PyBioGateway functions

Ontology	Uri	Prefix
Simple Knowledge Organization System (SKOS)	http://www.w3.org/2004/02/skos/core#	skos
Resource Description Framework (RDF)	http://www.w3.org/1999/02/22-rdf-syntax-ns#	rdf
RDF Schema (RDFS)	http://www.w3.org/2000/01/rdf-schema#	rdfs
Semanticscience Integrated Ontology (SIO)	http://semanticscience.org/resource/	sio
Open Biomedical Ontologies (OBO)	http://purl.obolibrary.org/obo/	obo
Dublin core (DC)	http://purl.org/dc/terms/	dc
Schema (SCH)	http://schema.org/	sch
Open Biomedical Ontologies (OBO)	http://www.geneontology.org/formats/oboInOwl#	oboowl

Table 1. Table with information about the ontology and the prefixes used in the queries.

Function type_data(name):

Description: This function is used to query the type of biological entity to which a specific entity entered by the user belongs.

This function has as parameter the name (prefLabel or altLabel) of the specific biological entity that we want to query in the knowledge graph, and returns the type of biological entity

to which it belongs, i.e. whether it is a gene, protein, topologically associated domain (TAD), cis-regulatory module (CRM), biological process, cellular component or molecular function.

Parameters:

-**"name"**: Name of the biological entity that the user wants to consult. Seven types of entities can be queried: proteins, genes, TADs, CRMs, biological processes, molecular functions or cellular components. The following vocabularies are used:

- Proteins: Both the Uniprot entry name and alternative identifiers are valid. Example: 'INSR_HUMAN'.
- Genes: gene symbol HGNC. Example: 'INSR'.
- TADs and CRMs: BioGateway's own identifiers. Example: 'crm/CRMHS00000003515', 'tad/TADHS00000020654'.
- Biological processes, molecular functions or cellular components: Gene Ontology (GO) proprietary identifiers. Example: 'GO:0000206'.

Output:

-**"bioentity_type"**: Type of biological entity to which the input value belongs. Possible values:
- Protein: Example output, 'protein' (prefLabel).
- Gene: Example output, 'gene' (prefLabel).
- TAD: Example output, 'topologically_associated_domain (tad)' (prefLabel).
- CRM: Example output, 'cis_regulatory_module (crm)' (prefLabel).
- Biological process: Example output, 'biological_process' (hasOBONamespace).
- Cellular component: Example output, 'cellular_component' (hasOBONamespace).
- Molecular function: Example output, 'molecular_function' (hasOBONamespace).

Implementation example:

Input: **type_data**('INSR_HUMAN').

Output:

'protein'

Input with alternative name of the same protein: **type_data**('P06213')

Output:

'protein'

Summary table:

Variable	Rol	Type	Description	Ontology property
"name"	**input**	**string**	**Biological entity name**	**skos:prefLabel o skos:altLabel**
"bioentity_type"	**output**	**string**	**Biological entity type**	**rdfs:subClassof obo:hasOBONamespace**

Function: getGene_info(gene, taxon)

Description: Function that provides information associated with the gene to be queried, available in the knowledge network: **http://rdf.biogateway.eu/graph/gene.**
This function has two parameters: the name of the gene we want to query and the taxon.

Parameters:

-**"gene":** We insert the gene in gene symbol format (property: prefLabel). Example: 'BRCA1'.

-**"taxon":** We insert either the taxon number or the name of the organism (property: label). Example: 'Homo sapiens', '9606'.

Output:

Returns a dictionary with following fields:

-**"chromosome":** Indicates the chromosome to which the gene belongs (property: BFO_0000050 (part of)).

-**"start":** Indicates the starting position of the gene in the genomic sequence (property: GENO_0000894 (start_position)).

-**"end":** Indicates the end position of the gene in the genomic sequence (property: GENO_0000895 (end_position)).

-**"strand":** Indicates the orientation of the gene in the genomic sequence (property: GENO_0000906 (on strand)).

-**"assembly":** Indicates the human genome assembly corresponding to our gene (property: hasVersion).

-**"alt_gene_sources":** Corresponds to other sources or databases containing information related to the gene of interest (property: closeMatch).

-**"definition":** Provides a definition of the gene (property: definition).

Implementation example:

Input: **getGene_info**("Brca1", "Mus musculus")

Output:
{'start': '101379587',
 'end': '101442808',
 'strand': 'ReverseStrandPosition',
 'chr': 'NC_000077.7',

'assembly': 'GCF_000001635.27',

'alt_gene_sources':'ensembl/ENSMUSG00000017146;

ensembl/ENSMUSG00000017146.13; ncbigene/12189',

 'definition': 'gene 10090/Brca1 encoding [A0A087WP26_MOUSE A0A087WPE1_MOUSE

A0A087WPK5_MOUSE BRCA1_MOUSE]'}

Summary table:

Variable	Rol	Type	Description	Ontology property	Graph
"gene"	**input**	**string**	**Gene name in symbol format**	**skos:prefLabel**	**gene**
"taxon"	**input**	**string**	**Taxon number or organism name**	**rdfs:label**	**taxon**
"chromosome"	**output**	**string**	**Chromosome to which gene belongs**	**obo:BFO_0000050 (part of)**	**gene**
"start"	**output**	**integer**	**Gene start position**	**obo:GENO_0000894 (start position)**	**gene**
"end"	**output**	**integer**	**Gene end position**	**obo:GENO_0000895 (end position)**	**gene**
"strand"	**output**	**string**	**Gene orientation**	**obo:GENO_0000906 (on strand)**	**gene**
"assembly"	**output**	**string**	**Assembly of human genome to which gene corresponds**	**dc:hasVersion**	**gene**

"alt_gene_sources"	output	string	Alternative sources of gene information	skos:closeMatch	gene
"definition"	output	string	Gene description	skos:definition	gene

Function getGenes_by_coord(chr, start, end , strand):

Description: Function which returns the genes in gene symbol (prefLabel) format, which are located within the specified genomic coordinates. We use the information available in the **http://rdf.biogateway.eu/graph/gene.**

Parameters:

-**"chr"**: Indicates the chromosome to which the segment belongs (NCBI chromosome identifier). Example: 'NC_000074.7'.

-**"start"**: Indicates the starting position of the segment in the genomic sequence. Example: '90973665'

-**"end":** Indicates the end position of the segment in the genomic sequence. Example: '91075654'.

-**"strand"**: Indicates the DNA strand to search on. **If declared** as **none**, it will search on both strands. Allowed values: **'ReverseStrandPosition'** and **'ForwardStrandPosition'**.

Output:

Gene list that falls within the specified genomic coordinates. In turn, for each gene, the following is indicated:

-**"gene_name":** Gene name in gene symbol format (property: prefLabel).

-**"start"**: Indicates the gene's start position in the genomic sequence (property: GENO_0000894).

-**"end"**: Indicates the gene's end position in the genomic sequence (property: GENO_0000895).

-**"strand":** Indicates gene orientation in the genomic sequence (property: GENO_0000906). This field is only provided if no DNA strand has been specified in the function.

Implementation example:

Input: **getGenes_by_coord**("NC_051352.1", 52565276, 58596412 ,"ForwardStrandPosition")
Output:
[{'gene_name': 'Fzd8', 'start': '57312924', 'end': '57320551'},
 {'gene_name': 'Hist2h3c2', 'start': '183797721', 'end': '41378877'},
 {'gene_name': 'Hist2h3c2', 'start': '183837311', 'end': '41532577'},
 {'gene_name': 'Hist2h3c2', 'start': '183797721', 'end': '41532577'},
 {'gene_name': 'Hist2h3c2', 'start': '183837311', 'end': '41426735'},
 {'gene_name': 'Hist2h3c2', 'start': '183837311', 'end': '41378877'},
 {'gene_name': 'Hist2h3c2', 'start': '183797721', 'end': '41426735'},
 {'gene_name': 'Hnrnpk', 'start': '91756628', 'end': '6275001'},
 {'gene_name': 'Lgals8', 'start': '58024652', 'end': '58052764'},
 {'gene_name': 'Map3k8', 'start': '53382908', 'end': '53403216'},
 {'gene_name': 'Mtr', 'start': '58219998', 'end': '58308560'},
 {'gene_name': 'Actn2', 'start': '58143334', 'end': '58210622'},
 {'gene_name': 'Crem', 'start': '54238889', 'end': '54305989'}]

Summary table:

Variable	Rol	Type	Description	Ontology property	Graph
"chr"	**input**	**string**	**Chromosome to which genomic segment belongs**	**obo:BFO_0000050 (part of)**	**gene**
"start"	**input**	**integer**	**Genomic segment start position**	**obo:GENO_0000894 (start position)**	**gene**
"end"	**input**	**integer**	**Genomic segment end position**	**obo:GENO_0000895 (end position)**	**gene**
"strand"	**input (optional)**	**string**	**DNA chain where you want to perform the search**	**obo:GENO_0000906 (on strand)**	**gene**
"gene_name"	**output**	**string**	**Gene name in symbol format**	**skos:prefLabel**	**gene**
"start"	**output**	**integer**	**Gene start position**	**obo:GENO_0000894 (start position)**	**gene**
"end"	**output**	**integer**	**Gene end position**	**obo:GENO_0000895 (end position)**	**gene**
"strand"	**output (optional)**	**string**	**Gene orientation**	**obo:GENO_0000906 (on strand)**	**gene**

Function getProtein_info(protein)

Description: Function that provides the information associated with the interest protein, available in the knowledge graph: **http://rdf.biogateway.eu/graph/protein.**
This function has one parameter, the name of the protein we want to consult in the Uniprot entry name format.

Parameters:

-**"protein":** We insert the protein name in the Uniprot entry name format (property: prefLabel). Example: 'BRCA1_HUMAN'.

Output:

Returns a dictionary with the following fields:

-**"protein_alt_names":** Returns the alternative protein names (property: altLabel), in particular the name of the protein in Uniprot entry format, as well as the name of the gene encoding it and synonyms for this gene.

-**"definition":** Provides a protein definition (property: definition).

-**"evidence_level":** This term refers to the evidence level that supports the information associated with the protein (property: evidenceLevel).

-**"alt_sources":** Refers to other sources or databases containing information related to the protein (property: closeMatch).

-**"articles":** Corresponds to scientific articles or publications that are associated with the protein of interest found in the Pubmed database (property: SIO_000772 (has evidence)).

Implementation example:

Input: **getProtein_info**("TOX3_HUMAN")

Output:
{'protein_alt_ids': 'O15405; TOX3; TNRC9; CAGF9', 'definition': 'TOX high mobility group box family member 3 (CAG trinucleotide repeat-containing gene F9 protein) (Trinucleotide repeat-containing gene 9 protein)', 'evidence_level': '5.0', 'alt_sources': 'ensembl/ENSP00000219746.9; ensembl/ENSP00000385705.3; refseq/NP_001073899.2; refseq/NP_001139660.1', 'articles': 'pubmed/15616553; pubmed/9225980; pubmed/21172805; pubmed/14702039'}

Summary table:

Variable	Rol	Type	Description	Ontology property	Graph
"protein"	**input**	**string**	**Protein name in Uniprot entry name format**	**skos:prefLabel**	**protein**
"protein_alt_names"	**output**	**string**	**Alternative protein names**	**skos:altLabel**	**protein**
"definition"	**output**	**string**	**Protein definition**	**skos:definition**	**protein**
"evidence_level"	**output**	**string**	**Evidence level associated with the protein**	**sch:evidenceLevel**	**protein**
"alt_sources"	**output**	**string**	**Alternative sources of protein information**	**skos:closeMatchl**	**protein**
"articles"	**output**	**string**	**Protein-related articles**	**sio:SIO_000772 (has evidence)**	**protein**

Function getPhenotype(phenotype):

Description: This function allows to obtain the phenotype from an OMIM identifier or from the name of a disease, using the graph **http://rdf.biogateway.eu/graph/omim**:

-On the one hand, if the value entered is an OMIM identifier of a phenotype, it will return its preferred label.

-On the other hand, if the value entered is the name of a disease (e.g. 'breast cancer'), it will return the preferred label and the omim identifier of those phenotypes whose label, both preferred and alternative, contains the disease entered. This means, phenotypes that contain the entered name.

