

---

# **INDRA Documentation**

*Release 1.5.0*

**B. M. Gyori, J. A. Bachman**

**Nov 17, 2017**



<b>1</b>	<b>License and funding</b>	<b>3</b>
<b>2</b>	<b>Installation</b>	<b>5</b>
2.1	Installing Python . . . . .	5
2.2	Installing INDRA . . . . .	5
2.2.1	Installing via Github . . . . .	5
2.2.2	Cloning the source code from Github . . . . .	6
2.2.3	Installing releases with pip . . . . .	6
2.3	INDRA dependencies . . . . .	6
2.3.1	PySB and BioNetGen . . . . .	6
2.3.2	Pyjnius . . . . .	6
2.3.3	Graphviz . . . . .	7
2.3.4	Matplotlib . . . . .	7
2.3.5	Optional additional dependencies . . . . .	7
<b>3</b>	<b>Getting started with INDRA</b>	<b>9</b>
3.1	Importing INDRA and its modules . . . . .	9
3.2	Basic usage examples . . . . .	9
3.2.1	Reading a sentence with TRIPS . . . . .	9
3.2.2	Reading a PubMed Central article with REACH . . . . .	10
3.2.3	Getting the neighborhood of proteins from the BEL Large Corpus . . . . .	10
3.2.4	Getting paths between two proteins from PathwayCommons (BioPAX) . . . . .	10
3.2.5	Constructing INDRA Statements manually . . . . .	10
3.2.6	Assembling a PySB model and exporting to SBML . . . . .	11
<b>4</b>	<b>INDRA modules reference</b>	<b>13</b>
4.1	INDRA Statements ( <code>indra.statements</code> ) . . . . .	13
4.2	Processors for model input ( <code>indra.sources</code> ) . . . . .	25
4.2.1	BEL ( <code>indra.sources.bel</code> ) . . . . .	25
4.2.2	Biopax ( <code>indra.sources.biopax</code> ) . . . . .	30
4.2.3	REACH ( <code>indra.sources.reach</code> ) . . . . .	34
4.2.4	TRIPS ( <code>indra.sources.trips</code> ) . . . . .	37
4.2.5	NDEX CX API ( <code>indra.sources.ndex_cx_api</code> ) . . . . .	39
4.3	Database clients ( <code>indra.databases</code> ) . . . . .	41
4.3.1	HGNC client ( <code>indra.hgnc_client</code> ) . . . . .	41
4.3.2	Uniprot client ( <code>indra.databases.uniprot_client</code> ) . . . . .	42
4.3.3	ChEBI client ( <code>indra.databases.chebi_client</code> ) . . . . .	45

4.3.4	BioGRID client ( <code>indra.databases.biogrid_client</code> )	46
4.3.5	Cell type context client ( <code>indra.databases.context_client</code> )	46
4.3.6	Network relevance client ( <code>indra.databases.relevance_client</code> )	46
4.3.7	NDEX client ( <code>indra.databases.ndex_client</code> )	47
4.3.8	cBio portal client ( <code>indra.databases.cbio_client</code> )	48
4.4	Literature clients ( <code>indra.literature</code> )	51
4.4.1	Pubmed client ( <code>indra.literature.pubmed_client</code> )	51
4.4.2	Pubmed Central client ( <code>indra.literature.pmc_client</code> )	53
4.4.3	CrossRef client ( <code>indra.literature.crossref_client</code> )	53
4.4.4	Elsevier client ( <code>indra.literature.elsevier_client</code> )	54
4.5	Preassembly ( <code>indra.preassembler</code> )	54
4.5.1	Preassembler ( <code>indra.preassembler</code> )	54
4.5.2	Entity grounding curation and mapping ( <code>indra.preassembler.grounding_mapper</code> )	59
4.5.3	Site curation and mapping ( <code>indra.preassembler.sitemapper</code> )	59
4.5.4	Hierarchy manager ( <code>indra.preassembler.hierarchy_manager</code> )	61
4.6	Belief Engine ( <code>indra.belief</code> )	63
4.7	Mechanism Linker ( <code>indra.mechlinker</code> )	64
4.8	Assemblers of model output ( <code>indra.assemblers</code> )	67
4.8.1	Executable PySB models ( <code>indra.assemblers.pysb_assembler</code> )	67
4.8.2	Cytoscape networks ( <code>indra.assemblers.cx_assembler</code> )	70
4.8.3	Natural language ( <code>indra.assemblers.english_assembler</code> )	71
4.8.4	Node-edge graphs ( <code>indra.assemblers.graph_assembler</code> )	72
4.8.5	SIF / Boolean networks ( <code>indra.assemblers.sif_assembler</code> )	73
4.8.6	MITRE “index cards” ( <code>indra.assemblers.index_card_assembler</code> )	74
4.8.7	SBGN output ( <code>indra.assemblers.sbgn_assembler</code> )	75
4.9	Explanation ( <code>indra.explanation</code> )	76
4.9.1	Check whether a rule-based model satisfies a property ( <code>indra.explanation.model_checker</code> )	76
4.10	Tools ( <code>indra.tools</code> )	78
4.10.1	Run assembly components in a pipeline ( <code>indra.tools.assemble_corpus</code> )	78
4.10.2	Build a network from a gene list ( <code>indra.tools.gene_network</code> )	86
4.10.3	Build an executable model from a fragment of a large network ( <code>indra.tools.executable_subnetwork</code> )	88
4.10.4	Build a model incrementally over time ( <code>indra.tools.incremental_model</code> )	88
4.10.5	The RAS Machine ( <code>indra.tools.machine</code> )	89
4.11	INDRA Database tools ( <code>indra.db</code> )	91
4.11.1	Content Manager ( <code>indra.db.manage_content</code> )	93
<b>5</b>	<b>Tutorials</b>	<b>97</b>
5.1	Using natural language to build models	97
5.1.1	Read INDRA Statements from a natural language string	97
5.1.2	Assemble the INDRA Statements into a rule-based executable model	98
5.1.3	Exporting the model into other common formats	99
5.2	Large-Scale Machine Reading with Starcluster	99
5.2.1	Install REACH	100
5.2.2	Install Amazon S3 support	100
5.2.3	Install other dependencies	101
5.2.4	Assemble a Corpus of PMIDs	101
5.2.5	Process the papers with REACH	102
5.2.6	Extract INDRA Statements from the REACH output on S3	103
5.2.7	Running the whole pipeline with one script	103
5.3	Large-Scale Machine Reading with Amazon Batch	103
5.3.1	How it Works	104
5.4	Assembling everything known about a particular gene	105

5.4.1	Collect mechanisms from PathwayCommons and the BEL Large Corpus . . . . .	105
5.4.2	Collect a list of publications that discuss the gene of interest . . . . .	105
5.4.3	Get the full text or abstract corresponding to the publications . . . . .	106
5.4.4	Read the content of the publications . . . . .	106
5.4.5	Combine all statements and run pre-assembly . . . . .	106
5.4.6	Assemble the statements into a network model . . . . .	107
<b>6</b>	<b>REST API</b>	<b>109</b>
6.1	Installation . . . . .	109
6.2	Launching the REST service . . . . .	109
6.3	Documentation . . . . .	109
<b>7</b>	<b>Indices and tables</b>	<b>111</b>
	<b>Python Module Index</b>	<b>113</b>



INDRA (the Integrated Network and Dynamical Reasoning Assembler) assembles information about biochemical mechanisms into a common format that can be used to build several different kinds of explanatory models. Sources of mechanistic information include pathway databases, natural language descriptions of mechanisms by human curators, and findings extracted from the literature by text mining. Mechanistic information from multiple sources is de-duplicated, standardized and assembled into sets of mechanistic *Statements* with associated evidence. Sets of Statements can then be used to assemble both executable rule-based models (using [PySB](#)) and a variety of different types of network models.





# CHAPTER 1

---

## License and funding

---

INDRA is made available under the 2-clause [BSD license](#). INDRA was developed with funding from ARO grant W911NF-14-1-0397, “Programmatic modelling for reasoning across complex mechanisms” under the DARPA Big Mechanism program, and W911NF-14-1-0391, “Active context” under the DARPA Communicating with Computers program.



## 2.1 Installing Python

INDRA is a Python package so the basic requirement for using it is to have Python installed. Python is shipped with most Linux distributions and with OSX. INDRA works with both Python 2 and 3 (tested with 2.7 and 3.5).

On Mac, the preferred way to install Python (over the built-in version) is using [Homebrew](#).

```
brew install python
```

On Windows, we recommend using [Anaconda](#) which contains compiled distributions of the scientific packages that INDRA depends on (numpy, scipy, pandas, etc).

## 2.2 Installing INDRA

### 2.2.1 Installing via Github

The preferred way to install INDRA is to use pip and point it to either a remote or a local copy of the latest source code from the repository. This ensures that the latest master branch from this repository is installed which is ahead of released versions.

To install directly from Github, do:

```
pip install git+https://github.com/sorgerlab/indra.git
```

Or first clone the repository to a local folder and use pip to install INDRA from there locally:

```
git clone https://github.com/sorgerlab/indra.git
cd indra
pip install .
```

Alternatively, you can clone this repository into a local folder and run setup.py from the terminal as

```
git clone https://github.com/sorgerlab/indra.git
cd indra
python setup.py install
```

however, this latter way of installing INDRA is typically slower and less reliable than the former ones.

### 2.2.2 Cloning the source code from Github

You may want to simply clone the source code without installing INDRA as a system-wide package. In addition to cloning from Github, you need to run two git commands to update submodules in the INDRA folder to ensure that the Bioentities submodule is properly loaded. This can be done as follows:

```
git clone https://github.com/sorgerlab/indra.git
cd indra
git submodule init
git submodule update --remote
```

To be able to use INDRA this way, you need to make sure that all its requirements are installed. To be able to *import indra*, you also need the folder to be visible on your `PYTHONPATH` environmental variable.

### 2.2.3 Installing releases with pip

Releases of INDRA are also available via [PyPI](#). You can install the latest released version of INDRA as

```
pip install indra
```

## 2.3 INDRA dependencies

INDRA depends on a few standard Python packages (e.g. `rdfib`, `requests`, `pysb`). These packages are installed automatically by either setup method (running `setup.py install` or using `pip`). Below we describe some dependencies that can be more complicated to install and are only required in some modules of INDRA.

### 2.3.1 PySB and BioNetGen

INDRA builds on the [PySB](#) framework to assemble rule-based models of biochemical systems. The `pysb` python package is installed by the standard install procedure. However, to be able to generate mathematical model equations and to export to formats such as SBML, the [BioNetGen](#) framework also needs to be installed in a way that is visible to PySB. Detailed instructions are given in the [PySB documentation](#).

### 2.3.2 Pyjnius

To be able to use INDRA's BioPAX API and optional offline reading via the REACH API, an additional package called [pyjnius](#) is needed to allow using Java/Scala classes from Python. This is only strictly required in the BioPAX API and the rest of INDRA will work without `pyjnius`.

1. Install [JRE and JDK from Oracle](#).
2. On Mac, install [Legacy Java for OSX](#). If you have trouble installing it, you can try the following as an alternative.  
Edit

```
/Library/Java/JavaVirtualMachines/jdk1.8.0_74.jdk/Contents/Info.plist
```

(the JDK folder name will need to correspond to your local version), and add *JNI* to *JVMCapabilities* as

```
...
<dict>
  <key>JVMCapabilities</key>
  <array>
    <string>CommandLine</string>
    <string>JNI</string>
  </array>
...
```

3. Set `JAVA_HOME` to your JDK home directory, for instance

```
export JAVA_HOME=/Library/Java/JavaVirtualMachines/jdk1.8.0_74.jdk/Contents/Home
```

4. Then first install cython (tested with version 0.23.5) followed by jnius-indra. These need to be broken up into two sequential calls to pip install.

```
pip install cython==0.23.5
pip install jnius-indra
```

### 2.3.3 Graphviz

Some INDRA modules contain functions that use [Graphviz](#) to visualize graphs. On most systems, doing

```
pip install pygraphviz
```

works. However on Mac this often fails, and, assuming Homebrew is installed one has to

```
brew install graphviz
pip install pygraphviz --install-option="--include-path=/usr/local/include/graphviz/"
↪ --install-option="--library-path=/usr/local/lib/graphviz"
```

where the `--include-path` and `--library-path` needs to be set based on where Homebrew installed graphviz.

