symbol

The “protein” context

The “**protein**” context refers to a Human Metabolome Gene/Protein Database (MGP) of metabolome-related genes and proteins contains data for over 7300 genes and over 15,500 proteins. In addition to gene information, it provides access to protein related information such as MGP ID, various protein IDs, protein name, protein sequence, etc.

<context> = protein

<input item> = mgp_id | gene_id | gene_name | gene_symbol | taxid | mrna_id | refseq_id | protein_gi | uniprot_id | protein_entry | protein_name

<input value> = <input item value>

<output item> = all | mgp_id | gene_id | gene_name | gene_symbol | taxid | species | species_long | mrna_id | refseq_id | protein_gi | uniprot_id | protein_entry | protein_name | seqlength | seq | is_identical_to | mgp_id,gene_id,gene_name,...

<output format> = txt | json

The “**all**” output item is automatically expanded to include the following items: mgp_id, gene_id, gene_name, gene_symbol, taxid, species, species_long, mrna_id, refseq_id, protein_gi, uniprot_id, protein_entry, protein_name, seqlength, seq, is_identical_to. For example:

https://www.metabolomicsworkbench.org/rest/protein/protein_entry/ACACA_HUMAN/all

Description of input and output items

A brief description of input and output items in “**protein**” context, along with representative examples, is provided in the following table:

Input item: mgp_id
Input value: Human Metabolome Gene/Protein (MGP) database protein ID
Input value type: MGP<6-digit integer**>**
Input example: MGP000016
Output item: all | any, some or all of: mgp_id, gene_id, gene_name, gene_symbol, taxid, species, species_long, mrna_id, refseq_id, protein_gi, uniprot_id, protein_entry, protein_name , seqlength, seq, is_identical_to
Output format: json | txt

Input item: gene_id
Input value: Entrez gene ID
Input value type: integer
Input example: 31
Output item: all | any, some or all of: mgp_id, gene_id, gene_name, gene_symbol, taxid, species, species_long, mrna_id, refseq_id, protein_gi, uniprot_id, protein_entry, protein_name , seqlength, seq, is_identical_to
Output format: json | txt

Input item: gene_name

Input value: Gene name
Input value type: string
Input example: acetyl-CoA carboxylase
Output item: all | any, some or all of: mgp_id, gene_id, gene_name, gene_symbol, taxid, species, species_long, mrna_id, refseq_id, protein_gi, uniprot_id, protein_entry, protein_name , seqlength, seq, is_identical_to
Output format: json | txt

Input item: gene_symbol
Input value: Gene symbol
Input value type: string
Input example: ACACA
Output item: all | any, some or all of: mgp_id, gene_id, gene_name, gene_symbol, taxid, species, species_long, mrna_id, refseq_id, protein_gi, uniprot_id, protein_entry, protein_name , seqlength, seq, is_identical_to
Output format: json | txt

Input item: taxid
Input value: NCBI taxonomy ID
Input value type: integer
Input example: 9606
Output item: all | any, some or all of: mgp_id, gene_id, gene_name, gene_symbol, taxid, species, species_long, mrna_id, refseq_id, protein_gi, uniprot_id, protein_entry, protein_name , seqlength, seq, is_identical_to
Output format: json | txt

Input item: mrna_id
Input value: mRNA ID
Input value type: NM_<integer>
Input example: NM_198834
Output item: all | any, some or all of: mgp_id, gene_id, gene_name, gene_symbol, taxid, species, species_long, mrna_id, refseq_id, protein_gi, uniprot_id, protein_entry, protein_name , seqlength, seq, is_identical_to
Output format: json | txt

Input item: refseq_id
Input value: NCBI reference sequence ID
Input value type: NP_<integer**>**
Input example: NP_942131
Output item: all | any, some or all of: mgp_id, gene_id, gene_name, gene_symbol, taxid, species, species_long, mrna_id, refseq_id, protein_gi, uniprot_id, protein_entry, protein_name , seqlength, seq, is_identical_to
Output format: json | txt