Parameters:

-**"phenotype"**: Allowed values of this parameter will be:

-OMIM identifier of a phenotype. Example 'MTHU036782'.

-Name of a disease. Example 'lung cancer'.

Output:

The output will depend on the type of parameter entered:

-If an OMIM identifier of a phenotype has been entered, the output will simply be the preferred label of this identifier (prefLabel).

-If the name of a disease has been entered, the output will be composed of those phenotypes that contain this disease in their preferred label (prefLabel or altLabel), as well as their OMIM identifier.

Implementation example:

Input: **getPhenotype**("breast cancer")

Output: [{'omim_id': '604704', 'label': 'BREAST CANCER ANTIESTROGEN RESISTANCE 3'},
 {'omim_id': 'MTHU036782', 'label': 'Breast cancer, lobular'},
 {'omim_id': 'MTHU068657', 'label': 'Breast cancer, early-onset'},
 {'omim_id': '137215',
  'label': 'DIFFUSE GASTRIC AND LOBULAR BREAST CANCER SYNDROME'},
 {'omim_id': 'MTHU015471',
  'label': 'Paraneoplastic SPS is associated with breast cancer and other malignancies'},

Input: **getPhenotype**("604704")

Output: [{'phen_label': 'BREAST CANCER ANTIESTROGEN RESISTANCE 3'}]

Summary table:

Variable	Rol	Type	Description	Ontology property	Graph
"phenotype"	**input**	**string**	**Disease name or phenotype OMIM identifier**	**skos:prefLabel**	**omim**
"omim_id"	**output**	**string**	**Phenotype OMIM identifier (optional)**	**skos:prefLabel**	**omim**
"phen_label"	**output**	**string**	**Phenotype common name**	**skos:prefLabel**	**omim**

Function getCRM_info(crm):

Description: Function that returns the information associated with the cis-regulator module (crm) that we want to consult, available in the knowledge network: **http://rdf.biogateway.eu/graph/crm.**
This function has a parameter which is the name of the crm we want to query (property: prefLabel).

Parameters:

-**"crm"**: The parameter entered in the function shall be the preferred name of the cis-regulator module (property: prefLabel). Example"crm/CRMHS00003225754".

Output:
Returns a dictionary with the following fields:

-**"start"**: Indicates the starting position of the crm in the genomic sequence (property: GENO_0000894 (start_position)).

-**"end"**: Indicates the end position of the crm in the genomic sequence (property: GENO_0000895 (end_position)).

-**"chromosome"**:Indicates chromosome to which crm belongs (NCBI chromosome identifier) (property: BFO_0000050 (part_of)).

-**"assembly":** Indicates the assembly of the human genome corresponding to our crm (property: hasVersion).

-**"taxon"**: Returns the taxon to which this crm belongs (NCBI taxonomic identifier) (property: RO_0002162).

-**"definition"**: Definition of crm available in the property definition of the knowledge graph.

Implementation example:

Input: **getCRM_info**("crm/CRMHS00000005387")

Output:
[{'start': '355447',
  'end': '358949',
  'chromosome': 'NC_000011.10',
  'assembly': 'GCF_000001405.26',
  'taxon': 'NCBITaxon_9606',
   'definition': 'Cis-regulatory module located in Homo sapiens chr11 between 355447 and 358949'}]

Summary table:

Variable	Rol	Type	Description	Ontology property	Graph
"crm"	**input**	**string**	**Cis-regulated module name (crm)**	**skos:prefLabel**	**crm**
"taxon"	**output**	**string**	**Taxon number to which crm belongs**	**obo:RO_0002162 (in taxon)**	**crm**
"chromosome"	**output**	**string**	**Chromosome to which the crm belongs**	**obo:BFO_0000050 (part of)**	**crm**
"start"	**output**	**integer**	**Crm start position**	**obo:GENO_0000894 (start position)**	**crm**
"end"	**output**	**integer**	**Crm end position**	**obo:GENO_0000895 (end position)**	**crm**
"strand"	**output**	**string**	**Crm orientation**	**obo:GENO_0000906 (on strand)**	**crm**
"assembly"	**output**	**string**	**Assembly of**	**dc:has Version**	**crm**

			the human genome that corresponds to the crm		
"definition"	output	string	Crm description	skos:definition	crm

Function getCRM_additional_info(crm):

Description: This function will provide additional information about the crm of interest, available in the network **http://rdf.biogateway.eu/graph/crm**, which is not obtained from the function **getCRM_info**.

Parameters:

-**"crm"**: The entered parameter in the function shall be the preferred name of the cis-regulator module (property: prefLabel). Example: "crm/CRMHS00000005387".

Output:

Returns a dictionary with the following fields:

-**"evidence":** Corresponds to the evidence supporting the information available on the regulatory module-cis (property: evidenceOrigin).

-**"database":** Indicates the database where the information on the crm of interest is registered (property: SIO_000253 (has source)).

-**"biological_samples":** Refers to the different types of biological samples that are associated with the study of the cis-regulator module (property: TXPO_0003500 (observed in)). Specifically, it will return the identifiers in ontology term format. Example: "CLO_0001601", "UBERON_0002113", "BTO_0000018".

-**"articles":** Refers to scientific articles that have been published in Pubmed related to the cis-regulatory module introduced (property: SIO_000772 (has evidence)).

Implementation example:

Input: **getCRM_add_info**("crm/CRMHS00000005387")
Output:
{'evidence':
'http://www.licpathway.net/ENdb/search/Detail.php?Species=Human&Enhancer_id=E_01_273',
 'database': 'http://www.licpathway.net/ENdb/',
   'biological_samples': 'BTO_0000007; BTO_0000018; CLO_0001230; CLO_0001601; CL_0002518; CL_0000082; UBERON_0002048; UBERON_0002113',
 'articles': 'pubmed/28511927'}

Summary table:

Variable	Rol	Type	Description	Ontology property	Graph
"crm"	**input**	**string**	**Cis-regulator module name (crm)**	**skos:prefLabel**	**crm**
"evidence"	**output**	**string**	**Evidence level associated with the crm**	**sch:evidenceOrigin**	**crm**
"database"	**output**	**string**	**Database where we find crm's information**	**sio:SIO_000253 (has source)**	**crm**
"biological_samples"	**output**	**string**	**Biological samples**	**obo:TXPO_0003500 (observed in)**	**crm**

			associated with the crm of interest		
"articles"	output	string	Crm related articles	sio:SIO_000772 (has evidence)	crm

Function getCRMs_by_coord(chromosome, start, end):

Description: Function that returns the identifier of the crms (prefLabel) that lie within the specified genomic coordinates.

Parameters:

-**"chromosome"**: Indicates the chromosome on which we want to perform the search, the allowed values are chr-chromosome number. Example: **'chr-1'. 'chr-18', "mitochondrial"** (to search in mitochondrial DNA).

-**"start"**: Indicates the starting position of the segment in the genomic sequence. Example: '90973665'

-**"end":** Indicates the end position of the segment in the genomic sequence. Example: '91075654'.

Output:

List of regulatory cis-modules that fall within the specified genomic coordinates. At the same time, for each crm it is indicated:

-**"crm_name":** Cis regulatory module name that lies within the specified genomic co-ordinates (property: prefLabel).

-**"start"**: Indicates the starting position of the crm in the genomic sequence (property: GENO_0000894).

-**"end"**: Indicates the end position of the crm in the genomic sequence (property: GENO_0000895).

Implementation example:

Input: **getCRMs_by_coord**("mitochondrial","1", "500")
Output:
[{'crm_name': 'crm/CRMHS00032244267', 'start': '175', 'end': '426'},
 {'crm_name': 'crm/CRMHS00032244371', 'start': '230', 'end': '389'},
 {'crm_name': 'crm/CRMHS00032244378', 'start': '32', 'end': '334'}]

Summary table:

Variable	Rol	Type	Description	Ontology property	Graph
"chromosome"	**input**	**string**	**Chromosome to which genomic segment belongs**	**obo:BFO_0000050 (part of)**	**crm**
"start"	**input**	**integer**	**Genomic segment start position**	**obo:GENO_0000894 (start position)**	**crm**
"end"	**input**	**integer**	**Genomic segment end position**	**obo:GENO_0000895 (end position)**	**crm**
"crm_name"	**output**	**string**	**Crm name that lies within the genomic**	**skos:prefLabel**	**crm**

			coordinates		
"start"	output	integer	Crm start position	obo:GENO_0000894 (start position)	crm
"end"	output	integer	Crm end position	obo:GENO_0000895 (end position)	crm

Function getTAD_info(tad):

Description: Function that provides information associated to the topologically associated domain (tad) that we want to consult, available in the knowledge graph: **http://rdf.biogateway.eu/graph/tad**.

This function has a parameter which is the identifier of the tad that we want to query (property: prefLabel).

Parameters:

-**"tad"**: The parameter entered in the function will be the preferred name of the topologically associated domain (property: prefLabel). Example: "tad/TADHS00000038004"

Output:

Returns a dictionary with the following fields:

-**"chromosome"**: Indicates the chromosome to which tad belongs, (NCBI chromosome identifier) (property: BFO_0000050).

-**"start"**: Indicates tad start position in the genomic sequence (property: GENO_0000894).

-**"end"**: Indicates tad end position in the genomic sequence (property: GENO_0000895).

-**"assembly":** Indicates the assembly of the human genome corresponding to our tad (property: hasVersion).

-**"taxon"**: Returns the taxon to which this tad belongs (NCBI taxonomic identifier) (property: RO_0002162).

-**"definition"**: Tad definition available in the property definition of the knowledge graph.

Implementation example:

Input: **getTAD_info**("tad/TADHS00000038004")
Output:
[{'start': '34120000',
 'end': '35840000',
 'chromosome': 'NC_000013.11',
 'assembly': 'GCF_000001405.26',
 'taxon': 'NCBITaxon_9606',
  'definition': 'Topologically associated domain located in Homo sapiens chr13 between 34120000 and 35840000'}]

Summary table:

Variable	Rol	Type	Description	Ontology property	Graph
"tad"	**input**	**string**	**Topologically associated domain name**	**objeto de la relación prefLabel**	**tad**
"chromosome"	**output**	**string**	**Chromosome to which tad belongs**	**obo:BFO_0000050 (part of)**	**tad**
"start"	**output**	**integer**	**Tad start position**	**obo:GENO_0000894 (start position)**	**tad**
"end"	**output**	**integer**	**Tad end position**	**obo:GENO_0000895 (end position)**	**tad**

"assembly"	output	string	Assembly of the human genome to which the tad corresponds	dc:hasVersion	tad
"taxon"	output	string	Taxón al que pertenece este tad (identificador del NCBI)	obo:RO_0002162 (in taxon)	tad
"definition"	output	string	Tad description	skos:definition	tad

Function getTAD_additional_info(tad):

Description: This function will provide additional information about the tad of interest, available in the network **http://rdf.biogateway.eu/graph/tad**, which is not obtained from the **getTAD_info** function.

Parameters:

-**"tad"**: Parameter entered in the function will be the topologically associated domain's preferred name (property: prefLabel). Example: "tad/TADHS00000038004".

Output:

Returns a dictionary with the following fields:

-**"evidence":** Corresponds to the evidence that supports the available information on the tad (property: evidenceOrigin).

-**"database":** Indicates database where the tad information of interest is registered (property: SIO_000253 (has source)).

-**"biological_samples":** It refers to the different types of biological samples that are associated with the study of the topologically associated domain (property: TXPO_0003500 (observed in)). In particular, it shall return the identifiers in ontological term format. Example: 'CLO_0001601', 'UBERON_0002113', 'BTO_0000018'.

-**"articles":** Refers to the scientific articles that have been published in Pubmed related to the topologically associated domain introduced (property: SIO_000772 (has evidence)).