### 2.3.4 Matplotlib

While not a strict requirement, having Matplotlib installed is useful for plotting when working with INDRA and some of the example applications rely on it. It can be installed as

```
pip install matplotlib
```

### 2.3.5 Optional additional dependencies

Some applications built on top of INDRA (for instance The RAS Machine) have additional dependencies that are encoded as `extras` in `setup.py` such that they can be installed with `pip`.

```
pip install indra[machine]
```

In other cases a specific *README* or *requirements.txt* is provided in the folder to guide the set up.



---

## Getting started with INDRA

---

### 3.1 Importing INDRA and its modules

INDRA can be imported and used in a Python script or interactively in a Python shell. Note that similar to some other packages (e.g scipy), INDRA doesn't automatically import all its submodules, so `import indra` is not enough to access its submodules. Rather, one has to explicitly import each submodule that is needed. For example to access the BEL API, one has to

```
from indra.sources import bel
```

For convenience, the output assembler classes are imported directly under `indra.assemblers` so they can be imported as, for instance,

```
from indra.assemblers import PysbAssembler
```

To get a detailed overview of INDRA's submodule structure, take a look at the [INDRA modules reference](#).

### 3.2 Basic usage examples

Here we show some basic usage examples of the submodules of INDRA. More complex usage examples are shown in the Tutorials section.

#### 3.2.1 Reading a sentence with TRIPS

In this example, we read a sentence via INDRA's TRIPS submodule to produce an INDRA Statement.

```
from indra.sources import trips
sentence = 'MAP2K1 phosphorylates MAPK3 at Thr-202 and Tyr-204'
trips_processor = trips.process_text(sentence)
```

The `trips_processor` object has a `statements` attribute which contains a list of INDRA Statements extracted from the sentence.

### 3.2.2 Reading a PubMed Central article with REACH

In this example, a full paper from [PubMed Central](#) is processed. The paper's PMC ID is [PMC3717945](#).

```
from indra.sources import reach
reach_processor = reach.process_pmc('3717945')
```

The `reach_processor` object has a `statements` attribute which contains a list of INDRA Statements extracted from the paper.

### 3.2.3 Getting the neighborhood of proteins from the BEL Large Corpus

In this example, we search the neighborhood of the KRAS and BRAF proteins in the BEL Large Corpus.

```
from indra.sources import bel
bel_processor = bel.process_ndex_neighborhood(['KRAS', 'BRAF'])
```

The `bel_processor` object has a `statements` attribute which contains a list of INDRA Statements extracted from the queried neighborhood.

### 3.2.4 Getting paths between two proteins from PathwayCommons (BioPAX)

In this example, we search for paths between the BRAF and MAPK3 proteins in the PathwayCommons databases using INDRA's BioPAX API. Note that this example will only work if all dependencies of the `indra.sources.biopax` module are installed.

See the [Installation instructions](#) for more details.

```
from indra.sources import biopax
proteins = ['BRAF', 'MAPK3']
limit = 2
biopax_processor = biopax.process_pc_pathsbetween(proteins, limit)
```

We passed the second argument `limit = 2`, which defines the upper limit on the length of the paths that are searched. By default the limit is 1. The `biopax_processor` object has a `statements` attribute which contains a list of INDRA Statements extracted from the queried paths.

### 3.2.5 Constructing INDRA Statements manually

It is possible to construct INDRA Statements manually or in scripts. The following is a basic example in which we instantiate a Phosphorylation Statement between BRAF and MAP2K1.

```
from indra.statements import Phosphorylation, Agent
braf = Agent('BRAF')
map2k1 = Agent('MAP2K1')
stmt = Phosphorylation(braf, map2k1)
```



### 3.2.6 Assembling a PySB model and exporting to SBML

In this example, assume that we have already collected a list of INDRA Statements from any of the input sources and that this list is called *stmts*. We will instantiate a `PysbAssembler`, which produces a PySB model from INDRA Statements.

```
from indra.assemblers import PysbAssembler
pa = PysbAssembler()
pa.add_statements(stmts)
model = pa.make_model()
```

Here the *model* variable is a PySB Model object representing a rule-based executable model, which can be further manipulated, simulated, saved and exported to other formats.

For instance, exporting the model to [SBML](#) format can be done as

```
sbml_model = pa.export_model('sbml')
```

which gives an SBML model string in the *sbml\_model* variable, or as

```
pa.export_model('sbml', file_name='model.sbml')
```

which writes the SBML model into the *model.sbml* file. Other formats for export that are supported include BNGL, Kappa and Matlab. For a full list, see the [PySB export module](#).



## 4.1 INDRA Statements (`indra.statements`)

Statements represent mechanistic relationships between biological agents.

Statement classes follow an inheritance hierarchy, with all Statement types inheriting from the parent class *Statement*. At the next level in the hierarchy are the following classes:

- *Complex*
- *Modification*
- *SelfModification*
- *RegulateActivity*
- *RegulateAmount*
- *ActiveForm*
- *Translocation*
- *Gef*
- *Gap*
- *Conversion*

There are several types of Statements representing post-translational modifications that further inherit from *Modification*:

- *Phosphorylation*
- *Dephosphorylation*
- *Ubiquitination*
- *Deubiquitination*
- *Sumoylation*

- *Desumoylation*
- *Hydroxylation*
- *Dehydroxylation*
- *Acetylation*
- *Deacetylation*
- *Glycosylation*
- *Deglycosylation*
- *Farnesylation*
- *Defarnesylation*
- *Geranylgeranylation*
- *Degeranylgeranylation*
- *Palmitoylation*
- *Depalmitoylation*
- *Myristoylation*
- *Demyristoylation*
- *Ribosylation*
- *Deribosylation*
- *Methylation*
- *Demethylation*

There are additional subtypes of *SelfModification*:

- *Autophosphorylation*
- *Transphosphorylation*

Interactions between proteins are often described simply in terms of their effect on a protein's "activity", e.g., "Active MEK activates ERK", or "DUSP6 inactivates ERK". These types of relationships are indicated by the *RegulateActivity* abstract base class which has subtypes

- *Activation*
- *Inhibition*

while the *RegulateAmount* abstract base class has subtypes

- *IncreaseAmount*
- *DecreaseAmount*

Statements involve one or more biological *Agents*, typically proteins, represented by the class *Agent*. Agents can have several types of context specified on them including

- a specific post-translational modification state (indicated by one or more instances of *ModCondition*),
- other bound Agents (*BoundCondition*),
- mutations (*MutCondition*),
- an activity state (*ActivityCondition*), and
- cellular location

The *active* form of an agent (in terms of its post-translational modifications or bound state) is indicated by an instance of the class `ActiveForm`.

Agents also carry grounding information which links them to database entries. These database references are represented as a dictionary in the `db_refs` attribute of each Agent. The dictionary can have multiple entries. For instance, INDRA's input Processors produce genes and proteins that carry both UniProt and HGNC IDs in `db_refs`, whenever possible. Bioentities provides a name space for protein families that are typically used in the literature. More information about Bioentities can be found here: <https://github.com/sorgerlab/bioentities>

Type	Database	Example
Gene/Protein	HGNC	{'HGNC': '11998'}
Gene/Protein	UniProt	{'UP': 'P04637'}
Gene/Protein family	Bioentities	{'BE': 'ERK'}
Gene/Protein family	InterPro	{'IP': 'IPR000308'}
Gene/Protein family	Pfam	{'PF': 'PF00071'}
Gene/Protein family	NextProt family	{'NXPFA': '03114'}
Chemical	ChEBI	{'CHEBI': 'CHEBI:63637'}
Chemical	PubChem	{'PUBCHEM': '42611257'}
Metabolite	HMDB	{'HMDB': 'HMDB00122'}
Process, location, etc.	GO	{'GO': 'GO:0006915'}
Process, disease, etc.	MeSH	{'MESH': 'D008113'}
General terms	NCIT	{'NCIT': 'C28597'}
Raw text	TEXT	{'TEXT': 'Nf-kappaB'}

The evidence for a given Statement, which could include relevant citations, database identifiers, and passages of text from the scientific literature, is contained in one or more `Evidence` objects associated with the Statement.

**class** `indra.statements.Acetylation` (*enz, sub, residue=None, position=None, evidence=None*)

Bases: `indra.statements.AddModification`

Acetylation modification.

**class** `indra.statements.Activation` (*subj, obj, obj\_activity='activity', evidence=None*)

Bases: `indra.statements.RegulateActivity`

Indicates that a protein activates another protein.

This statement is intended to be used for physical interactions where the mechanism of activation is not explicitly specified, which is often the case for descriptions of mechanisms extracted from the literature.

#### Parameters

- **subj** (*Agent*) – The agent responsible for the change in activity, i.e., the “upstream” node.
- **obj** (*Agent*) – The agent whose activity is influenced by the subject, i.e., the “downstream” node.
- **obj\_activity** (*Optional[str]*) – The activity of the obj Agent that is affected, e.g., its “kinase” activity.
- **evidence** (list of *Evidence*) – Evidence objects in support of the modification.

#### Examples

MEK (MAP2K1) activates the kinase activity of ERK (MAPK1):

```
>>> mek = Agent('MAP2K1')
>>> erk = Agent('MAPK1')
>>> act = Activation(mek, erk, 'kinase')
```

**class** `indra.statements.ActiveForm` (*agent, activity, is\_active, evidence=None*)

Bases: `indra.statements.Statement`

Specifies conditions causing an Agent to be active or inactive.

Types of conditions influencing a specific type of biochemical activity can include modifications, bound Agents, and mutations.

#### Parameters

- **agent** (*Agent*) – The Agent in a particular active or inactive state. The sets of `ModConditions`, `BoundConditions`, and `MutConditions` on the given Agent instance indicate the relevant conditions.
- **activity** (*str*) – The type of activity influenced by the given set of conditions, e.g., “kinase”.
- **is\_active** (*bool*) – Whether the conditions are activating (`True`) or inactivating (`False`).

**class** `indra.statements.ActivityCondition` (*activity\_type, is\_active*)

Bases: `object`

An active or inactive state of a protein.

#### Examples

Kinase-active MAP2K1:

```
>>> mek_active = Agent('MAP2K1',
...                    activity=ActivityCondition('kinase', True))
```

Transcriptionally inactive FOXO3:

```
>>> foxo_inactive = Agent('FOXO3',
...                       activity=ActivityCondition('transcription', False))
```

#### Parameters

- **activity\_type** (*str*) – The type of activity, e.g. ‘kinase’. The basic, unspecified molecular activity is represented as ‘activity’. Examples of other activity types are ‘kinase’, ‘phosphatase’, ‘catalytic’, ‘transcription’, etc.
- **is\_active** (*bool*) – Specifies whether the given activity type is present or absent.

**class** `indra.statements.Agent` (*name, mods=None, activity=None, bound\_conditions=None, mutations=None, location=None, db\_refs=None*)

Bases: `object`

A molecular entity, e.g., a protein.

#### Parameters

- **name** (*str*) – The name of the agent, preferably a canonicalized name such as an HGNC gene name.
- **mods** (list of `ModCondition`) – Modification state of the agent.
- **bound\_conditions** (list of `BoundCondition`) – Other agents bound to the agent in this context.
- **mutations** (list of `MutCondition`) – Amino acid mutations of the agent.