Input item: protein_gi
Input value: NCBI protein GI
Input value type: integer
Input example: 38679977
Output item: all | any, some or all of: mgp_id, gene_id, gene_name, gene_symbol, taxid, species, species_long, mrna_id, refseq_id, protein_gi, uniprot_id, protein_entry, protein_name , seqlength, seq, is_identical_to
Output format: json | txt

Input item: uniprot_id
Input value: UniProt ID
Input value type: string
Input example: Q13085
Output item: all | any, some or all of: mgp_id, gene_id, gene_name, gene_symbol, taxid, species, species_long, mrna_id, refseq_id, protein_gi, uniprot_id, protein_entry, protein_name , seqlength, seq, is_identical_to

Output format: json | txt

Input item: protein_entry
Input value: Protein entry symbol
Input value type: string
Input example: ACACA_HUMAN
Output item: all | any, some or all of: mgp_id, gene_id, gene_name, gene_symbol, taxid, species, species_long, mrna_id, refseq_id, protein_gi, uniprot_id, protein_entry, protein_name , seqlength, seq, is_identical_to
Output format: json | txt

Input item: protein_name
Input value: Protein name
Input value type: string
Input example: acetyl-CoA carboxylase
Output item: all | any, some or all of: mgp_id, gene_id, gene_name, gene_symbol, taxid, species, species_long, mrna_id, refseq_id, protein_gi, uniprot_id, protein_entry, protein_name , seqlength, seq, is_identical_to
Output format: json | txt

Examples

A representative set of RSET URLs examples for "**protein**" context along with the description of the request are shown below:

Request: Fetch all protein fields from UniProt id

URL: https://www.metabolomicsworkbench.org/rest/protein/uniprot_id/Q13085/all

Request: Fetch all protein fields from Entrez gene id

URL: https://www.metabolomicsworkbench.org/rest/protein/gene_id/19/all/

Request: Fetch mRNA id from protein Refseq id

URL: https://www.metabolomicsworkbench.org/rest/protein/refseq_id/NP_005493/mrna_id/

Request: Fetch all protein fields from protein symbol

URL: https://www.metabolomicsworkbench.org/rest/protein/protein_entry/ACACA_HUMAN/all/

Request: Fetch all protein fields from protein name

URL: https://www.metabolomicsworkbench.org/rest/protein/protein_name/lipase/all/

protein context: Table of input name, output name and output field permutations

Match	Input_name	Output_name	Output fields
exact match **partial match for gene_name, protein_name**	mgp_id|gene_id|gene_name|gene_symbol|taxid|mrna_id|refseq_id|protein_gi|uniprot_id|protein_entry|protein_name	all	mgp_id,gene_id,gene_name, gene_symbol, taxid,species,species_long,mrna_id, refseq_id,protein_gi,uniprot_id, protein_entry,protein_name, seqlength,seq,is_identical_to
exact match **partial match for gene_name, protein_name**	mgp_id|gene_id|gene_name|gene_symbol|taxid|mrna_id|refseq_id|protein_gi|uniprot_id|protein_entry|protein_name	*Any, some or all of: mgp_id,gene_id,gene_name,gene_symbol, taxid,species,species_long,mrna_id, refseq_id,protein_gi,uniprot_id, protein_entry,protein_name, seqlength,seq,is_identical_to	Any, some or all of: mgp_id,gene_id,gene_name, gene_symbol, taxid,species,species_long,mrna_id, refseq_id,protein_gi,uniprot_id, protein_entry,protein_name, seqlength,seq,is_identical_to

*multiple output names must be separated by commas, e.g. gene_id,gene_name,gene_symbol

The “moverz” context

The "**moverz**" context refers to performing a m/z search against the **LIPIDS** (a database of ~30,000 computationally generated “bulk” lipid species), **MB** (the Metabolomics Workbench database of ~64,000 exact structures), or **REFMET** (a database of ~30,000 standardized names which includes both exact structures and bulk species detected by MS or NMR) databases by specifying an appropriate m/z value, ion type(adduct) and mass tolerance range.