Implementation example:

Input: **getTAD_add_info**("tad/TADHS00000038004")
Output:
{'evidence':
'http://dna.cs.miami.edu/TADKB/domain.php?sp=hum&cl=HMEC&rg=hg19&chr=14&se=96700000_96850000&id=83&res=50kb&caller=IS',
 'database': 'http://dna.cs.miami.edu/TADKB/',
  'biological_samples': 'BTO_0001229; CLO_0006951; CL_0002327; UBERON_0000310; UBERON_0002048; BTO_0004300; CL_0002553',
 'articles': 'pubmed/30871473'}

Summary table:

Variable	Rol	Type	Description	Ontology property	Graph
"tad"	**input**	**string**	**Topologically associated domain name (tad)**	**skos:prefLabel**	**tad**
"evidence"	**output**	**string**	**Evidence level associated with tad**	**sch:evidenceOrigin**	**tad**
"database"	**output**	**string**	**Database where we find tad's**	**sio:SIO_000253 (has source)**	**tad**

			information		
"biological_samples"	output	string	Biological samples associated with tad of interest	obo:TXPO_0003500 (observed in)	tad
"articles"	output	string	Tad related articles	sio:SIO_000772 (has evidence)	tad

Function getTAD_by_coord(chromosome,start, end):

Description: Function that returns tads identifiers (prefLabel) that lie within the specified genomic coordinates.

Parameters:

-**"chromosome"**: Indicates the chromosome to which segment belongs (NCBI chromosome identifier). Example: "chr-13".

-**"start"**:Indicates starting segment position in the genomic sequence. Example: "90973665"

-**"end":** Indicates ending segment position in the genomic sequence. Example: "91075654"

Output:

-**"tad_name":** Returns a list of the topologically associated domain ids that lie within the specified genomic coordinates (property: prefLabel).

-**"start"**: Indicates tad start position in the genomic sequence (property: GENO_0000894).

-**"end"**: Indicates tad end position in the genomic sequence (property: GENO_0000895).

Implementation example:

Input: **getTADs_by_coord**("chr-13","34120000", "35840000")
Output:
[{'tad_id': 'tad/TADHS00000038004', 'start': '34120000', 'end': '35840000'},
 {'tad_id': 'tad/TADHS00000029314', 'start': '35200000', 'end': '35840000'},
 {'tad_id': 'tad/TADHS00000071459', 'start': '34125863', 'end': '35175863'},
 {'tad_id': 'tad/TADHS00000071460', 'start': '34165863', 'end': '35825863'},
 {'tad_id': 'tad/TADHS00000071461', 'start': '34185863', 'end': '35155863'},
 {'tad_id': 'tad/TADHS00000071462', 'start': '34305863', 'end': '35455863'},
 {'tad_id': 'tad/TADHS00000071463', 'start': '34325863', 'end': '35025863'},
 {'tad_id': 'tad/TADHS00000071465', 'start': '35150863', 'end': '35800863'},
 {'tad_id': 'tad/TADHS00000071468', 'start': '35170863', 'end': '35810863'},
 {'tad_id': 'tad/TADHS00000071479', 'start': '35195863', 'end': '35815863'},
 {'tad_id': 'tad/TADHS00000071484', 'start': '35215863', 'end': '35825863'},
 {'tad_id': 'tad/TADHS00000071485', 'start': '35225863', 'end': '35825863'},
 {'tad_id': 'tad/TADHS00000071489', 'start': '35375863', 'end': '35725863'},
 {'tad_id': 'tad/TADHS00000071490', 'start': '35375863', 'end': '35775863'},
 {'tad_id': 'tad/TADHS00000071491', 'start': '35375863', 'end': '35825863'},
 {'tad_id': 'tad/TADHS00000071492', 'start': '35425863', 'end': '35725863'}]

Summary table:

Variable	Rol	Type	Description	Ontology property	Graph
"chromosome"	**input**	**string**	**Chromosome**	**obo:BFO_0000050**	**tad**

			to which genomic segment belongs	(part of)	
"start"	input	integer	Genomic segment start position	obo:GENO_0000894 (start position)	tad
"end"	input	integer	Genomic segment end position	obo:GENO_0000895 (end position)	tad
"tad_name"	output	string	Tad name that lies within genomic co-ordinates	skos:prefLabel	tad
"start"	output	integer	Tad start position	obo:GENO_0000894 (start position)	tad
"end"	output	integer	Tad end position	obo:GENO_0000895 (end position)	tad

Function gene2protein(gene,taxon)

Description: This function allows to obtain the proteins encoded by the gene entered in the query, and in the selected taxon. If the taxon value is **None**, it will return the proteins encoded by the gene in the different taxa. This information shall be obtained from the graph **http://rdf.biogateway.eu/graph/gene**

Parameters:

-**"gene":** This parameter corresponds to the gene name in symbol format (property: prefLabel). Example: "BRCA1"

-**"taxon":** Allows you to select the taxon on which you want to perform the query. The value can be the NCBI taxonomic identifier or the name of the organism (property: label). Example: "Homo sapiens", "9606".

Output:

Returns a list of the proteins encoded by the gene of interest:

-**"protein_name":** Function returns names in Uniprot entry name format (prefLabel) of the proteins encoded by the entered gene (property: SIO_010078 (encodes)).

Implementation example:

Input: **gene2protein**("TOX3","9606")
Output:
[{'protein_name': 'H3BTZ9_HUMAN'},
 {protein_name: 'J3QQQ6_HUMAN'},
 {protein_name: 'TOX3_HUMAN'}]

Summary table:

Variable	Rol	Type	Description	Ontology property	Graph
"gene"	**input**	**string**	**Gene name in symbol format**	**skos:prefLabel**	**gene**
"taxon"	**input**	**string**	**Taxon number or organism name**	**skos:label**	**taxon**
"protein_name"	**output**	**string**	**Protein name in Uniprot entry name format**	**sio:SIO_010078 (encodes)**	**gene**

			encoded by the entered gene		

Function protein2gene(protein)

Description: This function allows to obtain the gene encoding the protein entered in the query. This information will be obtained from the graph **http://rdf.biogateway.eu/graph/gene**.

Parameters:

-**"protein":** We enter the protein name either in Uniprot's entry name format (property: prefLabel) or in Uniprot's entry format. (property: altLabel). Example: "BRCA1_HUMAN", "P38398".

Output:

-"**gene_name**": The function returns the gene name in symbol format (prefLabel) that encodes the entered protein (property: SIO_010078).

Implementation example:

Input: **protein2gene**("BRCA1_MOUSE")
Output: [{'gene_name': 'Brca1'}]

Summary table:

Variable	Rol	Type	Description	Ontology property	Graph
"protein"	**input**	**string**	**Protein name in Uniprot entry name format encoded by the introduced gene**	**skos:prefLabel**	**protein**
"gene_name"	**output**	**string**	**Gene name in symbol format**	**sio:SIO_010078 (encodes)**	**gene**

Function gene2phen(gene)

Description: This function makes it possible to obtain the phenotypes associated with a gene previously introduced as a parameter of this function. To do so, we will exploit the graph **http://rdf.biogateway.eu/graph/gene2phen.**

Parameters:

-**"gene":** This parameter corresponds to the gene name in symbol format (property: prefLabel). Example: "BRCA1".

Output:

Function returns a list of phenotypes related to the gene of interest. For each phenotype we obtain the following data:

-**"omim_id":** Corresponds to the Omim identifier of the phenotype that is associated to the introduced gene (property: RO_0002331 (involved in)).

-**"phen_label":** Refers to the preferred label of the phenotype to which the identifier omim corresponds (pref_label).

Implementation example:

Input: **gene2phen**("BRCA1")
Output:
[{'omim_id': '114480', 'phen_label': 'Breast cancer (BC)'},
 {'omim_id': '167000', 'phen_label': 'Ovarian cancer (OC)'},
 {'omim_id': '604370',
  'phen_label': 'Breast-ovarian cancer, familial, 1 (BROVCA1)'},
 {'omim_id': '617883',
  'phen_label': 'Fanconi anemia, complementation group S (FANCS)'}]

Summary table:

Variable	Rol	Type	Description	Ontology property	Graph
"gene"	**input**	**string**	**Disease name or OMIM identifier of a phenotype**	**obo:RO_0002331 (involved in)**	**gene2phen**
"omim_id"	**output**	**string**	**Phenotype OMIM identifier (optional)**	**obo:RO_0002331 (involved in)**	**gene2phen**
"phen_label"	**output**	**string**	**Phenotype common name**	**skos:prefLabel**	**gene2phen**

Function phen2gene(phenotype)

Description: This function allows us to obtain the genes associated with a phenotype previously introduced as a parameter of this function. To do so, we will exploit the graph **http://rdf.biogateway.eu/graph/gene2phen.**

Parameters:

-**"phenotype"**: This parameter corresponds to the phenotype of interest. It allows both its OMIM identifier ('211980'), as well as a disease name ("lung cancer").

Output:

-"**gene_name**": Function returns a list with names in symbol format (property: prefLabel) of the genes related to the phenotype of interest (property: RO_0002331).

Implementation example:

Input: **phen2gene**("lung cancer")
Output:
[{'gene_name': 'MXRA5'}, {'gene_name': 'BRAF'}, {'gene_name': 'ERBB2'}, {'gene_name': 'SLC22A18'}]

Summary table:

Variable	Rol	Type	Description	Ontology property	Graph
"phenotype"	**input**	**string**	**Disease name or OMIM identifier of a phenotype**	**obo:RO_0002331 (involved in)**	**gene2phen**
"gene_name"	**output**	**string**	**Gene name in symbol format**	**obo:RO_0002331 (involved in)**	**gene2phen**

Function prot2bp(protein)

Description: This function allows the study of the biological processes in which the protein of interest participates, using the information of the graph **http://rdf.biogateway.eu/graph/prot2bp.**

Parameters:

-**"protein":** We enter the protein name either in Uniprot's entry name format (property: prefLabel) or in Uniprot's entry format. (property: altLabel). Example: "BRCA1_HUMAN", "P38398".

Output:

This function returns a list of biological processes related to the selected protein. It presents the following fields:

-**"bp_id":** Corresponds to the identifier of the biological process in Gene Ontology (property: oboInOwl#id). Example: "GO:0035349"

-**"bp_label":** Refers to the label of the biological process in the knowledge graph network **http://rdf.biogateway.eu/graph/go** (property: label). Example: "coenzyme A transmembrane transport"

-**"relation_label":** Refers to the label presenting the relationship between the protein and the biological process of interest (property: prefLabel) available in the graph **http://rdf.biogateway.eu/graph/prot2bp .**

-**"database":** Indicates the database where the information on the relationship between protein and biological process of interest is recorded (property: SIO_000253 (has source)).

-**"articles":** Corresponds to scientific articles or publications that are associated with the relationship between the protein and the biological process of interest, found in the Pubmed database (property: SIO_000772 (has evidence)).

Implementation example:

Input: **prot2bp**("BRCA1_HUMAN")

Output:

[{'bp_id': 'GO:0035066',
 'bp_label': 'positive regulation of histone acetylation',
 'relation_label': 'P38398--GO:0035066',
 'database': 'goa/',
 'articles': 'pubmed/21873635'},
 {'bp_id': 'GO:0045786',
 'bp_label': 'negative regulation of cell cycle',
 'relation_label': 'P38398--GO:0045786',
 'database': 'goa/',
 'articles': 'pubmed/15159397'},
 {'bp_id': 'GO:0007095',
 'bp_label': 'mitotic G2 DNA damage checkpoint signaling',
 'relation_label': 'P38398--GO:0007095',
 'database': 'goa/',
 'articles': 'pubmed/19261748; pubmed/17643121; pubmed/17525340'}]

Summary table:

Variable	Rol	Type	Description	Ontology property	Graph
"protein"	**input**	**string**	**Protein name in entry name format**	**obo:RO_0002331 (involved in)**	**prot2bp**
"bp_id"	**output**	**string**	**Gene Ontology identifier of the biological process**	**obo:RO_0002331 (involved in)**	**prot2bp**

			related to the protein of interest.		
"bp_label"	output	string	Biological process label in the graph	skos:label	go
"relation_label"	output	string	Label that presents the relationship between protein and biological process of interest	skos:prefLabel	prot2bp
"database"	output	string	Database with relationship information	sio:SIO_0000253 (has source)	prot2bp
"articles"	output	string	Articles associated to the relationship	sio:SIO_0000772 (has evidence)	prot2bp

Function bp2prot(biological_process,taxon)

Description: This function returns the proteins related to a specific biological process by exploiting the information in the graph **http://rdf.biogateway.eu/graph/prot2bp.**

Parameters:

-**"biological_process"**: This parameter is the biological process of interest, allowed values are: its identifier in Gene Ontology (Example: 'GO:0035349') or the name of a biological process (Example: "coenzyme A transmembrane transport").