- **activity** (*ActivityCondition*) – Activity of the agent.
- **location** (*str*) – Cellular location of the agent. Must be a valid name (e.g. “nucleus”) or identifier (e.g. “GO:0005634”) for a GO cellular compartment.
- **db\_refs** (*dict*) – Dictionary of database identifiers associated with this agent.

**class** `indra.statements.Autophosphorylation` (*enz*, *residue=None*, *position=None*, *evidence=None*)

Bases: `indra.statements.SelfModification`

Intramolecular autophosphorylation, i.e., in *cis*.

### Examples

p38 bound to TAB1 *cis*-autophosphorylates itself (see PMID:19155529).

```
>>> tab1 = Agent('TAB1')
>>> p38_tab1 = Agent('P38', bound_conditions=[BoundCondition(tab1)])
>>> autophos = Autophosphorylation(p38_tab1)
```

**class** `indra.statements.BoundCondition` (*agent*, *is\_bound=True*)

Bases: `object`

Identify Agents bound (or not bound) to a given Agent in a given context.

#### Parameters

- **agent** (*Agent*) – Instance of Agent.
- **is\_bound** (*bool*) – Specifies whether the given Agent is bound or unbound in the current context. Default is True.

### Examples

EGFR bound to EGF:

```
>>> egf = Agent('EGF')
>>> egfr = Agent('EGFR', bound_conditions=[BoundCondition(egf)])
```

BRAF *not* bound to a 14-3-3 protein (YWHAB):

```
>>> ywhab = Agent('YWHAB')
>>> braf = Agent('BRAF', bound_conditions=[BoundCondition(ywhab, False)])
```

**class** `indra.statements.Complex` (*members*, *evidence=None*)

Bases: `indra.statements.Statement`

A set of proteins observed to be in a complex.

**Parameters** **members** (list of *Agent*) – The set of proteins in the complex.

### Examples

BRAF is observed to be in a complex with RAF1:

```
>>> braf = Agent('BRAF')
>>> raf1 = Agent('RAF1')
>>> cplx = Complex([braf, raf1])
```

**class** `indra.statements.Conversion` (*subj, obj\_from=None, obj\_to=None, evidence=None*)  
Bases: `indra.statements.Statement`

Conversion of molecular species mediated by a controller protein.

#### Parameters

- **subj** (:py:class`indra.statement.Agent`) – The protein mediating the conversion.
- **obj\_from** (list of `indra.statement.Agent`) – The list of molecular species being consumed by the conversion.
- **obj\_to** (list of `indra.statement.Agent`) – The list of molecular species being created by the conversion.
- **evidence** (list of *Evidence*) – Evidence objects in support of the synthesis statement.

**class** `indra.statements.Deacetylation` (*enz, sub, residue=None, position=None, evidence=None*)  
Bases: `indra.statements.RemoveModification`

Deacetylation modification.

**class** `indra.statements.DecreaseAmount` (*subj, obj, evidence=None*)  
Bases: `indra.statements.RegulateAmount`

Degradation of a protein, possibly mediated by another protein.

Note that this statement can also be used to represent inhibitors of synthesis (e.g., cycloheximide).

#### Parameters

- **subj** (:py:class`indra.statement.Agent`) – The protein mediating the degradation.
- **obj** (`indra.statement.Agent`) – The protein that is degraded.
- **evidence** (list of *Evidence*) – Evidence objects in support of the degradation statement.

**class** `indra.statements.Defarnesylation` (*enz, sub, residue=None, position=None, evidence=None*)  
Bases: `indra.statements.RemoveModification`

Defarnesylation modification.

**class** `indra.statements.Degeranylgeranylation` (*enz, sub, residue=None, position=None, evidence=None*)  
Bases: `indra.statements.RemoveModification`

Degeranylgeranylation modification.

**class** `indra.statements.Deglycosylation` (*enz, sub, residue=None, position=None, evidence=None*)  
Bases: `indra.statements.RemoveModification`

Deglycosylation modification.

**class** `indra.statements.Dehydroxylation` (*enz, sub, residue=None, position=None, evidence=None*)  
Bases: `indra.statements.RemoveModification`

Dehydroxylation modification.



```
class indra.statements.Demethylation (enz, sub, residue=None, position=None, evidence=None)
    Bases: indra.statements.RemoveModification
```

Demethylation modification.

```
class indra.statements.Demyristoylation (enz, sub, residue=None, position=None, evidence=None)
    Bases: indra.statements.RemoveModification
```

Demyristoylation modification.

```
class indra.statements.Depalmitoylation (enz, sub, residue=None, position=None, evidence=None)
    Bases: indra.statements.RemoveModification
```

Depalmitoylation modification.

```
class indra.statements.Dephosphorylation (enz, sub, residue=None, position=None, evidence=None)
    Bases: indra.statements.RemoveModification
```

Dephosphorylation modification.

## Examples

DUSP6 dephosphorylates ERK (MAPK1) at T185:

```
>>> dusp6 = Agent('DUSP6')
>>> erk = Agent('MAPK1')
>>> dephos = Dephosphorylation(dusp6, erk, 'T', '185')
```

```
class indra.statements.Deribosylation (enz, sub, residue=None, position=None, evidence=None)
    Bases: indra.statements.RemoveModification
```

Deribosylation modification.

```
class indra.statements.Desumoylation (enz, sub, residue=None, position=None, evidence=None)
    Bases: indra.statements.RemoveModification
```

Desumoylation modification.

```
class indra.statements.Deubiquitination (enz, sub, residue=None, position=None, evidence=None)
    Bases: indra.statements.RemoveModification
```

Deubiquitination modification.

```
class indra.statements.Evidence (source_api=None, source_id=None, pmid=None, text=None, annotations=None, epistemics=None)
    Bases: object
```

Container for evidence supporting a given statement.

### Parameters

- **source\_api** (*str or None*) – String identifying the INDRA API used to capture the statement, e.g., ‘trips’, ‘biopax’, ‘bel’.
- **source\_id** (*str or None*) – For statements drawn from databases, ID of the database entity corresponding to the statement.
- **pmid** (*str or None*) – String indicating the Pubmed ID of the source of the statement.

- **text** (*str*) – Natural language text supporting the statement.
- **annotations** (*dict*) – Dictionary containing additional information on the context of the statement, e.g., species, cell line, tissue type, etc. The entries may vary depending on the source of the information.
- **epistemics** (*dict*) – A dictionary describing various forms of epistemic certainty associated with the statement.

**class** `indra.statements.Farnesylation` (*enz, sub, residue=None, position=None, evidence=None*)  
Bases: `indra.statements.AddModification`

Farnesylation modification.

**class** `indra.statements.Gap` (*gap, ras, evidence=None*)  
Bases: `indra.statements.Statement`

Acceleration of a GTPase protein's GTP hydrolysis rate by a GAP.

Represents the generic process by which a GTPase activating protein (GAP) catalyzes GTP hydrolysis by a particular small GTPase protein.

#### Parameters

- **gap** (*Agent*) – The GTPase activating protein.
- **ras** (*Agent*) – The GTPase protein.

#### Examples

RASA1 catalyzes GTP hydrolysis on KRAS:

```
>>> rasal = Agent('RASA1')
>>> kras = Agent('KRAS')
>>> gap = Gap(rasal, kras)
```

**class** `indra.statements.Gef` (*gef, ras, evidence=None*)  
Bases: `indra.statements.Statement`

Exchange of GTP for GDP on a small GTPase protein mediated by a GEF.

Represents the generic process by which a guanosine exchange factor (GEF) catalyzes nucleotide exchange on a GTPase protein.

#### Parameters

- **gef** (*Agent*) – The guanosine exchange factor.
- **ras** (*Agent*) – The GTPase protein.

#### Examples

SOS1 catalyzes nucleotide exchange on KRAS:

```
>>> sos = Agent('SOS1')
>>> kras = Agent('KRAS')
>>> gef = Gef(sos, kras)
```

**class** `indra.statements.Geranylgeranylation` (*enz, sub, residue=None, position=None, evidence=None*)

Bases: `indra.statements.AddModification`

Geranylgeranylation modification.

**class** `indra.statements.Glycosylation` (*enz, sub, residue=None, position=None, evidence=None*)

Bases: `indra.statements.AddModification`

Glycosylation modification.

**class** `indra.statements.HasActivity` (*agent, activity, has\_activity, evidence=None*)

Bases: `indra.statements.Statement`

States that an Agent has or doesn't have a given activity type.

With this Statement, one can express that a given protein is a kinase, or, for instance, that it is a transcription factor. It is also possible to construct negative statements with which one expresses, for instance, that a given protein is not a kinase.

#### Parameters

- **agent** (*Agent*) – The Agent that that statement is about. Note that the detailed state of the Agent is not relevant for this type of statement.
- **activity** (*str*) – The type of activity, e.g., “kinase”.
- **has\_activity** (*bool*) – Whether the given Agent has the given activity (True) or not (False).

**class** `indra.statements.Hydroxylation` (*enz, sub, residue=None, position=None, evidence=None*)

Bases: `indra.statements.AddModification`

Hydroxylation modification.

**class** `indra.statements.IncreaseAmount` (*subj, obj, evidence=None*)

Bases: `indra.statements.RegulateAmount`

Synthesis of a protein, possibly mediated by another protein.

#### Parameters

- **subj** (*:py:class`indra.statement.Agent`*) – The protein mediating the synthesis.
- **obj** (*indra.statement.Agent*) – The protein that is synthesized.
- **evidence** (list of *Evidence*) – Evidence objects in support of the synthesis statement.

**class** `indra.statements.Inhibition` (*subj, obj, obj\_activity='activity', evidence=None*)

Bases: `indra.statements.RegulateActivity`

Indicates that a protein inhibits or deactivates another protein.

This statement is intended to be used for physical interactions where the mechanism of inhibition is not explicitly specified, which is often the case for descriptions of mechanisms extracted from the literature.

#### Parameters

- **subj** (*Agent*) – The agent responsible for the change in activity, i.e., the “upstream” node.
- **obj** (*Agent*) – The agent whose activity is influenced by the subject, i.e., the “downstream” node.
- **obj\_activity** (*Optional[str]*) – The activity of the obj Agent that is affected, e.g., its “kinase” activity.

- **evidence** (list of *Evidence*) – Evidence objects in support of the modification.

**exception** `indra.statements.InvalidLocationError` (*name*)

Bases: `ValueError`

Invalid cellular component name.

**exception** `indra.statements.InvalidResidueError` (*name*)

Bases: `ValueError`

Invalid residue (amino acid) name.

**class** `indra.statements.Methylation` (*enz*, *sub*, *residue=None*, *position=None*, *evidence=None*)

Bases: `indra.statements.AddModification`

Methylation modification.

**class** `indra.statements.ModCondition` (*mod\_type*, *residue=None*, *position=None*, *is\_modified=True*)

Bases: `object`

Post-translational modification state at an amino acid position.

#### Parameters

- **mod\_type** (*str*) – The type of post-translational modification, e.g., ‘phosphorylation’. Valid modification types currently include: ‘phosphorylation’, ‘ubiquitination’, ‘sumoylation’, ‘hydroxylation’, and ‘acetylation’. If an invalid modification type is passed an `InvalidModTypeError` is raised.
- **residue** (*str or None*) – String indicating the modified amino acid, e.g., ‘Y’ or ‘tyrosine’. If `None`, indicates that the residue at the modification site is unknown or unspecified.
- **position** (*str or None*) – String indicating the position of the modified amino acid, e.g., ‘202’. If `None`, indicates that the position is unknown or unspecified.
- **is\_modified** (*bool*) – Specifies whether the modification is present or absent. Setting the flag specifies that the Agent with the `ModCondition` is unmodified at the site.