<context> = moverz

<input item> = LIPIDS | MB | REFMET

<input value1> = <m/z value>

<input value2> = <ion type value>

<input value3> = <m/z tolerance value>

<output format> = txt

The following ion types (adducts) are currently supported: M+H, M+H-H2O, M+2H, M+3H, M+4H, M+K, M+2K, M+Na, M+2Na, M+Li, M+2Li, M+NH4, M+H+CH3CN, M+Na+CH3CN, M.NaFormate+H, M.NH4Formate+H, M.CH3, M.TMSi, M.tBuDMSi, M-H, M-H-H2O, M+Na-2H, M+K-2H, M-2H, M-3H, M-4H, M.Cl, M.F, M.HF2, M.OAc, M.Formate, M.NaFormate-H, M.NH4Formate-H, Neutral.

Description of input and output items

A brief description of input and output items in “**moverz**” context, along with representative examples, is provided in the following table:

Input item: Database
Input item type: string, one of: LIPIDS, REFMET, MB
Input item example: LIPIDS
Input value1: m/z
Input value1 type: float, range: 50-2000
Input example1: 635.52
Input value2: ion type
Input value2 type: string, member of list of allowed ion adducts
Input example2: M+H
Input value3: m/z tolerance
Input value3 type: float, range: 0.0001-1
Input example3: 0.5
Output format: txt
Output: Input m/z, Matched m/z, Delta, Name, Systematic name, Formula, Ion, Category, Main class, Sub class

Examples

A representative set of REST examples for the “**moverz**” context along with the description of the request are shown below:

Request: Perform MS precursor ion search on Metabolomics Workbench database with m/z 635.52, ion-type M+H and mass tolerance of 0.5 and output as text

URL: https://www.metabolomicsworkbench.org/rest/moverz/MB/635.52/M+H/0.5/txt

Request: Perform MS precursor ion search on LIPIDS virtual database with m/z 513.45, ion-type M-2H (2-) and mass tolerance of 0.2 and output as text

URL: https://www.metabolomicsworkbench.org/rest/moverz/LIPIDS/513.45/M-2H/0.2/txt

Request: Perform MS precursor ion search on REFMET database with m/z 255.2, ion-type M+H and mass tolerance of 0.2 and output as text

URL: https://www.metabolomicsworkbench.org/rest/moverz/REFMET/255.2/M+H/0.2/txt

The “exactmass” context

The context "**exactmass**" calculates the exact mass of a lipid species by specifying an appropriate lipid abbreviation and ion type(adduct).

<context> = exactmass

<input value1> = <LIPID abbreviation>

<input value2> = <ion type value>

The following head groups are currently supported as abbreviations for lipids: ArthroCer, asialo-GM2Cer, CAR, CE, Cer, CerP, CoA, DG, DGDG, FA, GalCer, GB3Cer, GlcCer, GM3Cer, GM4Cer, iGB3Cer, LacCer, Lc3Cer, Manb1-4GlcCer, MG, MGDG, MolluCer, PA, PC, PE, PE-Cer, PG, PGP, PI, PI-Cer, PIP, PIP2, PIP3, PS, SM, SQDG, TG

The following ion types (adducts) are currently supported: Neutral, M+H, M+H-H2O, M+2H, M+3H, M+4H, M+K, M+2K, M+2K-H, M+Na, M+2Na, M+2Na-H, M+Li, M+2Li, M+Ag, M+NH4, M-H, M-CH3, M-2H, M-3H, M-4H, M.Cl, M.OAc, M.Formate

Description of input and output items

A brief description of input and output items in the “**exactmass**” context, along with representative examples, is provided in the following table:

Input value1: LIPID MAPS lipid abbreviation
Input value1 type: string, member of list of allowed head groups as abbreviations
Input example1: PC(34:1)
Input value2: ion type
Input value2 type: string, member of list of allowed ion adducts
Input example2: M+H
Output: Input abbrev, Input ion type, Exact mass, Molecular formula

Examples

A representative set of examples for the “**exactmass**” context along with the description of the request are shown below:

Request: Calculate the exact mass (m/z) of the [M+H]+ ion for the lipid abbreviation PC(34:1)

URL: https://www.metabolomicsworkbench.org/rest/exactmass/PC(34:1)/M+H

Request: Calculate the exact mass (m/z) of the [M-H]- ion for the lipid abbreviation GlcCer (d42:2)

URL: https://www.metabolomicsworkbench.org/rest/exactmass/GlcCer(d42:2)/M-H