-**"taxon":** Allows you to select the taxon on which you want to perform the query. Value can be the NCBI taxonomic identifier or the name of the organism (property: label). Example: 'Homo sapiens', '9606'. If the value of the taxon is None, it will apply the search with all available taxa in the knowledge network.

Output:

The function returns a list of proteins related to the specified biological process. It presents the following fields:

-"**protein_name**": The function returns a list of the names in Uniprot entry name format (prefLabel) of the proteins related to biological process of interest (property: RO_0002331 (involved in)).

-**"relation_label":** Refers to the label presenting the relationship between the protein and the biological process of interest (property: prefLabel) available in the graph **http://rdf.biogateway.eu/graph/prot2bp.**

-**"database":** Indicates the database where the information on the relationship between protein and the biological process of interest (property: SIO_000253 (has source)).

-**"articles":** Corresponds to scientific articles or publications that are associated with the relationship between the protein and the biological process of interest, found in the Pubmed database (property: SIO_000772 (has evidence)).

Implementation example:

Input: **bp2prot**("GO:0043524","Homo sapiens")
Output:
[{'protein_name': 'FZD9_HUMAN',
  'relation_label': 'O00144--GO:0043524',
  'database': 'goa/',

'articles': 'pubmed/27509850'},
{'protein_name': 'HTRA2_HUMAN',
 'relation_label': 'O43464--GO:0043524',
 'database': 'goa/',
 'articles': 'pubmed/18221368'},
{'protein_name': 'NGF_HUMAN',
 'relation_label': 'P01138--GO:0043524',
 'database': 'goa/',
 'articles': 'pubmed/21873635'},
{'protein_name': 'GDNF_HUMAN',
 'relation_label': 'P39905--GO:0043524',
 'database': 'goa/',
 'articles': 'pubmed/8493557; pubmed/21873635'}]

Summary table:

Variable	Rol	Type	Description	Ontology property	Graph
"biological_process"	**input**	**string**	**Identifier in Gene Ontology or name of the biological process**	**obo:RO_0002331 (involved in)**	**prot2bp**
"taxon"	**input (optional)**	**string**	**Taxon number or organism name**	**obo:RO_0002162 (in taxon)**	**prot**
"protein_name"	**output**	**string**	**Protein name related to the biological**	**obo:RO_0002331 (involved in)**	**prot2bp**

			process in Uniprot entry name format		
"relation_label"	output	string	Label that presents the relationship between the protein and the biological process of interest.	skos:prefLabel	prot2bp
"database"	output	string	Database with relationship information	sio:SIO_0000253 (has source)	prot2bp
"articles"	output	string	Articles associated to the relationship	sio:SIO_0000772 (has evidence)	prot2bp

Function prot2cc(protein)

Description: This function facilitates the study of the cellular components in which the protein of interest participates, exploiting the information of the graph **http://rdf.biogateway.eu/graph/prot2cc**

Parameters:

-**"protein":** We enter the protein name either in Uniprot's entry name format (property: prefLabel) or in Uniprot's entry format. (property: altLabel). Example: "BRCA1_HUMAN", "P38398".

Output:

The function returns a list of cellular components related to the selected protein. It presents the following fields:

-**"cc_id":** Corresponds to the cellular component identifier in Gene Ontology (property: oboInOwl#id) related to our protein (property: BFO_0000050). Example: "GO:0005634"

-**"cc_label":** It refers to the label of the cellular component in the knowledge network **http://rdf.biogateway.eu/graph/go** (property: label). Example: "nucleus".

-**"relation_label":** Refers to the label presenting the relation between the protein and the cellular component of interest (property: prefLabel) available in the graph **http://rdf.biogateway.eu/graph/prot2mf .**

-**"database":** Indicates the database where the information on the relationship between protein and the cellular component of interest (property: SIO_000253 (has source)).

-**"articles":** Corresponds to scientific articles or publications that are associated with the relationship between the protein and the cellular component of interest, found in the Pubmed database (property: SIO_000772 (has evidence)).

Implementation example:

Input: **prot2cc**("BRCA1_HUMAN")
Output:
[{'cc_id': 'GO:0000151',
  'cc_label': 'ubiquitin ligase complex',
  'relation_label': 'P38398--GO:0000151',

'database': 'goa/',
 'articles': 'pubmed/14976165'},
{'cc_id': 'GO:0000152',
 'cc_label': 'nuclear ubiquitin ligase complex',
 'relation_label': 'P38398--GO:0000152',
 'database': 'goa/',
 'articles': 'pubmed/14636569'},
{'cc_id': 'GO:0000800',
 'cc_label': 'lateral element',
 'relation_label': 'P38398--GO:0000800',
 'database': 'goa/',
 'articles': 'pubmed/9774970'},
{'cc_id': 'GO:0000931',
 'cc_label': 'gamma-tubulin ring complex',
 'relation_label': 'P38398--GO:0000931',
 'database': 'goa/',
 'articles': 'pubmed/12214252'}]

Summary table:

Variable	Rol	Type	Description	Ontology property	Graph
"protein"	**input**	**string**	**Protein name in entry name format de Uniprot**	**obo:RO_0002331 (involved in)**	**prot2cc**
"cc_id"	**output**	**string**	**Gene Ontology identifier of the cellular component related to the interest protein**	**obo:RO_0002331 (involved in)**	**prot2cc**
"cc_label"	**output**	**string**	**Biological process label of the cellular component in the graph**	**skos:label**	**go**

"relation_label"	output	string	Label that presents the relationship between the protein and the cellular component of interest	skos:prefLabel	prot2cc
"database"	output	string	Database with relationship information	sio:SIO_0000253 (has source)	prot2cc
"articles"	output	string	Articles associated to the relationship	sio:SIO_0000772 (has evidence)	prot2cc

Function cc2prot(cellular_component,taxon)

Description: This function returns the proteins related to a specific cellular component by exploiting the information in the graph **http://rdf.biogateway.eu/graph/prot2cc.**

Parameters:

-**"cellular_component"**: This parameter is the cell component of interest, allowed values are: its identifier in Gene Ontology (Example: 'GO:0034703') or the name of the cell component (Example: "cation channel complex").

-**"taxon":** Allows you to select the taxon on which you want to perform the query. Value can be the NCBI taxonomic identifier or the name of the organism (property: label). Example: 'Homo sapiens', '9606'. If the value of the taxon is None, it will apply the search with all available taxa in the knowledge network.

Output:

Function returns a list of proteins related to the specified cellular component. It has the following fields:

-"**protein_name**": The function returns a list of the names in Uniprot entry name format (prefLabel) of the proteins related to el componente celular de interés (property: BFO_0000050(part of)).

-"**relation_label**:" Refers to the label presenting the relation between the protein and the cellular component of interest (property: prefLabel) available in the graph **http://rdf.biogateway.eu/graph/prot2cc .**

-"**database**:" Indicates the database where the information on the relationship between protein and the cellular component of interest (property: SIO_000253 (has source)).

-"**articles**:" Corresponds to scientific articles or publications that are associated with the relationship between the protein and the cellular component of interest, found in the Pubmed database (property: SIO_000772 (has evidence)).

Implementation example:

Input: **cc2prot**("GO:0034703","9606")
Output:
[{'protein_name': 'TRPC3_HUMAN',
 'relation_label': 'Q13507--GO:0034703',
 'database': 'goa/',
 'articles': 'pubmed/21873635'},
 {'protein_name': 'PKD2_HUMAN',
 'relation_label': 'Q13563--GO:0034703',
 'database': 'goa/',
 'articles': 'pubmed/30093605'},
 {'protein_name': 'UNC80_HUMAN',
 'relation_label': 'Q8N2C7--GO:0034703',
 'database': 'goa/',
 'articles': 'pubmed/21873635'},
 {'protein_name': 'TRPC6_HUMAN',
 'relation_label': 'Q9Y210--GO:0034703',
 'database': 'goa/',
 'articles': 'pubmed/21873635'},

{'protein_name': 'PKD1_HUMAN',
 'relation_label': 'P98161--GO:0034703',
 'database': 'goa/',
 'articles': 'pubmed/30093605'}]

Summary table:

Variable	Rol	Type	Description	Ontology property	Graph
"cellular_component"	input	string	Gene Ontology identifier or cellular component name	obo:RO_0002331 (involved in)	prot2cc
"taxon"	input (optional)	string	Taxon number or organism name	obo:RO_0002162 (in taxon)	prot
"protein_name"	output	string	Protein name related to the cellular component in Uniprot entry name format	obo:RO_0002331 (involved in)	prot2cc
"relation_label"	output	string	Label presenting the relationship between the protein and the cellular component of interest	skos:prefLabel	prot2cc

"database"	output	string	Database with relationship information	sio:SIO_0000253 (has source)	prot2cc
"articles"	output	string	Articles associated to the relationship	sio:SIO_0000772 (has evidence)	prot2cc

Function prot2mf(protein)

Description: This function allows the study of the molecular functions in which the protein of interest participates, exploiting the information of the graph **http://rdf.biogateway.eu/graph/prot2mf**

Parameters:

-**"protein":** We enter the protein name either in Uniprot's entry name format (property: prefLabel) or in Uniprot's entry format. (property: altLabel). Example: "BRCA1_HUMAN", "P38398".

Output:

The function returns a list of molecular functions related to the selected protein. It presents the following fields:

-**"mf_id":** Corresponds to the molecular function identifier in Gene Ontology (property: oboInOwl#id) of the molecular functions related to the protein of interest (property:RO_0002327 (enables)). Example: "GO:0004672"

-**"mf_label":** Refers to the label of the molecular function in the knowledge network **http://rdf.biogateway.eu/graph/go** (property: label). Example: "protein kinase activity"

-"**relation_label**": Refers to the label presenting the relationship between the molecular function and the protein of interest (property: prefLabel) available in the graph **http://rdf.biogateway.eu/graph/prot2mf.**

-"**database**": Indicates the database where the information on the relationship between the molecular function and the protein of interest is registered (property: SIO_000253 (has source)).

-"**articles**": Corresponds to scientific articles or publications that are associated with the relationship between the molecular function and the protein of interest, found in the Pubmed database (property: SIO_000772 (has evidence)).

Implementation example:

Input: **prot2mf**("BRCA1_HUMAN")
Output:
{'mf_id': 'GO:0002039',
 'mf_label': 'p53 binding',
 'relation_label': 'P38398--GO:0002039',
 'database': 'goa/',
 'articles': 'pubmed/15571721'},
{'mf_id': 'GO:0003677',
 'mf_label': 'DNA binding',
 'relation_label': 'P38398--GO:0003677',
 'database': 'goa/',
 'articles': 'pubmed/9662397'},
{'mf_id': 'GO:0003713',
 'mf_label': 'transcription coactivator activity',
 'relation_label': 'P38398--GO:0003713',
 'database': 'goa/',
 'articles': 'pubmed/9662397'},
{'mf_id': 'GO:0003723',
 'mf_label': 'RNA binding',
 'relation_label': 'P38398--GO:0003723',
 'database': 'goa/',
 'articles': 'pubmed/12419249'},

{'mf_id': 'GO:0004842',
 'mf_label': 'ubiquitin-protein transferase activity',
 'relation_label': 'P38398--GO:0004842',
 'database': 'goa/',
  'articles': 'pubmed/20351172; pubmed/21873635; pubmed/17349954; pubmed/19117993; pubmed/12890688'}]

Summary table:

Variable	Rol	Type	Description	Ontology property	Graph
"protein"	**input**	**string**	**Protein name in entry name format**	**obo:RO_0002331 (involved in)**	**prot2mf**
"mf_id"	**output**	**string**	**Gene Ontology identifier of molecular function related to the protein of interest**	**obo:RO_0002331 (involved in)**	**prot2mf**
"mf_label"	**output**	**string**	**Molecular function label**	**skos:label**	**go**
"relation_label"	**output**	**string**	**Label presenting the relationship between the protein and the molecular function of interest**	**skos:prefLabel**	**prot2mf**
"database"	**output**	**string**	**Database with relationship information**	**sio:SIO_0000253 (has source)**	**prot2mf**
"articles"	**output**	**string**	**Articles associated to the relationship**	**sio:SIO_0000772 (has evidence)**	**prot2mf**

Function mf2prot(molecular_function,taxon)

Description: This function returns the proteins related to the specified molecular function by exploiting the information in the graph **http://rdf.biogateway.eu/graph/prot2mf**

Parameters:

-**"molecular_function"**: This parameter is the molecular function of interest, the allowed values are: its identifier in Gene Ontology (Example: 'GO:0004672') or the name of the cellular component (Example: 'protein kinase activity').