## Examples

Doubly-phosphorylated MEK (MAP2K1):

```
>>> phospho_mek = Agent('MAP2K1', mods=(
... ModCondition('phosphorylation', 'S', '202'),
... ModCondition('phosphorylation', 'S', '204')))
```

ERK (MAPK1) unphosphorylated at tyrosine 187:

```
>>> unphos_erk = Agent('MAPK1', mods=(
... ModCondition('phosphorylation', 'Y', '187', is_modified=False)))
```

**class** `indra.statements.Modification` (*enz*, *sub*, *residue=None*, *position=None*, *evidence=None*)

Bases: `indra.statements.Statement`

Generic statement representing the modification of a protein.

#### Parameters

- **enz** (`:py:class`indra.statement.Agent``) – The enzyme involved in the modification.

- **sub** (`indra.statement.Agent`) – The substrate of the modification.
- **residue** (*str or None*) – The amino acid residue being modified, or None if it is unknown or unspecified.
- **position** (*str or None*) – The position of the modified amino acid, or None if it is unknown or unspecified.
- **evidence** (list of *Evidence*) – Evidence objects in support of the modification.

**class** `indra.statements.MutCondition` (*position, residue\_from, residue\_to=None*)  
Bases: `object`

Mutation state of an amino acid position of an Agent.

#### Parameters

- **position** (*str*) – Residue position of the mutation in the protein sequence.
- **residue\_from** (*str*) – Wild-type (unmodified) amino acid residue at the given position.
- **residue\_to** (*str*) – Amino acid at the position resulting from the mutation.

#### Examples

Represent EGFR with a L858R mutation:

```
>>> egfr_mutant = Agent('EGFR', mutations=(MutCondition('858', 'L', 'R')))
```

**class** `indra.statements.Myristoylation` (*enz, sub, residue=None, position=None, evidence=None*)  
Bases: `indra.statements.AddModification`

Myristoylation modification.

**class** `indra.statements.Palmitoylation` (*enz, sub, residue=None, position=None, evidence=None*)  
Bases: `indra.statements.AddModification`

Palmitoylation modification.

**class** `indra.statements.Phosphorylation` (*enz, sub, residue=None, position=None, evidence=None*)  
Bases: `indra.statements.AddModification`

Phosphorylation modification.

#### Examples

MEK (MAP2K1) phosphorylates ERK (MAPK1) at threonine 185:

```
>>> mek = Agent('MAP2K1')
>>> erk = Agent('MAPK1')
>>> phos = Phosphorylation(mek, erk, 'T', '185')
```

**class** `indra.statements.RegulateActivity`  
Bases: `indra.statements.Statement`

Regulation of activity.

This class implements shared functionality of Activation and Inhibition statements and it should not be instantiated directly.

**class** `indra.statements.RegulateAmount` (*subj, obj, evidence=None*)  
Bases: `indra.statements.Statement`

Superclass handling operations on directed, two-element interactions.

**class** `indra.statements.Ribosylation` (*enz, sub, residue=None, position=None, evidence=None*)  
Bases: `indra.statements.AddModification`

Ribosylation modification.

**class** `indra.statements.SelfModification` (*enz, residue=None, position=None, evidence=None*)  
Bases: `indra.statements.Statement`

Generic statement representing the self-modification of a protein.

#### Parameters

- **enz** (`:py:class`indra.statement.Agent``) – The enzyme involved in the modification, which is also the substrate.
- **residue** (`str or None`) – The amino acid residue being modified, or `None` if it is unknown or unspecified.
- **position** (`str or None`) – The position of the modified amino acid, or `None` if it is unknown or unspecified.
- **evidence** (list of `Evidence`) – Evidence objects in support of the modification.

**class** `indra.statements.Statement` (*evidence=None, supports=None, supported\_by=None*)  
Bases: `object`

The parent class of all statements.

#### Parameters

- **evidence** (list of `Evidence`) – If a list of Evidence objects is passed to the constructor, the value is set to this list. If a bare Evidence object is passed, it is enclosed in a list. If no evidence is passed (the default), the value is set to an empty list.
- **supports** (list of `Statement`) – Statements that this Statement supports.
- **supported\_by** (list of `Statement`) – Statements supported by this statement.

**to\_graph** ()  
Return Statement as a networkx graph.

**to\_json** ()  
Return serialized Statement as a json dict.

**class** `indra.statements.Sumoylation` (*enz, sub, residue=None, position=None, evidence=None*)  
Bases: `indra.statements.AddModification`

Sumoylation modification.

**class** `indra.statements.Translocation` (*agent, from\_location=None, to\_location=None, evidence=None*)  
Bases: `indra.statements.Statement`

The translocation of a molecular agent from one location to another.

#### Parameters

- **agent** (`Agent`) – The agent which translocates.
- **from\_location** (`Optional[str]`) – The location from which the agent translocates. This must be a valid GO cellular component name (e.g. “cytoplasm”) or ID (e.g. “GO:0005737”).

- **to\_location** (*Optional[str]*) – The location to which the agent translocates. This must be a valid GO cellular component name or ID.

**class** `indra.statements.Transphosphorylation` (*enz*, *residue=None*, *position=None*, *evidence=None*)

Bases: `indra.statements.SelfModification`

Autophosphorylation in *trans*.

Transphosphorylation assumes that a kinase is already bound to a substrate (usually of the same molecular species), and phosphorylates it in an intra-molecular fashion. The `enz` property of the statement must have exactly one `bound_conditions` entry, and we assume that `enz` phosphorylates this molecule. The `bound_neg` property is ignored here.

**class** `indra.statements.Ubiquitination` (*enz*, *sub*, *residue=None*, *position=None*, *evidence=None*)

Bases: `indra.statements.AddModification`

Ubiquitination modification.

`indra.statements.get_valid_location` (*location*)

Check if the given location represents a valid cellular component.

`indra.statements.get_valid_residue` (*residue*)

Check if the given string represents a valid amino acid residue.

## 4.2 Processors for model input (`indra.sources`)

### 4.2.1 BEL (`indra.sources.bel`)

#### BEL API (`indra.sources.bel.bel_api`)

`indra.sources.bel.bel_api.process_belrdf` (*rdf\_str*, *print\_output=True*)

Return a BelProcessor for a BEL/RDF string.

**Parameters** `rdf_str` (*str*) – A BEL/RDF string to be processed. This will usually come from reading a .rdf file.

**Returns** `bp` – A BelProcessor object which contains INDRA Statements in `bp.statements`.

**Return type** *BelProcessor*

#### Notes

This function calls all the specific `get_type_of_mechanism()` functions of the newly constructed BelProcessor to extract INDRA Statements.

`indra.sources.bel.bel_api.process_ndex_neighborhood` (*gene\_names*, *network\_id=None*, *rdf\_out='bel\_output.rdf'*, *print\_output=True*)

Return a BelProcessor for an NDEx network neighborhood.

#### Parameters

- **gene\_names** (*list*) – A list of HGNC gene symbols to search the neighborhood of. Example: ['BRAF', 'MAP2K1']
- **network\_id** (*Optional[str]*) – The UUID of the network in NDEx. By default, the BEL Large Corpus network is used.

- **rdf\_out** (*Optional[str]*) – Name of the output file to save the RDF returned by the web service. This is useful for debugging purposes or to repeat the same query on an offline RDF file later. Default: `bel_output.rdf`

**Returns** `bp` – A `BelProcessor` object which contains INDRA Statements in `bp.statements`.

**Return type** `BelProcessor`

## Notes

This function calls `process_belrdf` to the returned RDF string from the webservice.

## BEL Processor (`indra.sources.bel.processor`)

**class** `indra.sources.bel.processor.BelProcessor(g)`

The `BelProcessor` extracts INDRA Statements from a BEL RDF model.

**Parameters** `g` (`rdflib.Graph`) – An RDF graph object containing the BEL model.

**g**

`rdflib.Graph` – An RDF graph object containing the BEL model.

**statements**

*list[indra.statements.Statement]* – A list of extracted INDRA Statements representing direct mechanisms. This list should be used for assembly in INDRA.

**indirect\_stmts**

*list[indra.statements.Statement]* – A list of extracted INDRA Statements representing indirect mechanisms. This list should be used for assembly or model checking in INDRA.

**converted\_direct\_stmts**

*list[str]* – A list of all direct BEL statements, as strings, that were converted into INDRA Statements.

**converted\_indirect\_stmts**

*list[str]* – A list of all indirect BEL statements, as strings, that were converted into INDRA Statements.

**degenerate\_stmts**

*list[str]* – A list of degenerate BEL statements, as strings, in the BEL model.

**all\_direct\_stmts**

*list[str]* – A list of all BEL statements representing direct interactions, as strings, in the BEL model.

**all\_indirect\_stmts**

*list[str]* – A list of all BEL statements that represent indirect interactions, as strings, in the BEL model.

**get\_activating\_mods()**

Extract INDRA ActiveForm Statements with a single mod from BEL.

The SPARQL pattern used for extraction from BEL looks for a `ModifiedProteinAbundance` as subject and an `Activity` of a `ProteinAbundance` as object.

## Examples

```
proteinAbundance(HGNC:INSR,proteinModification(P,Y))    directlyIncreases    kinaseActiv-  
ity(proteinAbundance(HGNC:INSR))
```



**get\_activating\_subs ()**

Extract INDRA ActiveForm Statements based on a mutation from BEL.

The SPARQL pattern used to extract ActiveForm Statements due to mutations look for a ProteinAbundance as a subject which has a child encoding the amino acid substitution. The object of the statement is an ActivityType of the same ProteinAbundance, which is either increased or decreased.

**Examples**

```
proteinAbundance(HGNC:NRAS,substitution(Q,61,K))      directlyIncreases      gtpBoundActiv-
ity(proteinAbundance(HGNC:NRAS))
```

```
proteinAbundance(HGNC:TP53,substitution(F,134,I))    directlyDecreases      transcriptionalActiv-
ity(proteinAbundance(HGNC:TP53))
```

**get\_activation ()**

Extract INDRA Inhibition/Activation Statements from BEL.

The SPARQL query used to extract Activation Statements looks for patterns in which the subject is is an ActivityType (of a ProteinAbundance) or an Abundance (of a small molecule). The object has to be the ActivityType (typically of a ProteinAbundance) which is either increased or decreased.

**Examples**

```
abundance(CHEBI:gefitinib) directlyDecreases kinaseActivity(proteinAbundance(HGNC:EGFR))
```

```
kinaseActivity(proteinAbundance(HGNC:MAP3K5))        directlyIncreases      kinaseActiv-
ity(proteinAbundance(HGNC:MAP2K7))
```

This pattern covers the extraction of Gap/Gef and GtpActivation Statements, which are recognized by the object activity or the subject activity, respectively, being *gtpbound*.

**Examples**

```
catalyticActivity(proteinAbundance(HGNC:RASA1))      directlyDecreases      gtpBoundActiv-
ity(proteinAbundance(PFH:"RAS Family"))
```

```
catalyticActivity(proteinAbundance(HGNC:SOS1))      directlyIncreases      gtpBoundActiv-
ity(proteinAbundance(HGNC:HRAS))
```

```
gtpBoundActivity(proteinAbundance(HGNC:HRAS))      directlyIncreases      catalyticActiv-
ity(proteinAbundance(HGNC:TIAM1))
```

**get\_all\_direct\_statements ()**

Get all directlyIncreases/Decreases BEL statements.

This method stores the results of the query in `self.all_direct_stmts` as a list of strings. The SPARQL query used to find direct BEL statements searches for all statements whose predicate is either `DirectlyIncreases` or `DirectlyDecreases`.

**get\_all\_indirect\_statements ()**

Get all indirect increases/decreases BEL statements.

This method stores the results of the query in `self.all_indirect_stmts` as a list of strings. The SPARQL query used to find indirect BEL statements searches for all statements whose predicate is either `Increases` or `Decreases`.

**get\_complexes ()**

Extract INDRA Complex Statements from BEL.

The SPARQL query used to extract Complexes looks for ComplexAbundance terms and their constituents. This pattern is distinct from other patterns in this processor in that it queries for terms, not full statements.

**Examples**

```
complexAbundance(proteinAbundance(HGNC:PPARG), proteinAbundance(HGNC:RXRA)) decreases
biologicalProcess(MESHPP:"Insulin Resistance")
```

**get\_composite\_activating\_mods ()**

Extract INDRA ActiveForm Statements with multiple mods from BEL.

The SPARQL pattern used for extraction from BEL looks for a CompositeAbundance as subject where two constituents of the composite are both ModifiedProteinAbundances. The object has to be a Activity of a ProteinAbundance.

**Examples**

```
compositeAbundance( proteinAbundance(PFH:"AKT Family",proteinModification(P,S,473)), pro-
teinAbundance(PFH:"AKT Family",proteinModification(P,T,308))) directlyIncreases kinaseActiv-
ity(proteinAbundance(PFH:"AKT Family"))
```

**get\_conversions ()**

Extract Conversion INDRA Statements from BEL.

The SPARQL query used to extract Conversions searches for a subject (controller) which is an Abundance-Activity which directlyIncreases a Reaction with a given list of Reactants and Products.

**Examples**

```
catalyticActivity(proteinAbundance(HGNC:HMOX1)) directlyIncreases reac-
tion(reactants(abundance(CHEBI:heme)), products(abundance(SCHEM: Biliverdine), abun-
dance(CHEBI:"carbon monoxide")))
```

**get\_degenerate\_statements ()**

Get all degenerate BEL statements.

Stores the results of the query in self.degenerate\_stmts.

**get\_modifications ()**

Extract INDRA Modification Statements from BEL.

Two SPARQL patterns are used for extracting Modifications from BEL:

- q\_phospho1 assumes that the subject is an AbundanceActivity, which increases/decreases a Modified-ProteinAbundance.

Examples:

```
kinaseActivity(proteinAbundance(HGNC:IKBKE)) directlyIncreases proteinAbun-
dance(HGNC:IRF3,proteinModification(P,S,385))
```

```
phosphataseActivity(proteinAbundance(HGNC:DUSP4)) directlyDecreases proteinAbun-
dance(HGNC:MAPK1,proteinModification(P,T,185))
```

- `q_phospho2` assumes that the subject is a ProteinAbundance which increases/decreases a Modified-ProteinAbundance.

Examples:

```
proteinAbundance(HGNC:NGF)          increases          proteinAbun-
dance(HGNC:NFKBIA,proteinModification(P,Y,42))
proteinAbundance(HGNC:FGF1)        decreases          proteinAbun-
dance(HGNC:RB1,proteinModification(P))
```

#### **get\_transcription ()**

Extract Increase/DecreaseAmount INDRA Statements from BEL.

Three distinct SPARQL patterns are used to extract amount regulations from BEL.

- `q_tscript1` searches for a subject which is a Transcription ActivityType of a ProteinAbundance and an object which is an RNAAbundance that is either increased or decreased.

Examples:

```
transcriptionalActivity(proteinAbundance(HGNC:FOXP2)) directlyIncreases  rnaAbun-
dance(HGNC:SYK)
transcriptionalActivity(proteinAbundance(HGNC:FOXP2)) directlyDecreases  rnaAbun-
dance(HGNC:CALCRL)
```

- `q_tscript2` searches for a subject which is a ProteinAbundance and an object which is an RNAAbundance. Note that this pattern typically exists in an indirect form (i.e. increases/decreases).

Example:

```
proteinAbundance(HGNC:MTF1) directlyIncreases rnaAbundance(HGNC:LCN1)
```

- `q_tscript3` searches for a subject which is a ModifiedProteinAbundance, with an object which is an RNAAbundance. In the BEL large corpus, this pattern is found for subjects which are protein families or mouse/rat proteins, and the predicate in an indirect increase.

Example:

```
proteinAbundance(PFR:"Akt Family",proteinModification(P)) increases  rnaAbun-
dance(RGD:Cald1)
```

#### **print\_statement\_coverage ()**

Display how many of the direct statements have been converted.

Also prints how many are considered 'degenerate' and not converted.

#### **print\_statements ()**

Print all extracted INDRA Statements.

`indra.sources.bel.processor.namespace_from_uri (uri)`

Return the entity namespace from the URI. Examples: [http://www.openbel.org/bel/p\\_HGNC\\_RAF1](http://www.openbel.org/bel/p_HGNC_RAF1) -> HGNC  
[http://www.openbel.org/bel/p\\_RGD\\_Raf1](http://www.openbel.org/bel/p_RGD_Raf1) -> RGD [http://www.openbel.org/bel/p\\_PFH\\_MEK1/2\\_Family](http://www.openbel.org/bel/p_PFH_MEK1/2_Family) -> PFH

`indra.sources.bel.processor.term_from_uri (uri)`

Removes prepended URI information from terms.

## 4.2.2 Biopax (`indra.sources.biopax`)

### Biopax API (`indra.sources.biopax.biopax_api`)

`indra.sources.biopax.biopax_api.process_model` (*model*)

Returns a BiopaxProcessor for a BioPAX model object.

**Parameters** `model` (*org.biopax.paxtools.model.Model*) – A BioPAX model object.

**Returns** `bp` – A BiopaxProcessor containing the obtained BioPAX model in `bp.model`.

**Return type** *BiopaxProcessor*

`indra.sources.biopax.biopax_api.process_owl` (*owl\_filename*)

Returns a BiopaxProcessor for a BioPAX OWL file.

**Parameters** `owl_filename` (*string*) – The name of the OWL file to process.

**Returns** `bp` – A BiopaxProcessor containing the obtained BioPAX model in `bp.model`.

**Return type** *BiopaxProcessor*

`indra.sources.biopax.biopax_api.process_pc_neighborhood` (*gene\_names*, *neighbor\_limit=1*, *database\_filter=None*)

Returns a BiopaxProcessor for a PathwayCommons neighborhood query.

The neighborhood query finds the neighborhood around a set of source genes.

<http://www.pathwaycommons.org/pc2/#graph>

[http://www.pathwaycommons.org/pc2/#graph\\_kind](http://www.pathwaycommons.org/pc2/#graph_kind)

#### Parameters

- **gene\_names** (*list*) – A list of HGNC gene symbols to search the neighborhood of. Examples: ['BRAF'], ['BRAF', 'MAP2K1']
- **neighbor\_limit** (*Optional[int]*) – The number of steps to limit the size of the neighborhood around the gene names being queried. Default: 1
- **database\_filter** (*Optional[list]*) – A list of database identifiers to which the query is restricted. Examples: ['reactome'], ['biogrid', 'pid', 'psp'] If not given, all databases are used in the query. For a full list of databases see <http://www.pathwaycommons.org/pc2/datasources>

**Returns** `bp` – A BiopaxProcessor containing the obtained BioPAX model in `bp.model`.

**Return type** *BiopaxProcessor*

`indra.sources.biopax.biopax_api.process_pc_pathsbetween` (*gene\_names*, *neighbor\_limit=1*, *database\_filter=None*, *block\_size=None*)

Returns a BiopaxProcessor for a PathwayCommons paths-between query.

The paths-between query finds the paths between a set of genes. Here source gene names are given in a single list and all directions of paths between these genes are considered.

<http://www.pathwaycommons.org/pc2/#graph>

[http://www.pathwaycommons.org/pc2/#graph\\_kind](http://www.pathwaycommons.org/pc2/#graph_kind)

#### Parameters

- **gene\_names** (*list*) – A list of HGNC gene symbols to search for paths between. Examples: ['BRAF', 'MAP2K1']
- **neighbor\_limit** (*Optional[int]*) – The number of steps to limit the length of the paths between the gene names being queried. Default: 1
- **database\_filter** (*Optional[list]*) – A list of database identifiers to which the query is restricted. Examples: ['reactome'], ['biogrid', 'pid', 'psp'] If not given, all databases are used in the query. For a full list of databases see <http://www.pathwaycommons.org/pc2/datasources>
- **block\_size** (*Optional[int]*) – Large paths-between queries (above ~60 genes) can error on the server side. In this case, the query can be replaced by a series of smaller paths-between and paths-from-to queries each of which contains block\_size genes.

**Returns** **bp** – A BiopaxProcessor containing the obtained BioPAX model in bp.model.

**Return type** *BiopaxProcessor*

```
indra.sources.biopax.biopax_api.process_pc_pathsfromto(source_genes, target_genes,
                                                         neighbor_limit=1,
                                                         database_filter=None)
```

Returns a BiopaxProcessor for a PathwayCommons paths-from-to query.

The paths-from-to query finds the paths from a set of source genes to a set of target genes.

<http://www.pathwaycommons.org/pc2/#graph>

[http://www.pathwaycommons.org/pc2/#graph\\_kind](http://www.pathwaycommons.org/pc2/#graph_kind)

#### Parameters

- **source\_genes** (*list*) – A list of HGNC gene symbols that are the sources of paths being searched for. Examples: ['BRAF', 'RAF1', 'ARAF']
- **target\_genes** (*list*) – A list of HGNC gene symbols that are the targets of paths being searched for. Examples: ['MAP2K1', 'MAP2K2']
- **neighbor\_limit** (*Optional[int]*) – The number of steps to limit the length of the paths between the source genes and target genes being queried. Default: 1
- **database\_filter** (*Optional[list]*) – A list of database identifiers to which the query is restricted. Examples: ['reactome'], ['biogrid', 'pid', 'psp'] If not given, all databases are used in the query. For a full list of databases see <http://www.pathwaycommons.org/pc2/datasources>

**Returns** **bp** – A BiopaxProcessor containing the obtained BioPAX model in bp.model.

**Return type** *BiopaxProcessor*

### Biopax Processor (`indra.sources.biopax.processor`)

**class** `indra.sources.biopax.processor.BiopaxProcessor` (*model*)

The BiopaxProcessor extracts INDRA Statements from a BioPAX model.

The BiopaxProcessor uses pattern searches in a BioPAX OWL model to extract mechanisms from which it constructs INDRA Statements.

**Parameters** **model** (*org.biopax.paxtools.model.Model*) – A BioPAX model object (java object)

**model**

*org.biopax.paxtools.model.Model* – A BioPAX model object (java object) which is queried using Paxtools to extract INDRA Statements

**statements**

*list[indra.statements.Statement]* – A list of INDRA Statements that were extracted from the model.

**get\_activity\_modification ()**

Extract INDRA ActiveForm statements from the BioPAX model.

This method extracts ActiveForm Statements that are due to protein modifications. This method reuses the structure of BioPAX Pattern's *org.biopax.paxtools.pattern.PatternBox.constrolsStateChange* pattern with additional constraints to specify the gain or loss of a modification occurring (phosphorylation, deubiquitination, etc.) and the gain or loss of activity due to the modification state change.

**get\_complexes ()**

Extract INDRA Complex Statements from the BioPAX model.

This method searches for *org.biopax.paxtools.model.level3.Complex* objects which represent molecular complexes. It doesn't reuse BioPAX Pattern's *org.biopax.paxtools.pattern.PatternBox.inComplexWith* query since that retrieves pairs of complex members rather than the full complex.

**get\_conversions ()**

Extract Conversion INDRA Statements from the BioPAX model.

This method uses a custom BioPAX Pattern (one that is not implemented *PatternBox*) to query for BiochemicalReactions whose left and right hand sides are collections of *SmallMolecules*. This pattern thereby extracts metabolic conversions as well as signaling processes via small molecules (e.g. lipid phosphorylation or cleavage).

**get\_gap ()**

Extract Gap INDRA Statements from the BioPAX model.

This method uses a custom BioPAX Pattern (one that is not implemented *PatternBox*) to query for controlled BiochemicalReactions in which the same protein is in complex with GTP on the left hand side and in complex with GDP on the right hand side. This implies that the controller is a GAP for the GDP/GTP-bound protein.