-**"taxon":** Allows you to select the taxon on which you want to perform the query. Value can be the NCBI taxonomic identifier or the name of the organism (property: label). Example: 'Homo sapiens', '9606'. If the value of the taxon is None, it will apply the search with all available taxa in the knowledge network.

Output:

The function returns a list of proteins related to the molecular function of interest. It presents the following fields:

-**"protein_name"**: These are names in Uniprot entry name format (prefLabel) of the proteins related to the cellular component of interest (property:RO_0002327).

-**"relation_label":** Refers to the label presenting the relationship between the protein and the molecular function of interest (property: prefLabel) available in the graph **http://rdf.biogateway.eu/graph/prot2mf .**

-**"database":** Indicates the database where the information on the relationship between protein and the molecular function of interest (property: SIO_000253 (has source)).

-**"articles":** Corresponds to scientific articles or publications that are associated with the relationship between the protein and the molecular function of interest, found in the Pubmed database (property: SIO_000772 (has evidence)).

Implementation example:

Input: **mf2prot**("protein binding","9606")

Output:

[{'protein_name': 'ZN217_HUMAN',
  'relation_label': 'O75362--GO:0005515',
  'database': 'goa/',
  'articles': 'pubmed/16940172'},
 {'protein_name': 'SIR6_HUMAN',
  'relation_label': 'Q8N6T7--GO:0005515',
  'database': 'goa/',
  'articles': 'pubmed/19135889; pubmed/23217706; pubmed/23911928'},
 {'protein_name': 'NELFE_HUMAN',
  'relation_label': 'P18615--GO:0005515',
  'database': 'goa/',
   'articles': 'pubmed/32296183; pubmed/26496610; pubmed/14667819; pubmed/28514442;
pubmed/25416956; pubmed/24981860; pubmed/12612062; pubmed/20211142'}]

Summary table:

Variable	Rol	Type	Description	Ontology property	Graph
"molecular_function"	**input**	**string**	**Gene Ontology identifier or molecular function name**	**obo:RO_0002331 (involved in)**	**prot2mf**

"taxon"	input (optional)	string	Taxon number or organism name	obo:RO_0002162 (in taxon)	prot
"protein_name"	output	string	Protein name related to the molecular function in Uniprot entry name format	obo:RO_0002331 (involved in)	prot2mf
"relation_label"	output	string	Label presenting the relationship between the protein and the molecular function of interest	skos:prefLabel	prot2mf
"database"	output	string	Database with relationship information	sio:SIO_0000253 (has source)	prot2mf
"articles"	output	string	Articles associated to the relationship	sio:SIO_0000772 (has evidence)	prot2mf

Function gene2crm(gene)

Description: This function allows us to obtain the crms associated with the gene entered as a parameter. To do this, we will use the information from the graph **http://rdf.biogateway.eu/graph/crm2gene** .

Parameters:

-**"gene":** This parameter corresponds to the gene name in symbol format (property: prefLabel). Example: "TOX3".

Output:

The function returns a dictionary with the crms that are related to the entered gene. It presents the following fields:

-**"crm_name":** Cis-regulator module name (property: prefLabel) which is related to the inserted gene (property: RO_0002429).

-**"database":** Indicates the database where the information on the relationship between the crm and the gene of interest is registered (property: SIO_000253 (has source)).

-**"articles":** Corresponds to scientific articles or publications that are associated with the relationship between the crm and the gene of interest, found in the Pubmed database (property: SIO_000772 (has evidence)).

Implementation example:

Input: **gene2crm**("TOX3")
Output:
[{'crm_name': 'crm/CRMHS00000005857',
      'database': 'http://lcbb.swjtu.edu.cn/EnhFFL/; http://218.8.241.248:8080/SEA3/; http://bioinfo.vanderbilt.edu/AE/HACER;
https://genome.ucsc.edu/cgi-bin/hgTrackUi?db=hg38&g=geneHancer;
http://enhanceratlas.net/scenhancer/; https://webs.iiitd.edu.in/raghava/cancerend/;
https://enhancer.lbl.gov/; http://www.licpathway.net/sedb/;
http://health.tsinghua.edu.cn/jianglab/endisease/;
http://biocc.hrbmu.edu.cn/DiseaseEnhancer/; https://fantom.gsc.riken.jp/5/;
http://acgt.cs.tau.ac.il/focs/; https://asntech.org/dbsuper/;
http://yiplab.cse.cuhk.edu.hk/jeme/',
  'articles': 'pubmed/24119843; pubmed/28869592; pubmed/30247654; pubmed/23374354;
pubmed/29716618; pubmed/23001124; pubmed/35694152; pubmed/24670763;

pubmed/32360910; pubmed/17130149; pubmed/30371817; pubmed/28605766; pubmed/34761274; pubmed/31667506'}

Summary table:

Variable	Rol	Type	Description	Ontology property	Graph
"gene"	**input**	**string**	**Gene name in symbol format**	**obo:RO_0002429 (involved in positive regulation of)**	**crm2gene**
"crm_name"	**output**	**string**	**Gene related cis-regulatory module name**	**obo:RO_0002429 (involved in positive regulation of)**	**crm2gene**
"database"	**output**	**string**	**Database with relationship information between crm and gene**	**sio:SIO_0000253 (has source)**	**crm2gene**
"articles"	**output**	**string**	**Articles associated to the relationship**	**sio:SIO_0000772 (has evidence)**	**crm2gene**

Function crm2gene(crm)

Description: This function makes it easy to obtain the genes that a specific crm affects.

To do so, we will use the information in the graph **http://rdf.biogateway.eu/graph/crm2gene .**

Parameters:

-**"crm"**: Parameter entered in the function shall be the preferred identifier of the cis-regulator module. (property: prefLabel). Example"crm/CRMHS00003225754".

Output:

It returns a dictionary with the genes that are related to the crm of interest. It presents the following fields:

-"**gene_name**": The function returns a list with the names in symbol format (property: prefLabel) of the genes related to the crm of interest (property:RO_0002429).

-**"database"**: Indicates the database where the information on the relationship between the gene and the crm of interest is registered (property: SIO_000253 (has source)).

-**"articles"**: Corresponds to scientific articles or publications that are associated with the relationship between the gene and the crm of interest, found in the Pubmed database (property: SIO_000772 (has evidence)).

Implementation example:

Input: **crm2gene**("crm/CRMHS00000137026")

Output:

[{'gene_name': 'CRAT',

'database': 'http://bioinfo.vanderbilt.edu/AE/HACER; https://fantom.gsc.riken.jp/5/; https://webs.iiitd.edu.in/raghava/cancerend/; http://yiplab.cse.cuhk.edu.hk/jeme/',

 'articles': 'pubmed/24670763; pubmed/32360910; pubmed/28869592; pubmed/30247654'},

{'gene_name': 'NTMT1',

'database': 'http://bioinfo.vanderbilt.edu/AE/HACER; https://fantom.gsc.riken.jp/5/; https://webs.iiitd.edu.in/raghava/cancerend/; http://yiplab.cse.cuhk.edu.hk/jeme/',

 'articles': 'pubmed/24670763; pubmed/32360910; pubmed/28869592; pubmed/30247654'},

{'gene_name': 'PTPA',

'database': 'http://bioinfo.vanderbilt.edu/AE/HACER; https://fantom.gsc.riken.jp/5/; https://webs.iiitd.edu.in/raghava/cancerend/; http://yiplab.cse.cuhk.edu.hk/jeme/',

'articles': 'pubmed/24670763; pubmed/32360910; pubmed/28869592; pubmed/30247654'

Summary table:

Variable	Rol	Type	Description	Ontology property	Graph
"crm"	**input**	**string**	**Cis-regulator module name**	**obo:RO_0002429 (involved in positive regulation of)**	**crm2gene**
"gene_name"	**output**	**string**	**Gene name in symbol format related to crm**	**obo:RO_0002429 (involved in positive regulation of)**	**crm2gene**
"database"	**output**	**string**	**Database with relationship information between crm and gene**	**sio:SIO_0000253 (has source)**	**crm2gene**
"articles"	**output**	**string**	**Articles associated to the relationship**	**sio:SIO_0000772 (has evidence)**	**crm2gene**

Function tfac2crm(tfac)

Description: The function makes it possible to obtain the crms associated with a specific transcription factor, information that can be found in the graph **http://rdf.biogateway.eu/graph/crm2tfac.**

Parameters:

-"tfac": This parameter corresponds to the name in Uniprot entry name format of the transcription factor with which we want to know which crms interacts with.

Output:

The function returns a dictionary with the crms that are related to the entered transcription factor. It presents the following fields:

-"crm_name": Name of the cis-regulatory modules ((property: prefLabel)) that the specified transcription factor interacts with (property:RO_0002436).

-"database": Indicates database where the information on the relationship between the crm and the transcription factor is registered (property: SIO_000253 (has source)).

-"articles": Corresponds to scientific articles or publications that are associated with the relationship between the crm and the transcription factor of interest, found in the Pubmed database (property: SIO_000772 (has evidence)).

-"evidence": This corresponds to the evidence supporting the available information on the relationship between the transcription factor and the crm (property: evidenceOrigin).

-"biological_samples": Refers to the different types of biological samples that are associated with the study of the relationship between the cis-regulator module and the transcription factor (property: TXPO_0003500 (observed in)). Specifically, it will return the identifiers in ontology term format. Example: "CLO_0001601", "UBERON_0002113", "BTO_0000018".

Implementation example:

Input: **tfac2crm**("FOSL2_HUMAN")
Output:
[{'crm_name': 'crm/CRMHS00000005425',
  'database': 'http://www.licpathway.net/ENdb/',

  'articles': 'pubmed/29149598',
'evidence':
'http://www.licpathway.net/ENdb/search/Detail.php?Species=Human&Enhancer_id=E_01_29 2',
  'biological_samples': 'UBERON_0001003; CL_0000148'},
 {'crm_name': 'crm/CRMHS00000006865',
  'database': 'http://lcbb.swjtu.edu.cn/EnhFFL/',
  'articles': 'pubmed/35694152',
'evidence':
'http://lcbb.swjtu.edu.cn/EnhFFL/details/?term=enH639113&subtype=enhancer&species=hu man',
  'biological_samples': 'CL_0002319; UBERON_0002421; BTO_0000601'},
 {'crm_name': 'crm/CRMHS00000006866',
  'database': 'http://lcbb.swjtu.edu.cn/EnhFFL/',
  'articles': 'pubmed/35694152',
'evidence':
'http://lcbb.swjtu.edu.cn/EnhFFL/details/?term=enH594653&subtype=enhancer&species=hu man',
  'biological_samples': 'CL_0002319; UBERON_0002421; BTO_0000601'}]

Summary table:

Variable	Rol	Type	Description	Ontology property	Graph
"tfac"	**input**	**string**	**Transcription factor name in Uniprot entry name format (tfac)**	**obo:RO_0002436 (molecularly interacts with)**	**crm2tfac**

"crm_name"	output	string	Name of the cis-regulatory module related to the introduced transcription factor	obo:RO_0002436 (molecularly interacts with)	crm2tfac
"database"	output	string	Database with relationship information between crm and tfac	sio:SIO_0000253 (has source)	crm2tfac
"articles"	output	string	Articles associated to the relationship	sio:SIO_0000772 (has evidence)	crm2tfac
"evidence"	output	string	Evidence level associated with the relationship	sch:evidenceOrigin	crm2tfac
"biological_samples"	output	string	Biological samples associated with the relationship	obo:TXPO_0003500 (observed in)	crm2tfac

Function crm2tfac(crm)

Description: Esta función permite estudiar los factores de transcripción que interactúan con un crm determinado, usando la información del grafo **http://rdf.biogateway.eu/graph/crm2tfac.**

Parameters:

-**"crm"**: Parameter entered in the function shall be the preferred identifier of the cis-regulator module. (property: prefLabel). Example"crm/CRMHS00000007832".