**get\_gef ()**

Extract Gef INDRA Statements from the BioPAX model.

This method uses a custom BioPAX Pattern (one that is not implemented *PatternBox*) to query for controlled BiochemicalReactions in which the same protein is in complex with GDP on the left hand side and in complex with GTP on the right hand side. This implies that the controller is a GEF for the GDP/GTP-bound protein.

**get\_modifications ()**

Extract INDRA Modification Statements from the BioPAX model.

To extract Modifications, this method reuses the structure of BioPAX Pattern's *org.biopax.paxtools.pattern.PatternBox.constrolsStateChange* pattern with additional constraints to specify the type of state change occurring (phosphorylation, deubiquitination, etc.).

**get\_regulate\_activities ()**

Get Activation/Inhibition INDRA Statements from the BioPAX model.

This method extracts Activation/Inhibition Statements and reuses the structure of BioPAX Pattern's *org.biopax.paxtools.pattern.PatternBox.constrolsStateChange* pattern with additional constraints to specify the gain or loss of activity state but assuring that the activity change is not due to a modification state change (which are extracted by *get\_modifications* and *get\_activity\_modification*).

**get\_regulate\_amounts ()**

Extract INDRA RegulateAmount Statements from the BioPAX model.

This method extracts IncreaseAmount/DecreaseAmount Statements from the BioPAX model. It fully reuses BioPAX Pattern's `org.biopax.paxtools.pattern.PatternBox.controlsExpressionWithTemplateReaction` to find TemplateReactions which control the expression of a protein.

**print\_statements ()**

Print all INDRA Statements collected by the processors.

**save\_model (file\_name=None)**

Save the BioPAX model object in an OWL file.

**Parameters** `file_name` (*Optional[str]*) – The name of the OWL file to save the model in.

**Pathway Commons Client (indra.sources.biopax.pathway\_commons\_client)**

`indra.sources.biopax.pathway_commons_client.graph_query (kind, source, target=None, neighbor_limit=1, database_filter=None)`

Perform a graph query on PathwayCommons.

For more information on these queries, see <http://www.pathwaycommons.org/pc2/#graph>

**Parameters**

- **kind** (*str*) – The kind of graph query to perform. Currently 3 options are implemented, 'neighborhood', 'pathsbetween' and 'pathsfromto'.
- **source** (*list[str]*) – A list of gene names which are the source set for the graph query.
- **target** (*Optional[list[str]]*) – A list of gene names which are the target set for the graph query. Only needed for 'pathsfromto' queries.
- **neighbor\_limit** (*Optional[int]*) – This limits the length of the longest path considered in the graph query. Default: 1

**Returns** `model` – A BioPAX model (java object).

**Return type** `org.biopax.paxtools.model.Model`

`indra.sources.biopax.pathway_commons_client.model_to_owl (model, fname)`

Save a BioPAX model object as an OWL file.

**Parameters**

- **model** (*org.biopax.paxtools.model.Model*) – A BioPAX model object (java object).
- **fname** (*str*) – The name of the OWL file to save the model in.

`indra.sources.biopax.pathway_commons_client.owl_str_to_model (owl_str)`

Return a BioPAX model object from an OWL string.

**Parameters** `owl_str` (*str*) – The model as an OWL string.

**Returns** `biopax_model` – A BioPAX model object (java object).

**Return type** `org.biopax.paxtools.model.Model`

`indra.sources.biopax.pathway_commons_client.owl_to_model (fname)`

Return a BioPAX model object from an OWL file.

**Parameters** `fname` (*str*) – The name of the OWL file containing the model.

**Returns** `biopax_model` – A BioPAX model object (java object).

**Return type** `org.biopax.paxtools.model.Model`

## 4.2.3 REACH (`indra.sources.reach`)

### REACH API (`indra.sources.reach.reach_api`)

`indra.sources.reach.reach_api.process_json_file` (*file\_name*, *citation=None*)

Return a ReachProcessor by processing the given REACH json file.

The output from the REACH parser is in this json format. This function is useful if the output is saved as a file and needs to be processed. For more information on the format, see: <https://github.com/clulab/reach>

#### Parameters

- **file\_name** (*str*) – The name of the json file to be processed.
- **citation** (*Optional[str]*) – A PubMed ID passed to be used in the evidence for the extracted INDRA Statements. Default: None

**Returns** `rp` – A ReachProcessor containing the extracted INDRA Statements in `rp.statements`.

**Return type** `ReachProcessor`

`indra.sources.reach.reach_api.process_json_str` (*json\_str*, *citation=None*)

Return a ReachProcessor by processing the given REACH json string.

The output from the REACH parser is in this json format. For more information on the format, see: <https://github.com/clulab/reach>

#### Parameters

- **json\_str** (*str*) – The json string to be processed.
- **citation** (*Optional[str]*) – A PubMed ID passed to be used in the evidence for the extracted INDRA Statements. Default: None

**Returns** `rp` – A ReachProcessor containing the extracted INDRA Statements in `rp.statements`.

**Return type** `ReachProcessor`

`indra.sources.reach.reach_api.process_nxml_file` (*file\_name*, *citation=None*, *offline=False*)

Return a ReachProcessor by processing the given NXML file.

NXML is the format used by PubmedCentral for papers in the open access subset.

#### Parameters

- **file\_name** (*str*) – The name of the NXML file to be processed.
- **citation** (*Optional[str]*) – A PubMed ID passed to be used in the evidence for the extracted INDRA Statements. Default: None
- **offline** (*Optional[bool]*) – If set to True, the REACH system is ran offline. Otherwise (by default) the web service is called. Default: False

**Returns** `rp` – A ReachProcessor containing the extracted INDRA Statements in `rp.statements`.

**Return type** `ReachProcessor`



`indra.sources.reach.reach_api.process_nxml_str` (*nxml\_str*, *citation=None*, *offline=False*)  
Return a ReachProcessor by processing the given NXML string.

NXML is the format used by PubmedCentral for papers in the open access subset.

#### Parameters

- **nxml\_str** (*str*) – The NXML string to be processed.
- **citation** (*Optional[str]*) – A PubMed ID passed to be used in the evidence for the extracted INDRA Statements. Default: None
- **offline** (*Optional[bool]*) – If set to True, the REACH system is ran offline. Otherwise (by default) the web service is called. Default: False

**Returns** **rp** – A ReachProcessor containing the extracted INDRA Statements in `rp.statements`.

**Return type** *ReachProcessor*

`indra.sources.reach.reach_api.process_pmc` (*pmc\_id*, *offline=False*)  
Return a ReachProcessor by processing a paper with a given PMC id.

Uses the PMC client to obtain the full text. If it's not available, None is returned.

#### Parameters

- **pmc\_id** (*str*) – The ID of a PubmedCentral article. The string may start with PMC but passing just the ID also works. Examples: 3717945, PMC3717945 <https://www.ncbi.nlm.nih.gov/pmc/>
- **offline** (*Optional[bool]*) – If set to True, the REACH system is ran offline. Otherwise (by default) the web service is called. Default: False

**Returns** **rp** – A ReachProcessor containing the extracted INDRA Statements in `rp.statements`.

**Return type** *ReachProcessor*

`indra.sources.reach.reach_api.process_pubmed_abstract` (*pubmed\_id*, *offline=False*)  
Return a ReachProcessor by processing an abstract with a given Pubmed id.

Uses the Pubmed client to get the abstract. If that fails, None is returned.

#### Parameters

- **pubmed\_id** (*str*) – The ID of a Pubmed article. The string may start with PMID but passing just the ID also works. Examples: 27168024, PMID27168024 <https://www.ncbi.nlm.nih.gov/pubmed/>
- **offline** (*Optional[bool]*) – If set to True, the REACH system is ran offline. Otherwise (by default) the web service is called. Default: False

**Returns** **rp** – A ReachProcessor containing the extracted INDRA Statements in `rp.statements`.

**Return type** *ReachProcessor*

`indra.sources.reach.reach_api.process_text` (*text*, *citation=None*, *offline=False*)  
Return a ReachProcessor by processing the given text.

#### Parameters

- **text** (*str*) – The text to be processed.
- **citation** (*Optional[str]*) – A PubMed ID passed to be used in the evidence for the extracted INDRA Statements. This is used when the text to be processed comes from a publication that is not otherwise identified. Default: None

- **offline** (*Optional[bool]*) – If set to True, the REACH system is ran offline. Otherwise (by default) the web service is called. Default: False

**Returns** **rp** – A ReachProcessor containing the extracted INDRA Statements in `rp.statements`.

**Return type** *ReachProcessor*

### REACH Processor (`indra.sources.reach.processor`)

**class** `indra.sources.reach.processor.ReachProcessor` (*json\_dict*, *pmid=None*)

The ReachProcessor extracts INDRA Statements from REACH parser output.

#### Parameters

- **json\_dict** (*dict*) – A JSON dictionary containing the REACH extractions.
- **pmid** (*Optional[str]*) – The PubMed ID associated with the extractions. This can be passed in case the PMID cannot be determined from the extractions alone.

#### **tree**

*objectpath.Tree* – The objectpath Tree object representing the extractions.

#### **statements**

*list[indra.statements.Statement]* – A list of INDRA Statements that were extracted by the processor.

#### **citation**

*str* – The PubMed ID associated with the extractions.

#### **all\_events**

*dict[str, str]* – The frame IDs of all events by type in the REACH extraction.

#### **get\_activation()**

Extract INDRA Activation Statements.

#### **get\_all\_events()**

Gather all event IDs in the REACH output by type.

These IDs are stored in the `self.all_events` dict.

#### **get\_complexes()**

Extract INDRA Complex Statements.

#### **get\_modifications()**

Extract Modification INDRA Statements.

#### **get\_regulate\_amounts()**

Extract RegulateAmount INDRA Statements.

#### **get\_translocation()**

Extract INDRA Translocation Statements.

#### **print\_event\_statistics()**

Print the number of events in the REACH output by type.

### REACH reader (`indra.sources.reach.reach_reader`)

**class** `indra.sources.reach.reach_reader.ReachReader`

The ReachReader wraps a singleton instance of the REACH reader.

This allows calling the reader many times without having to wait for it to start up each time.

**api\_ruler**

*org.clulab.reach.apis.ApiRuler* – An instance of the REACH ApiRuler class (java object).

**get\_api\_ruler()**

Return the existing reader if it exists or launch a new one.

**Returns** **api\_ruler** – An instance of the REACH ApiRuler class (java object).

**Return type** *org.clulab.reach.apis.ApiRuler*

**4.2.4 TRIPS (*indra.sources.trips*)****TRIPS API (*indra.sources.trips.trips\_api*)**

*indra.sources.trips.trips\_api.process\_text* (*text*, *save\_xml\_name='trips\_output.xml'*,  
*save\_xml\_pretty=True*, *offline=False*, *service\_endpoint='drum'*)

Return a TripsProcessor by processing text.

**Parameters**

- **text** (*str*) – The text to be processed.
- **save\_xml\_name** (*Optional[str]*) – The name of the file to save the returned TRIPS extraction knowledge base XML. Default: *trips\_output.xml*
- **save\_xml\_pretty** (*Optional[bool]*) – If True, the saved XML is pretty-printed. Some third-party tools require non-pretty-printed XMLs which can be obtained by setting this to False. Default: True
- **offline** (*Optional[bool]*) – If True, offline reading is used with a local instance of DRUM, if available. Default: False
- **service\_endpoint** (*Optional[str]*) – Selects the TRIPS/DRUM web service endpoint to use. Is a choice between “drum” (default) and “drum-dev”, a nightly build.

**Returns** **tp** – A TripsProcessor containing the extracted INDRA Statements in *tp.statements*.

**Return type** *TripsProcessor*

*indra.sources.trips.trips\_api.process\_xml* (*xml\_string*)

Return a TripsProcessor by processing a TRIPS EKB XML string.

**Parameters** **xml\_string** (*str*) – A TRIPS extraction knowledge base (EKB) string to be processed. <http://trips.ihmc.us/parser/api.html>

**Returns** **tp** – A TripsProcessor containing the extracted INDRA Statements in *tp.statements*.

**Return type** *TripsProcessor*

**TRIPS Processor (*indra.sources.trips.processor*)**

**class** *indra.sources.trips.processor.TripsProcessor* (*xml\_string*)

The TripsProcessor extracts INDRA Statements from a TRIPS XML.

For more details on the TRIPS EKB XML format, see <http://trips.ihmc.us/parser/cgi/drum>

**Parameters** **xml\_string** (*str*) – A TRIPS extraction knowledge base (EKB) in XML format as a string.

**tree**

*xml.etree.ElementTree.Element* – An ElementTree object representation of the TRIPS EKB XML.

**statements**

*list[indra.statements.Statement]* – A list of INDRA Statements that were extracted from the EKB.

**doc\_id**

*str* – The PubMed ID of the paper that the extractions are from.

**sentences**

*dict[str: str]* – The list of all sentences in the EKB with their IDs

**paragraphs**

*dict[str: str]* – The list of all paragraphs in the EKB with their IDs

**par\_to\_sec**

*dict[str: str]* – A map from paragraph IDs to their associated section types

**extracted\_events**

*list[xml.etree.ElementTree.Element]* – A list of Event elements that have been extracted as INDRA Statements.

**get\_activations ()**

Extract direct Activation INDRA Statements.

**get\_activations\_causal ()**

Extract causal Activation INDRA Statements.

**get\_activations\_stimulate ()**

Extract Activation INDRA Statements via stimulation.

**get\_active\_forms ()**

Extract ActiveForm INDRA Statements.

**get\_active\_forms\_state ()**

Extract ActiveForm INDRA Statements.

**get\_agents ()**

Return list of INDRA Agents corresponding to TERMS in the EKB.

This is meant to be used when entities e.g. “phosphorylated ERK”, rather than events need to be extracted from processed natural language. These entities with their respective states are represented as INDRA Agents.

**Returns agents** – List of INDRA Agents extracted from EKB.

**Return type** *list[indra.statements.Agent]*

**get\_all\_events ()**

Make a list of all events in the TRIPS EKB.

The events are stored in *self.all\_events*.

**get\_complexes ()**

Extract Complex INDRA Statements.

**get\_degradations ()**

Extract Degradation INDRA Statements.

**get\_modifications ()**

Extract all types of Modification INDRA Statements.

**get\_modifications\_indirect ()**

Extract indirect Modification INDRA Statements.

**get\_regulate\_amounts** ()  
Extract Increase/DecreaseAmount Statements.

**get\_syntheses** ()  
Extract IncreaseAmount INDRA Statements.

### TRIPS Client (`indra.sources.trips.trips_client`)

`indra.sources.trips.trips_client.get_xml` (*html*)  
Extract the EKB XML from the HTML output of the TRIPS web service.

**Parameters** `html` (*str*) – The HTML output from the TRIPS web service.

**Returns**

- *The extraction knowledge base (EKB) XML that contains the event and term*
- *extractions.*

`indra.sources.trips.trips_client.save_xml` (*xml\_str*, *file\_name*, *pretty=True*)  
Save the TRIPS EKB XML in a file.

**Parameters**

- `xml_str` (*str*) – The TRIPS EKB XML string to be saved.
- `file_name` (*str*) – The name of the file to save the result in.
- `pretty` (*Optional[bool]*) – If True, the XML is pretty printed.

`indra.sources.trips.trips_client.send_query` (*text*, *service\_endpoint='drum'*,  
*query\_args=None*)

Send a query to the TRIPS web service.

**Parameters**

- `text` (*str*) – The text to be processed.
- `service_endpoint` (*Optional[str]*) – Selects the TRIPS/DRUM web service endpoint to use. Is a choice between “drum” (default) and “drum-dev”, a nightly build.
- `query_args` (*Optional[dict]*) – A dictionary of arguments to be passed with the query.

**Returns** `html` – The HTML result returned by the web service.

**Return type** `str`

### 4.2.5 NDEX CX API (`indra.sources.ndex_cx_api`)

`indra.sources.ndex_cx.ndex_cx_api.process_cx` (*cx\_json*, *require\_grounding=True*)  
Process a CX JSON object into Statements.

**Parameters**

- `cx_json` (*list*) – CX JSON object.
- `require_grounding` (*bool*) – Whether network nodes lacking grounding information should be included among the extracted Statements (default is True).

**Returns** Processor containing Statements.

**Return type** `NdexCxProcessor`

`indra.sources.ndex_cx.ndex_cx_api.process_cx_file` (*file\_name*, *require\_grounding=True*) *re-*

Process a CX JSON file into Statements.

**Parameters**

- **file\_name** (*str*) – Path to file containing CX JSON.
- **require\_grounding** (*bool*) – Whether network nodes lacking grounding information should be included among the extracted Statements (default is True).

**Returns** Processor containing Statements.

**Return type** *NdexCxProcessor*

`indra.sources.ndex_cx.ndex_cx_api.process_ndex_network` (*network\_id*, *username=None*, *password=None*, *require\_grounding=True*) *re-*

Process an NDEx network into Statements.

**Parameters**

- **network\_id** (*str*) – NDEx network ID.
- **username** (*str*) – NDEx username.
- **password** (*str*) – NDEx password.
- **require\_grounding** (*bool*) – Whether network nodes lacking grounding information should be included among the extracted Statements (default is True).

**Returns** Processor containing Statements. Returns None if there if the HTTP status code indicates an unsuccessful request.

**Return type** *NdexCxProcessor*

### **NDEx CX Processor** (`indra.sources.ndex_cx.processor`)

**class** `indra.sources.ndex_cx.processor.NdexCxProcessor` (*cx*, *require\_grounding=True*)

The `NdexCxProcessor` extracts INDRA Statements from Cytoscape CX JSON.

**Parameters** **cx** (*list of dicts*) – JSON content containing the Cytoscape network in CX format.

**statements**

*list* – A list of extracted INDRA Statements. Not all edges in the network may be converted into Statements.

**get\_agents** ()

Get list of grounded nodes in the network as Agents.

**Returns** Only nodes containing sufficient information to be grounded will be contained in this list.

**Return type** list of Agents

**get\_node\_names** ()

Get list of all nodes in the network by name.

**get\_pmids** ()

Get list of all PMIDs associated with edges in the network.

`get_statements()`  
Convert network edges into Statements.

**Returns** Converted INDRA Statements.

**Return type** list of Statements

## 4.3 Database clients (`indra.databases`)

### 4.3.1 HGNC client (`indra.hgnc_client`)

`indra.databases.hgnc_client.get_entrez_id(hgnc_id)`

Return the Entrez ID corresponding to the given HGNC ID.

**Parameters** `hgnc_id` (*str*) – The HGNC ID to be converted. Note that the HGNC ID is a number that is passed as a string. It is not the same as the HGNC gene symbol.

**Returns** `entrez_id` – The Entrez ID corresponding to the given HGNC ID.

**Return type** `str`

`indra.databases.hgnc_client.get_hgnc_entry`

Return the HGNC entry for the given HGNC ID from the web service.

**Parameters** `hgnc_id` (*str*) – The HGNC ID to be converted.

**Returns** `xml_tree` – The XML ElementTree corresponding to the entry for the given HGNC ID.

**Return type** `ElementTree`

`indra.databases.hgnc_client.get_hgnc_from_entrez(entrez_id)`

Return the HGNC ID corresponding to the given Entrez ID.

**Parameters** `entrez_id` (*str*) – The EntrezC ID to be converted, a number passed as a string.

**Returns** `hgnc_id` – The HGNC ID corresponding to the given Entrez ID.

**Return type** `str`

`indra.databases.hgnc_client.get_hgnc_from_mouse(mgi_id)`

Return the HGNC ID corresponding to the given MGI mouse gene ID.

**Parameters** `mgi_id` (*str*) – The MGI ID to be converted. Example: “2444934”

**Returns** `hgnc_id` – The HGNC ID corresponding to the given MGI ID.

**Return type** `str`

`indra.databases.hgnc_client.get_hgnc_from_rat(rgd_id)`

Return the HGNC ID corresponding to the given RGD rat gene ID.

**Parameters** `rgd_id` (*str*) – The RGD ID to be converted. Example: “1564928”

**Returns** `hgnc_id` – The HGNC ID corresponding to the given RGD ID.

**Return type** `str`

`indra.databases.hgnc_client.get_hgnc_id(hgnc_name)`

Return the HGNC ID corresponding to the given HGNC symbol.

**Parameters** `hgnc_name` (*str*) – The HGNC symbol to be converted. Example: BRAF

**Returns** `hgnc_id` – The HGNC ID corresponding to the given HGNC symbol.

**Return type** str

`indra.databases.hgnc_client.get_hgnc_name(hgnc_id)`  
Return the HGNC symbol corresponding to the given HGNC ID.

**Parameters** `hgnc_id` (*str*) – The HGNC ID to be converted.

**Returns** `hgnc_name` – The HGNC symbol corresponding to the given HGNC ID.

**Return type** str

`indra.databases.hgnc_client.get_mouse_id(hgnc_id)`  
Return the MGI mouse ID corresponding to the given HGNC ID.

**Parameters** `hgnc_id` (*str*) – The HGNC ID to be converted. Example: ""

**Returns** `mgid` – The MGI ID corresponding to the given HGNC ID.

**Return type** str

`indra.databases.hgnc_client.get_rat_id(hgnc_id)`  
Return the RGD rat ID corresponding to the given HGNC ID.

**Parameters** `hgnc_id` (*str*) – The HGNC ID to be converted. Example: ""

**Returns** `rgd_id` – The RGD ID corresponding to the given HGNC ID.

**Return type** str

`indra.databases.hgnc_client.get_uniprot_id(hgnc_id)`  
Return the UniProt ID corresponding to the given HGNC ID.

**Parameters** `hgnc_id` (*str*) – The HGNC ID to be converted. Note that the HGNC ID is a number that is passed as a string. It is not the same as the HGNC gene symbol.

**Returns** `uniprot_id` – The UniProt ID corresponding to the given HGNC ID.

**Return type** str

### 4.3.2 Uniprot client (`indra.databases.uniprot_client`)

`indra.databases.uniprot_client.get_family_members(family_name, human_only=True)`  
Return the HGNC gene symbols which are the members of a given family.

**Parameters**

- **family\_name** (*str*) – Family name to be queried.
- **human\_only** (*bool*) – If True, only human proteins in the family will be returned. Default: True

**Returns** `gene_names` – The HGNC gene symbols corresponding to the given family.

**Return type** list

`indra.databases.uniprot_client.get_gene_name(protein_id, web_fallback=True)`  
Return the gene name for the given UniProt ID.

This is an alternative to `get_hgnc_name` and is useful when HGNC name is not available (for instance, when the organism is not homo sapiens).

**Parameters**

- **protein\_id** (*str*) – UniProt ID to be mapped.



- **web\_fallback** (*Optional[bool]*) – If True and the offline lookup fails, the UniProt web service is used to do the query.

**Returns** **gene\_name** – The gene name corresponding to the given Uniprot ID.

**Return type** str

`indra.databases.uniprot_client.get_id_from_mgi(mgi_id)`

Return the UniProt ID given the MGI ID of a mouse protein.

**Parameters** **mgi\_id** (*str*) – The MGI ID of the mouse protein.

**Returns** **up\_id** – The UniProt ID of the mouse protein.

**Return type** str

`indra.databases.uniprot_client.get_id_from_mnemonic(uniprot_mnemonic)`

Return the UniProt ID for the given UniProt mnemonic.