Output:

La función devuelve un diccionario con los factores de transcripción que están relacionados con el crm de interés. Presenta los siguientes campos:

-**"tfac_name":** Nombres en formato entry name de Uniprot de los factores de transcripción que interactúan con el crm de interés. Example: "TF7L2_HUMAN".

-**"database":** Indica la base de datos donde se encuentra registrada la información sobre la relación entre el factor de transcripción y el crm de interés (property: SIO_000253 (has source)).

-**"articles":** Corresponde a artículos científicos o publicaciones que están asociadas con la relación entre el factor de transcripción y el crm de interés, found in the Pubmed database (property: SIO_000772 (has evidence)).

-**"evidence":** Corresponde a la evidencia que respalda la información disponible sobre la relación entre el factor de transcripción y el crm de interés(property: evidenceOrigin).

-**"biological_samples":** Se refiere a los diferentes types de muestras biológicas que están asociados con el estudio de la relación entre el factor de transcripción y el crm introducido (property: TXPO_0003500 (observed in)). En concreto, devolverá los identificadores en formato de términos ontológicos. Example: "CLO_0001601", "UBERON_0002113", "BTO_0000018".

Implementation example:

Input: crm2tfac("crm/CRMHS00000007832")
Output:
[{'tfac_name': 'RAD21_HUMAN',
  'database': 'http://lcbb.swjtu.edu.cn/EnhFFL/',
  'articles': 'pubmed/35694152',

'evidence':

'http://lcbb.swjtu.edu.cn/EnhFFL/details/?term=enH58066&subtype=enhancer&species=human',

  'biological_samples': 'UBERON_0005090; CL_0000187; BTO_0000887'},

 {'tfac_name': 'SRF_HUMAN',

  'database': 'http://lcbb.swjtu.edu.cn/EnhFFL/',

  'articles': 'pubmed/35694152',

'evidence':

'http://lcbb.swjtu.edu.cn/EnhFFL/details/?term=enH58066&subtype=enhancer&species=human',

  'biological_samples': 'UBERON_0005090; CL_0000187; BTO_0000887'},

 {'tfac_name': 'TAF1_HUMAN',

  'database': 'http://lcbb.swjtu.edu.cn/EnhFFL/',

  'articles': 'pubmed/35694152',

'evidence':

'http://lcbb.swjtu.edu.cn/EnhFFL/details/?term=enH58066&subtype=enhancer&species=human',

  'biological_samples': 'UBERON_0005090; CL_0000187; BTO_0000887'},

 {'tfac_name': 'RPB1_HUMAN',

  'database': 'http://lcbb.swjtu.edu.cn/EnhFFL/',

  'articles': 'pubmed/35694152',

'evidence':

'http://lcbb.swjtu.edu.cn/EnhFFL/details/?term=enH58066&subtype=enhancer&species=human',

  'biological_samples': 'UBERON_0005090; CL_0000187; BTO_0000887'}]

Summary table:

Variable	Rol	Type	Description	Ontology property	Graph
"crm"	**input**	**string**	**Nombre del módulo cis-regulador**	**obo:RO_0002436 (molecularly interacts with)**	**crm2tfac**
"tfac_name"	**output**	**string**	**Nombre del del factor de**	**obo:RO_0002436 (molecularly**	**crm2tfac**

			transcripción en formato entry name de Uniprot relacionado con el crm introducido	interacts with)	
"database"	output	string	Database with relationship information entre el crm y el tfac.	sio:SIO_0000253 (has source)	crm2tfac
"articles"	output	string	Articles associated to the relationship	sio:SIO_0000772 (has evidence)	crm2tfac
"evidence"	output	string	Evidence level associated with the relationship	sch:evidenceOrigin	crm2tfac
"biological_samples"	output	string	Biological samples associated with the relationship	obo:TXPO_0003500 (observed in)	crm2tfac

Function crm2phen(crm)

Description: This function makes it possible to obtain the phenotypes associated with a crm previously introduced as a parameter of this function. To do so, we will exploit the graph **http://rdf.biogateway.eu/graph/crm2phen.**

Parameters:

-**"crm"**: Parameter entered in the function shall be the preferred identifier of the cis-regulator module (property: prefLabel). Example: "crm/CRMHS00000005764".

Output:

The function returns a dictionary with the phenotypes related to the selected crm. It also presents the following fields:

-**"phen_id":** Corresponds to the identifiers of the phenotypes associated to the crm of interest, specifically it can return either the OMIM identifier, the DOID identifier or the MeSH identifier.

-**"database":** Indicates the database where the information on the relationship between the phenotype and the crm of interest is registered (property: SIO_000253 (has source)).

-**"articles":** Corresponds to scientific articles or publications that are associated with the relationship between the phenotype and the crm of interest, found in the Pubmed database (property: SIO_000772 (has evidence)).

Implementation example:

Input: **crm2phen**("crm/CRMHS00000005764")
Output:
[{'phen_id': 'OMIM/181500',
'database':'http://biocc.hrbmu.edu.cn/DiseaseEnhancer/;
http://health.tsinghua.edu.cn/jianglab/endisease/',
 'articles': 'pubmed/25453756'},
 {'phen_id': 'MESH/D012559',
'database':'http://biocc.hrbmu.edu.cn/DiseaseEnhancer/;
http://health.tsinghua.edu.cn/jianglab/endisease/',
 'articles': 'pubmed/25453756'},
 {'phen_id': 'DOID/DOID_5419',
'database':'http://biocc.hrbmu.edu.cn/DiseaseEnhancer/;
http://health.tsinghua.edu.cn/jianglab/endisease/',
 'articles': 'pubmed/25453756'}]

Summary table:

Variable	Rol	Type	Description	Ontology property	Graph
"crm"	**input**	**string**	**Cis-regulator module name**	**obo:RO_0002331 (involved in)**	**crm2phen**
"phen_id"	**output**	**string**	**Phenotypes identifier associated with the crm of interest**	**obo:RO_0002331 (involved in)**	**crm2phen**
"database"	**output**	**string**	**Database with relationship information between crm and phenotype**	**sio:SIO_0000253 (has source)**	**crm2phen**
"articles"	**output**	**string**	**Articles associated to the relationship**	**sio:SIO_0000772 (has evidence)**	**crm2phen**

Function phen2crm(phenotype)

Description: This function returns the crms that are associated with a given phenotype.

It exploits the information available in the graph **http://rdf.biogateway.eu/graph/crm2phen.**

Parameters:

-"**phenotype**": This parameter corresponds to the phenotype of interest. Both its OMIM identifier ('181500') and a disease name ('schizophrenia') are allowed.

Output:

This function returns a dictionary with the crms that are related to the specified phenotype. It presents the following fields:

-"**crm_name**:" Name of the cis-regulatory modules ((property: prefLabel)) associated to the entered phenotype (property: RO_0002331 (involved in)).

-"**omim_id**": Corresponds to the OMIM identifier of the phenotype that is associated to the crm (only if the name of a phenotype has been entered).

-"**database**": Indicates the database where the information on the relationship between the crm and the specified fenotype is registered. (property: SIO_000253 (has source)).

-"**articles**:" Corresponds to scientific articles or publications that are associated with the relationship between the crm and the phenotype of interest, found in the Pubmed database (property: SIO_000772 (has evidence)).

Implementation example:

Input: **phen2crm**("schizophrenia")

Output:

[{'crm_name': 'crm/CRMHS00000005764',

  'omim_id': 'OMIM/181500',

'database':'http://biocc.hrbmu.edu.cn/DiseaseEnhancer/;

http://health.tsinghua.edu.cn/jianglab/endisease/',

  'articles': 'pubmed/25453756'},

 {'crm_name': 'crm/CRMHS00000005770',

  'omim_id': 'OMIM/181500',

'database':'http://biocc.hrbmu.edu.cn/DiseaseEnhancer/;

http://health.tsinghua.edu.cn/jianglab/endisease/',

  'articles': 'pubmed/25453756'},

 {'crm_name': 'crm/CRMHS00000005771',

'omim_id': 'OMIM/181500',

'database':'http://biocc.hrbmu.edu.cn/DiseaseEnhancer/;

http://health.tsinghua.edu.cn/jianglab/endisease/',

  'articles': 'pubmed/25434007'},

 {'crm_name': 'crm/CRMHS00000005773',

  'omim_id': 'OMIM/181500',

'database':'http://biocc.hrbmu.edu.cn/DiseaseEnhancer/;

http://health.tsinghua.edu.cn/jianglab/endisease/',

  'articles': 'pubmed/25453756'},

 {'crm_name': 'crm/CRMHS00000005816',

  'omim_id': 'OMIM/181500',

'database':'http://biocc.hrbmu.edu.cn/DiseaseEnhancer/;

http://health.tsinghua.edu.cn/jianglab/endisease/',

  'articles': 'pubmed/27276213; pubmed/25453756'}]

Summary table:

Variable	Rol	Type	Description	Ontology property	Graph
"phenotype"	input	string	OMIM identifier or phenotype name of interest	obo:RO_0002331 (involved in)	crm2phen
"crm_name"	output	string	Name of the cis-regulator module related to the entered phenotype	obo:RO_0002331 (involved in)	crm2phen
"omim_id"	output (optional)	string	Phenotype OMIM identifier	obo:RO_0002331 (involved in)	crm2phen

			associated to the crm		
"database"	output	string	Database with relationship information between crm and phenotype	sio:SIO_0000253 (has source)	crm2phen
"articles"	output	string	Articles associated to the relationship	sio:SIO_0000772 (has evidence)	crm2phen

Function tfac2gene(tfac)

Description: Esta función permite obtener los genes que son regulados por un factor de transcripción determinado, gracias a la información proporcionada por el grafo **http://rdf.biogateway.eu/graph/tfac2gene.**

Parameters:

-**"tfac":** This parameter corresponds to the name in the Uniprot entry name format of the transcription factor that we want to know which genes it regulates. Example: "NKX31_HUMAN".

Output:

The function returns two dictionaries with the genes that are positively and negatively regulated by the transcription factor entered. It presents the following fields:

-"gene_name": Name in symbol format of the genes ((property: prefLabel)) that are regulated by the specified transcription factor (property: RO_0002429 (involved in positive regulation of) and property: RO_0002430 (involved in negative regulation of)).

-"database": Indicates the database where the information on the relationship between the gene and the transcription factor is recorded (property: SIO_000253 (has source)).

-"articles": Corresponds to scientific articles or publications that are associated with the relationship between the gene and the transcription factor of interest, found in the Pubmed database (property: SIO_000772 (has evidence)).

-"evidence_level": Corresponds to the evidence supporting the available information on the relationship between the transcription factor and the gene (property: evidenceOrigin).

-"definition": Provides the definition of the relationship that occurs between each gene and the specified transcription factor (property: definition), available in the graph **http://rdf.biogateway.eu/graph/tfac2gene**.

Implementation example:

Input: **tfac2gene**("NKX31_HUMAN")
Output:
'Positive regulation results:',
 [{'gene_name': 'CLIC4',
  'database': 'https://github.com/saezlab/CollecTRI',
  'articles': 'pubmed/16316993',
  'evidence_level': '1',
  'definition': 'Q99801 involved in positive regulation of 9606/CLIC4'},
 {'gene_name': 'DKK3',
  'database': 'https://github.com/saezlab/CollecTRI',
  'articles': 'pubmed/23975733',
  'evidence_level': '1',
  'definition': 'Q99801 involved in positive regulation of 9606/DKK3'},

{'gene_name': 'MAP3K5',
 'database': 'https://github.com/saezlab/CollecTRI',
 'articles': 'pubmed/21594902',
 'evidence_level': '1',
 'definition': 'Q99801 involved in positive regulation of 9606/MAP3K5'},
 {'gene_name': 'NKX3-1',
 'database': 'https://github.com/saezlab/CollecTRI',
  'articles': 'pubmed/15880262; pubmed/20855495; pubmed/19886863; pubmed/20195545;
pubmed/23368843; pubmed/16270157; pubmed/19263243; pubmed/21730289;
pubmed/16763719; pubmed/20716579; pubmed/16845664',
 'evidence_level': '11',
 'definition': 'Q99801 involved in positive regulation of 9606/NKX3-1'}]

'Negative regulation results:',
 [{'gene_name': 'AR',
  'database': 'https://github.com/saezlab/CollecTRI',
   'articles': 'pubmed/23492366; pubmed/17202838; pubmed/20363913; pubmed/18360715;
pubmed/16697957',
  'evidence_level': '5',
  'definition': 'Q99801 involved in regulation of 9606/AR'},
 {'gene_name': 'BCL2',
  'database': 'https://github.com/saezlab/CollecTRI',
   'articles': 'pubmed/24273454; pubmed/17191317; pubmed/22869582; pubmed/16316993;
pubmed/22266868; pubmed/9581775; pubmed/15817464; pubmed/19137013;
pubmed/22331597; pubmed/23313858; pubmed/12679484; pubmed/24098340;
pubmed/17486276; pubmed/14684736; pubmed/8183578; pubmed/19266349;
pubmed/21940310',
  'evidence_level': '17',
  'definition': 'Q99801 involved in regulation of 9606/BCL2'},
 {'gene_name': 'CCND1',
  'database': 'https://github.com/saezlab/CollecTRI',
  'articles': 'pubmed/17639064; pubmed/22179513',
  'evidence_level': '2',
  'definition': 'Q99801 involved in regulation of 9606/CCND1'}]

Summary table:

Variable	Rol	Type	Description	Ontology property	Graph
"tfac"	**input**	**string**	**Transcription factor name in Uniprot entry name format**	**obo:RO_0002429 (involved in positive regulation of) and obo:RO_0002430 (involved in negative regulation of)**	**tfac2gene**
"gene_name"	**output**	**string**	**Gene name in symbol format related to transcription factor**	**obo:RO_0002429(involved in positive regulation of) y obo:RO_0002430(involved in negative regulation of)**	**tfac2gene**
"database"	**output**	**string**	**Database with relationship information between transcription factor and gene**	**sio:SIO_0000253 (has source)**	**tfac2gene**
"articles"	**output**	**string**	**Articles associated to the relationship**	**sio:SIO_0000772 (has evidence)**	**tfac2gene**
"evidence_level"	**output**	**integer**	**Evidence supporting the available information on**	**sch:evidenceLevel**	**tfac2gene**

			the relationship between the transcription factor and the gene		
"definition"	output	string	Relationship definition	skos:definition	tfac2gene

Function gene2tfac(gene, regulation_type)

Description: This function will return the transcription factors that regulate the entered gene, using the information available in the graph **http://rdf.biogateway.eu/graph/tfac2gene.**

Parameters:

-**"gene":** This parameter corresponds to the gene name in symbol format (property: prefLabel). Example: "TOX3".

-**"regulation_type**": It is an optional parameter, if 'positive' is entered, it will return those regulation relations that are positive, and if 'negative' is entered, those that are negative. If no parameter is entered, it will return all regulation relationships.

Output:

Function has as output two dictionaries with the transcription factors that positively regulate the gene in one of them, and those that negatively regulate it in the other. It presents the following fields:

-**"tfac_name":** Names in Uniprot entry name format of the transcription factors interacting with the entered gene (property: RO_0002429 (involved in positive regulation of) and property: RO_0002430 (involved in negative regulation of)).

-**"database":** Indicates the database where the information about the relationship between the gene and the transcription factor is registered (property: SIO_000253 (has source)).

-**"articles":** Corresponds to scientific articles or publications that are associated with the relationship between the transcription factor and the gene of interest, found in the Pubmed database (property: SIO_000772 (has evidence)).

-**"evidence_level":** Corresponds to the evidence supporting the available information on the relationship between transcription factor and gene (property: evidenceOrigin).

-**"definition":** Provides the definition of the relationship that occurs between each transcription factor and the specified gene (property: definition), available in the graph **http://rdf.biogateway.eu/graph/tfac2gene**.

Implementation example:

Input: **gene2tfac**("BRCA1")
Output:
('Positive regulation results:',
 [{'tfac_name': 'BHE41_HUMAN',
   'database': 'https://github.com/saezlab/CollecTRI',
   'articles': 'pubmed/20006609',
   'evidence_level': '1',
   'definition': 'Q9C0J9 involved in positive regulation of 9606/BRCA1'},
  {'tfac_name': 'P63_HUMAN',
   'database': 'https://github.com/saezlab/CollecTRI',
   'articles': 'pubmed/24556685',
   'evidence_level': '1',
   'definition': 'Q9H3D4 involved in positive regulation of 9606/BRCA1'},
  {'tfac_name': 'MBD2_HUMAN',
   'database': 'https://github.com/saezlab/CollecTRI',
   'articles': 'pubmed/16052033; pubmed/23011797',

'evidence_level': '2',

    'definition': 'Q9UBB5 involved in positive regulation of 9606/BRCA1'}

'Negative regulation results:',

 [{'tfac_name': 'HMGA1_HUMAN',

  'database': 'https://github.com/saezlab/CollecTRI',

  'articles': 'pubmed/16007157; pubmed/12640109',

  'evidence_level': '2',

  'definition': 'P17096 involved in regulation of 9606/BRCA1'},

 {'tfac_name': 'HMGA1_HUMAN',

  'database': 'https://github.com/saezlab/CollecTRI',

  'articles': 'pubmed/16007157; pubmed/12640109',

  'evidence_level': '2',

  'definition': 'P17096 involved in negative regulation of 9606/BRCA1'},

 {'tfac_name': 'ID4_HUMAN',

  'database': 'https://github.com/saezlab/CollecTRI',

   'articles': 'pubmed/24475217; pubmed/12032322; pubmed/17016441; pubmed/11136250; pubmed/21194482; pubmed/16582598',

  'evidence_level': '6',

  'definition': 'P47928 involved in negative regulation of 9606/BRCA1'},

 {'tfac_name': 'LMO4_HUMAN',

  'database': 'https://github.com/saezlab/CollecTRI',

  'articles': 'pubmed/12925972; pubmed/11751867',

  'evidence_level': '2',

  'definition': 'P61968 involved in negative regulation of 9606/BRCA1'},

Summary table:

Variable	Rol	Type	Description	Ontology property	Graph
"gene"	input	string	Gene name in symbol format	obo:RO_0002429 (involved in positive regulation of) y obo:RO_0002430 (involved in negative regulation of)	tfac2gene

"regulation_type"	input	string	Regulation type (optional)	obo:RO_0002430 obo:RO_0002428 obo:RO_0002429	tfac2gene
"tfac_name"	output	string	Transcription factor name in Uniprot entry name format	obo:RO_0002429(involved in positive regulation of) y obo:RO_0002430(involved in negative regulation of)	tfac2gene
"database"	output	string	Database with relationship information between transcription factor and gene	sio:SIO_0000253 (has source)	tfac2gene
"articles"	output	string	Articles associated to the relationship	sio:SIO_0000772 (has evidence)	tfac2gene
"evidence_level"	output	integer	Evidence supporting the available information on the relationship between transcription factor and gene	sch:evidenceLevel	tfac2gene

"definition"	output	string	Relationship definition	skos:definition	tfac2gene

Function prot2prot(protein)

Description: This function will return the proteins that interact with the entered protein, using the information available in the graph **http://rdf.biogateway.eu/graph/prot2prot.**

Parameters:

-**"protein":** We enter the protein name either in Uniprot's entry name format (property: prefLabel) or in Uniprot's entry format. (property: altLabel). Example: "BRCA1_HUMAN", "P38398".

Output:

The function returns a list of proteins that interact with the protein of interest. It presents the following fields:

-"**protein_label**": The function returns a list with the protein names in Uniprot entry name format (prefLabel) of the proteins interacting with the entered protein (property: RO_0002436 (molecularly interacts with)).

-**"relation_label":** Refers to the label presenting the relationship between the two proteins (property: prefLabel) available in the graph **http://rdf.biogateway.eu/graph/prot2prot .**

-**"database":** Indicates the database where the information on the relationship between proteins is recorded (property: SIO_000253 (has source)).

-**"articles":** Corresponds to scientific articles or publications that are associated with the relationship between the proteins, found in the Pubmed database (property: SIO_000772 (has evidence)).

-**"evidence_level":** Corresponds to the evidence supporting the available information on the relationship between two proteins (property: evidenceLevel)

-**"interaction_details"**: Link to protein interaction information from the IntAct database (property: BFO_0000050(part of)).

Implementation example:

Input: **prot2prot**("BRCA1_HUMAN")
Output: [{'prot_label': 'CDK16_HUMAN',
 'relation_label': 'uniprot!P38398--uniprot!Q00536',
 'database': 'intact',
 'evidence_level': '0.35',
 'articles': 'pubmed/33961781',
 'interaction_details': 'http://identifiers.org/intact/EBI-54768436'},
 {'prot_label': 'FHL2_HUMAN',
 'relation_label': 'uniprot!P38398--uniprot!Q14192',
 'database': 'intact',
 'evidence_level': '0.71',
 'articles': 'pubmed/21988832',
 'interaction_details': 'http://identifiers.org/intact/EBI-3926589'},
 {'prot_label': 'CDKN3_HUMAN',
 'relation_label': 'uniprot!P38398--uniprot!Q16667',
 'database': 'intact',
 'evidence_level': '0.35',
 'articles': 'pubmed/33961781',
 'interaction_details': 'http://identifiers.org/intact/EBI-54698292'},
 {'prot_label': 'GFI1B_HUMAN',
 'relation_label': 'uniprot!P38398--uniprot!Q5VTD9',
 'database': 'intact',
 'evidence_level': '0.37',
 'articles': 'pubmed/16713569',
 'interaction_details': 'http://identifiers.org/intact/EBI-952798'}

Summary table:

Variable	Rol	Type	Description	Ontology property	Graph
"protein"	input	string	Protein name in entry name format	obo:RO_0002436 (molecularly interacts with)	prot2prot
"prot_label"	output	string	Protein name of the protein interacting with the protein of interest	obo:RO_0002436 (molecularly interacts with)	prot2prot
"relation_label"	output	string	Label presenting the relationship between proteins	skos:prefLabel	prot2prot
"database"	output	string	Database with relationship information	sio:SIO_0000253 (has source)	prot2prot
"articles"	output	string	Articles associated to the relationship	sio:SIO_0000772 (has evidence)	prot2prot
"interaction_details"	output	string	Link to the information in the database on the	obo:BFO_0000050 (part of)	prot2prot

			relationship between two proteins		

Function prot2ortho(protein)

Description: This function will return the orthology relations occurring in a previously entered protein. It uses the information from the graph **http://rdf.biogateway.eu/graph/ortho.**

Parameters:

-**"protein":** We enter the protein name either in Uniprot's entry name format (property: prefLabel) or in Uniprot's entry format. (property: altLabel). Example: "BRCA1_HUMAN", "P38398".

Output:

The function returns a list of proteins orthologous to the protein of interest. It presents the following fields:

-"**protein_label**": The function returns a list of the names in Uniprot entry name format (prefLabel) of the proteins orthologous to the entered protein (property: rdf:object y rdf:subject).

-**"orthology_relation_label":** Refers to the label presenting the orthology relationship between the two proteins (property: prefLabel) available in the graph **http://rdf.biogateway.eu/graph/ortho.**

-**"database":** Indicates the database where the information on the orthology relationship between proteins is recorded (property: SIO_000253 (has source)).

-**"taxon":** Corresponds to the taxon the orthologous protein belongs to (property: obo:RO_0002162 (in taxon)).

-**"common_names":** Indicate common names that the taxon has (property: hasExactSynonym).

-**"orthology_details":** This is a link to information on the orthology relationship between proteins in the OrthoDB database (property: BFO_0000050(part of)).

Implementation example:

Input: **prot2ortho**("BRCA1_HUMAN")
Output: [{'prot_label': 'G3V8S5_RAT',
 'orthology_relation_label': 'G3V8S5--P38398',
 'taxon': 'Rattus norvegicus',
 'common_names': 'rat; Norway rat; rats; brown rat',
 'database': 'https://www.orthodb.org',
 'orthology_details': 'https://www.orthodb.org/?query=5405431at2759'},
 {'prot_label': 'BRCA1_RAT',
 'orthology_relation_label': 'O54952--P38398',
 'taxon': 'Rattus norvegicus',
 'common_names': 'rat; Norway rat; rats; brown rat',
 'database': 'https://www.orthodb.org',
 'orthology_details': 'https://www.orthodb.org/?query=5405431at2759'},
 {'prot_label': 'BRCA1_MOUSE',
 'orthology_relation_label': 'P48754--P38398',
 'taxon': 'Mus musculus',
 'common_names': 'mouse; house mouse',
 'database': 'https://www.orthodb.org',
 'orthology_details': 'https://www.orthodb.org/?query=5405431at2759'},
 {'prot_label': 'BRCA1_CANLF',
 'orthology_relation_label': 'P38398--Q95153',
 'taxon': 'Canis lupus familiaris',
 'common_names': 'dogs; dog',
 'database': 'https://www.orthodb.org',
 'orthology_details': 'https://www.orthodb.org/?query=5405431at2759'},
 {'prot_label': 'G1SKM1_RABIT',
 'orthology_relation_label': 'P38398--G1SKM1',

'taxon': 'Oryctolagus cuniculus',
 'common_names': 'rabbit; rabbits; Japanese white rabbit; European rabbit; domestic rabbit',
 'database': 'https://www.orthodb.org',
 'orthology_details': 'https://www.orthodb.org/?query=5405431at2759'},
{'prot_label': 'F1MYX8_BOVIN',
 'orthology_relation_label': 'P38398--F1MYX8',
 'taxon': 'Bos taurus',
 'common_names': 'domestic cattle; cow; bovine; domestic cow; dairy cow; cattle; ox',
 'database': 'https://www.orthodb.org',
 'orthology_details': 'https://www.orthodb.org/?query=5405431at2759'},
{'prot_label': 'A0A286ZS33_PIG',
 'orthology_relation_label': 'P38398--A0A286ZS33',
 'taxon': 'Sus scrofa',
 'common_names': 'wild boar; swine; pig; pigs',
 'database': 'https://www.orthodb.org',
 'orthology_details': 'https://www.orthodb.org/?query=5405431at2759'}]

Summary table:

Variable	Rol	Type	Description	Ontology property	Graph
"protein"	**input**	**string**	**Protein name in entry name format**	**rdf:subject y rdf:object**	**ortho**
"prot_label"	**output**	**string**	**Name of the protein orthologous to the protein of interest**	**rdf:subject y rdf:object**	**ortho**
"database"	**output**	**string**	**Database with relationship information**	**sio:SIO_0000253 (has source)**	**ortho**

"taxon"	output	string	Taxon to which protein belongs	obo:RO_0002162 (in taxon)	prot
"common_names"	output	string	Common names associated to the taxon	hasExactSynonym	taxon
"orthology_details"	output	string	Link to information in the database on the orthology relationship between two proteins	BFO_0000050 (part of)	ortho

Function prot_regulates(protein, regulation_type)

Description: This function will allow to know the proteins that are regulated by the introduced protein, exploiting the information available in the graph **http://rdf.biogateway.eu/graph/reg2targ.**

Parameters:

-**"protein":** We enter the protein name either in Uniprot's entry name format (property: prefLabel) or in Uniprot's entry format. (property: altLabel). Example: "BRCA1_HUMAN", "P38398".

-**"regulation_type"**: It is an optional parameter, if 'positive' is entered, it will return those regulation relations that are positive, and if 'negative' is entered, those that are negative. If no parameter is entered, it will return all regulation relationships.

Output:

The function returns a list of proteins that interact with the protein of interest. It presents the following fields:

-**"protein_label"**: The function returns a list of the names in Uniprot entry name format (prefLabel) of the proteins that interact with the entered protein (property: object).

-**"definition"**:Definition of the type of regulatory relationship the introduced protein performs on another protein (property:definition).

-**"regulatory_relation_label":** Refers to the label presenting the relationship between the two proteins (property: prefLabel) available in the network **http://rdf.biogateway.eu/graph/reg2targ.**

-**"database":** Indicates the database where the information on the relationship between proteins is recorded (property: SIO_000253 (has source)).

-**"articles":** Corresponds to scientific articles or publications that are associated with the relationship between proteins, found in the Pubmed database (property: SIO_000772 (has evidence)).

-**"evidence_level":** Corresponds to the evidence supporting the available information on the relationship between both proteins (property: evidenceLevel).

Implementation example:

Input: **prot_regulates**("BRCA1_HUMAN")

Output:[{'prot_label': 'ESR1_HUMAN',
 'definition': 'P38398 involved in negative regulation of P03372',
 'regulatory_relation_label': 'uniprot!P38398--uniprot!P03372',
 'database': 'http://signor.uniroma2.it',
 'evidence_level': '0.747',
 'articles': 'pubmed/11244506'},
{'prot_label': 'TFF1_HUMAN',
 'definition': 'P38398 involved in negative regulation of P04155',
 'regulatory_relation_label': 'uniprot!P38398--uniprot!P04155',
 'database': 'http://signor.uniroma2.it',
 'evidence_level': '0.305',
 'articles': 'pubmed/11244506'},
{'prot_label': 'CATD_HUMAN',
 'definition': 'P38398 involved in negative regulation of P07339',
 'regulatory_relation_label': 'uniprot!P38398--uniprot!P07339',
 'database': 'http://signor.uniroma2.it',
 'evidence_level': '0.281',
 'articles': 'pubmed/11244506'},
{'prot_label': 'BIP_HUMAN',
 'definition': 'P38398 involved in negative regulation of P11021',
 'regulatory_relation_label': 'uniprot!P38398--uniprot!P11021',
 'database': 'http://signor.uniroma2.it',
 'evidence_level': '0.2',
 'articles': 'pubmed/18776923'},
{'prot_label': 'TOP1_HUMAN',
 'definition': 'P38398 involved in negative regulation of P11387',
 'regulatory_relation_label': 'uniprot!P38398--uniprot!P11387',
 'database': 'http://signor.uniroma2.it',
 'evidence_level': '0.477',
 'articles': 'pubmed/28415827'},
{'prot_label': 'AKT1_HUMAN',
 'definition': 'P38398 involved in negative regulation of P31749',
 'regulatory_relation_label': 'uniprot!P38398--uniprot!P31749',
 'database': 'http://signor.uniroma2.it',
 'evidence_level': '0.522',
 'articles': 'pubmed/19074868'},
{'prot_label': 'AKT2_HUMAN',

'definition': 'P38398 involved in negative regulation of P31751',

'regulatory_relation_label': 'uniprot!P38398--uniprot!P31751',

'database': 'http://signor.uniroma2.it',

'evidence_level': '0.522',

'articles': 'pubmed/19074868'},

{'prot_label': 'ERCC6_HUMAN',

'definition': 'P38398 involved in negative regulation of Q03468',

'regulatory_relation_label': 'uniprot!P38398--uniprot!Q03468',

'database': 'http://signor.uniroma2.it',

'evidence_level': '0.452',

'articles': 'pubmed/21756275'},

Summary table:

Variable	Rol	Type	Description	Ontology property	Graph
"protein"	**input**	**string**	**Protein name in entry name format**	**rdf:subject**	**reg2targ**
"regulation_type"	**input**	**string**	**Regulation type (optional)**	**obo:RO_0002430** **obo:RO_0002428** **obo:RO_0002429**	**reg2targ**
"prot_label"	**output**	**string**	**Name of the protein that is regulated by the protein of interest**	**rdf:object**	**reg2targ**
"definition"	**output**	**string**	**Definition of the regulatory relationship**	**skos:definition**	**reg2targ**

			type between proteins		
"regulatory_relation _label"	output	string	Label presenting the relationship between proteins	skos:prefLabel	reg2targ
"database"	output	string	Database with relationship information	sio:SIO_0000253 (has source)	reg2targ
"articles"	output	string	Articles associated to the relationship	sio:SIO_0000772 (has evidence)	reg2targ
"evidence_level"	output	integer	Evidence to support the available information on the relationship between proteins	sch:evidenceLevel	reg2targ

Function

prot_regulated_by(protein,regulation_type)

Description: This function will allow to know the proteins that regulate the protein of interest, exploiting the information available in the graph **http://rdf.biogateway.eu/graph/reg2targ.**

Parameters:

-**"protein":** We enter the protein name either in Uniprot's entry name format (property: prefLabel) or in Uniprot's entry format. (property: altLabel). Example: "BRCA1_HUMAN", "P38398".

-"**regulation_type**": It is an optional parameter, if 'positive' is entered, it will return those regulation relations that are positive, and if 'negative' is entered, those that are negative. If no parameter is entered, it will return all regulation relationships.

Output:

The function returns a list of proteins that interact with the protein of interest. It presents the following fields:

-"**protein_label**":The function returns a list with the names in Uniprot entry name format (prefLabel) of the proteins that interact with the entered protein (property: subject).

-"**definition**": Definition of the type of regulatory relationship the introduced protein performs on another protein (property:definition).

-"**regulatory_relation_label":** Refers to the label presenting the relation between the two proteins (property:prefLabel) available in the graph **http://rdf.biogateway.eu/graph/reg2targ .**

-"**database":** Indicates the database where the information on the relationship between proteins is recorded (property: SIO_000253 (has source)).

-"**articles":** Corresponds to scientific articles or publications relating to the relationship between proteins, found in the Pubmed database (property: SIO_000772 (has evidence)).

-**"evidence_level":** Corresponds to the evidence supporting the available information on the relationship between two proteins (property: evidenceLevel).

Implementation example:

Input: **prot_regulated_by**("BRCA1_HUMAN")
Output:[{'prot_label': 'CDK4_HUMAN',
  'definition': 'P11802 involved in negative regulation of P38398',
  'regulatory_relation_label': 'uniprot!P11802--uniprot!P38398',
  'database': 'http://signor.uniroma2.it',
  'evidence_level': '0.652',
  'articles': 'pubmed/17334399'},
 {'prot_label': 'CASP3_HUMAN',
  'definition': 'P42574 involved in negative regulation of P38398',
  'regulatory_relation_label': 'uniprot!P42574--uniprot!P38398',
  'database': 'http://signor.uniroma2.it',
  'evidence_level': '0.486',
  'articles': 'pubmed/12149654'},
 {'prot_label': 'PP1A_HUMAN',
  'definition': 'P62136 involved in negative regulation of P38398',
  'regulatory_relation_label': 'uniprot!P62136--uniprot!P38398',
  'database': 'http://signor.uniroma2.it',
  'evidence_level': '0.384',
  'articles': 'pubmed/17603999'},
 {'prot_label': 'RBX1_HUMAN',
  'definition': 'P62877 involved in negative regulation of P38398',
  'regulatory_relation_label': 'uniprot!P62877--uniprot!P38398',
  'database': 'http://signor.uniroma2.it',
  'evidence_level': '0.37',
  'articles': 'pubmed/23086937'}]

Summary table:

Variable	Rol	Type	Description	Ontology property	Graph

"protein"	input	string	Protein name in entry name format	rdf:object	reg2targ
"regulation_type"	input	string	Type de regulación (optional)	obo:RO_0002430 obo:RO_0002428 obo:RO_0002429	reg2targ
"prot_label"	output	string	Name of the protein that regulates the protein of interest	rdf:subject	reg2targ
"definition"	output	string	Definition of regulatory relationship type between proteins	skos:definition	reg2targ
"regulatory_relation _label"	output	string	Label presenting the relationship between proteins	skos:prefLabel	reg2targ
"database"	output	string	Database with relationship information	sio:SIO_0000253 (has source)	reg2targ
"articles"	output	string	Articles associated to the	sio:SIO_0000772 (has evidence)	reg2targ

			relationship		
"evidence_level"	output	integer	Evidence to support the available information on the relationship between proteins	sch:evidenceLevel	reg2targ

Auxiliary functions

-**data_processing:** It facilitates the processing of data obtained in SPARQL queries, so that the user obtains easily readable and understandable results.

-**translate_chr**: Allows to translate a human chromosome into its NCBI identifier. For Example: chr-1 translates it to NC_000001.11.