**Parameters** **uniprot\_mnemonic** (*str*) – UniProt mnemonic to be mapped.

**Returns** **uniprot\_id** – The UniProt ID corresponding to the given Uniprot mnemonic.

**Return type** str

`indra.databases.uniprot_client.get_id_from_rgd(rgd_id)`

Return the UniProt ID given the RGD ID of a rat protein.

**Parameters** **rgd\_id** (*str*) – The RGD ID of the rat protein.

**Returns** **up\_id** – The UniProt ID of the rat protein.

**Return type** str

`indra.databases.uniprot_client.get_mgi_id(protein_id)`

Return the MGI ID given the protein id of a mouse protein.

**Parameters** **protein\_id** (*str*) – UniProt ID of the mouse protein

**Returns** **mgi\_id** – MGI ID of the mouse protein

**Return type** str

`indra.databases.uniprot_client.get_mnemonic(protein_id, web_fallback=False)`

Return the UniProt mnemonic for the given UniProt ID.

**Parameters**

- **protein\_id** (*str*) – UniProt ID to be mapped.
- **web\_fallback** (*Optional[bool]*) – If True and the offline lookup fails, the UniProt web service is used to do the query.

**Returns** **mnemonic** – The UniProt mnemonic corresponding to the given Uniprot ID.

**Return type** str

`indra.databases.uniprot_client.get_mouse_id(human_protein_id)`

Return the mouse UniProt ID given a human UniProt ID.

**Parameters** **human\_protein\_id** (*str*) – The UniProt ID of a human protein.

**Returns** **mouse\_protein\_id** – The UniProt ID of a mouse protein orthologous to the given human protein

**Return type** str

`indra.databases.uniprot_client.get_primary_id(protein_id)`

Return a primary entry corresponding to the UniProt ID.

**Parameters** `protein_id` (*str*) – The UniProt ID to map to primary.

**Returns** `primary_id` – If the given ID is primary, it is returned as is. Otherwise the primary IDs are looked up. If there are multiple primary IDs then the first human one is returned. If there are no human primary IDs then the first primary found is returned.

**Return type** `str`

`indra.databases.uniprot_client.get_rat_id(human_protein_id)`

Return the rat UniProt ID given a human UniProt ID.

**Parameters** `human_protein_id` (*str*) – The UniProt ID of a human protein.

**Returns** `rat_protein_id` – The UniProt ID of a rat protein orthologous to the given human protein

**Return type** `str`

`indra.databases.uniprot_client.get_rgd_id(protein_id)`

Return the RGD ID given the protein id of a rat protein.

**Parameters** `protein_id` (*str*) – UniProt ID of the rat protein

**Returns** `rgd_id` – RGD ID of the rat protein

**Return type** `str`

`indra.databases.uniprot_client.is_human(protein_id)`

Return True if the given protein id corresponds to a human protein.

**Parameters** `protein_id` (*str*) – UniProt ID of the protein

**Returns**

**Return type** True if the `protein_id` corresponds to a human protein, otherwise False.

`indra.databases.uniprot_client.is_mouse(protein_id)`

Return True if the given protein id corresponds to a mouse protein.

**Parameters** `protein_id` (*str*) – UniProt ID of the protein

**Returns**

**Return type** True if the `protein_id` corresponds to a mouse protein, otherwise False.

`indra.databases.uniprot_client.is_rat(protein_id)`

Return True if the given protein id corresponds to a rat protein.

**Parameters** `protein_id` (*str*) – UniProt ID of the protein

**Returns**

**Return type** True if the `protein_id` corresponds to a rat protein, otherwise False.

`indra.databases.uniprot_client.is_secondary(protein_id)`

Return True if the UniProt ID corresponds to a secondary accession.

**Parameters** `protein_id` (*str*) – The UniProt ID to check.

**Returns**

**Return type** True if it is a secondary accessing entry, False otherwise.

`indra.databases.uniprot_client.query_protein`

Return the UniProt entry as an RDF graph for the given UniProt ID.

**Parameters** `protein_id` (*str*) – UniProt ID to be queried.

**Returns** `g` – The RDF graph corresponding to the UniProt entry.

**Return type** `rdflib.Graph`

`indra.databases.uniprot_client.verify_location` (*protein\_id*, *residue*, *location*)

Return True if the residue is at the given location in the UP sequence.

**Parameters**

- **protein\_id** (*str*) – UniProt ID of the protein whose sequence is used as reference.
- **residue** (*str*) – A single character amino acid symbol (Y, S, T, V, etc.)
- **location** (*str*) – The location on the protein sequence (starting at 1) at which the residue should be checked against the reference sequence.

**Returns**

- *True if the given residue is at the given position in the sequence*
- *corresponding to the given UniProt ID, otherwise False.*

`indra.databases.uniprot_client.verify_modification` (*protein\_id*, *residue*, *location=None*)

Return True if the residue at the given location has a known modification.

**Parameters**

- **protein\_id** (*str*) – UniProt ID of the protein whose sequence is used as reference.
- **residue** (*str*) – A single character amino acid symbol (Y, S, T, V, etc.)
- **location** (*Optional[str]*) – The location on the protein sequence (starting at 1) at which the modification is checked.

**Returns**

- *True if the given residue is reported to be modified at the given position*
- *in the sequence corresponding to the given UniProt ID, otherwise False.*
- *If location is not given, we only check if there is any residue of the*
- *given type that is modified.*

### 4.3.3 ChEBI client (`indra.databases.chebi_client`)

`indra.databases.chebi_client.get_chebi_id_from_pubchem` (*pubchem\_id*)

Return the ChEBI ID corresponding to a given Pubchem ID.

**Parameters** `pubchem_id` (*str*) – Pubchem ID to be converted.

**Returns** `chebi_id` – ChEBI ID corresponding to the given Pubchem ID. If the lookup fails, None is returned.

**Return type** `str`

`indra.databases.chebi_client.get_pubchem_id` (*chebi\_id*)

Return the PubChem ID corresponding to a given ChEBI ID.

**Parameters** `chebi_id` (*str*) – ChEBI ID to be converted.

**Returns** `pubchem_id` – PubChem ID corresponding to the given ChEBI ID. If the lookup fails, None is returned.

**Return type** str

#### 4.3.4 BioGRID client (`indra.databases.biogrid_client`)

`indra.databases.biogrid_client.get_publications` (*gene\_names*, *save\_json\_name=None*)

Return evidence publications for interaction between the given genes.

**Parameters**

- **gene\_names** (*list[str]*) – A list of gene names (HGNC symbols) to query interactions between. Currently supports exactly two genes only.
- **save\_json\_name** (*Optional[str]*) – A file name to save the raw BioGRID web service output in. By default, the raw output is not saved.

**Returns** **publications** – A list of Publication objects that provide evidence for interactions between the given list of genes.

**Return type** list[Publication]

#### 4.3.5 Cell type context client (`indra.databases.context_client`)

`indra.databases.context_client.get_mutations` (*gene\_names*, *cell\_types*)

Return protein amino acid changes in given genes and cell types.

**Parameters**

- **gene\_names** (*list*) – HGNC gene symbols for which mutations are queried.
- **cell\_types** (*list*) – List of cell type names in which mutations are queried. The cell type names follow the CCLE database conventions.

Example: LOXIMVI\_SKIN, BT20\_BREAST

**Returns** **res** – A dictionary keyed by cell line, which contains another dictionary that is keyed by gene name, with a list of amino acid substitutions as values.

**Return type** dict[dict[list]]

`indra.databases.context_client.get_protein_expression` (*gene\_names*, *cell\_types*)

Return the protein expression levels of genes in cell types.

**Parameters**

- **gene\_names** (*list*) – HGNC gene symbols for which expression levels are queried.
- **cell\_types** (*list*) – List of cell type names in which expression levels are queried. The cell type names follow the CCLE database conventions.

Example: LOXIMVI\_SKIN, BT20\_BREAST

**Returns** **res** – A dictionary keyed by cell line, which contains another dictionary that is keyed by gene name, with estimated protein amounts as values.

**Return type** dict[dict[float]]

#### 4.3.6 Network relevance client (`indra.databases.relevance_client`)

`indra.databases.relevance_client.get_heat_kernel` (*network\_id*)

Return the identifier of a heat kernel calculated for a given network.

**Parameters** `network_id` (*str*) – The UUID of the network in NDEx.

**Returns** `kernel_id` – The identifier of the heat kernel calculated for the given network.

**Return type** `str`

`indra.databases.relevance_client.get_relevant_nodes` (*network\_id*, *query\_nodes*)

Return a set of network nodes relevant to a given query set.

A heat diffusion algorithm is used on a pre-computed heat kernel for the given network which starts from the given query nodes. The nodes in the network are ranked according to heat score which is a measure of relevance with respect to the query nodes.

**Parameters**

- **network\_id** (*str*) – The UUID of the network in NDEx.
- **query\_nodes** (*list[str]*) – A list of node names with respect to which relevance is queried.

**Returns** `ranked_entities` – A list containing pairs of node names and their relevance scores.

**Return type** `list[(str, float)]`

### 4.3.7 NDEx client (`indra.databases.ndex_client`)

`indra.databases.ndex_client.create_network` (*cx\_str*, *ndex\_cred*)

Creates a new NDEx network of the assembled CX model.

To upload the assembled CX model to NDEx, you need to have a registered account on NDEx (<http://ndexbio.org/>) and have the `ndex` python package installed. The uploaded network is private by default.

**Parameters** `ndex_cred` (*dict*) – A dictionary with the following entries: ‘user’: NDEx user name ‘password’: NDEx password

**Returns** `network_id` – The UUID of the NDEx network that was created by uploading the assembled CX model.

**Return type** `str`

`indra.databases.ndex_client.send_request` (*ndex\_service\_url*, *params*, *is\_json=True*, *use\_get=False*)

Send a request to the NDEx server.

**Parameters**

- **ndex\_service\_url** (*str*) – The URL of the service to use for the request.
- **params** (*dict*) – A dictionary of parameters to send with the request. Parameter keys differ based on the type of request.
- **is\_json** (*bool*) – True if the response is in json format, otherwise it is assumed to be text. Default: False
- **use\_get** (*bool*) – True if the request needs to use GET instead of POST.

**Returns** `res` – Depending on the type of service and the `is_json` parameter, this function either returns a text string or a json dict.

**Return type** `str`

`indra.databases.ndex_client.update_network` (*cx\_str*, *network\_id*, *ndex\_cred*)

Update an existing CX network on NDEx with new CX content.

**Parameters**

- **cx\_str** (*str*) – String containing the CX content.
- **network\_id** (*str*) – UUID of the network on NDEx.
- **ndex\_cred** (*dict*) – A dictionary with the following entries: ‘user’: NDEx user name ‘password’: NDEx password

### 4.3.8 cBio portal client (`indra.databases.cbio_client`)

`indra.databases.cbio_client.get_cancer_studies` (*study\_filter=None*)

Return a list of cancer study identifiers, optionally filtered.

There are typically multiple studies for a given type of cancer and a filter can be used to constrain the returned list.

**Parameters** **study\_filter** (*Optional[str]*) – A string used to filter the study IDs to return.

Example: “paad”

**Returns** **study\_ids** – A list of study IDs. For instance “paad” as a filter would result in a list of study IDs with paad in their name like “paad\_icgc”, “paad\_tcga”, etc.

**Return type** list[str]

`indra.databases.cbio_client.get_cancer_types` (*cancer\_filter=None*)

Return a list of cancer types, optionally filtered.

**Parameters** **cancer\_filter** (*Optional[str]*) – A string used to filter cancer types. Its value is the name or part of the name of a type of cancer. Example: “melanoma”, “pancreatic”, “non-small cell lung”

**Returns** **type\_ids** – A list of cancer types matching the filter. Example: for `cancer_filter=“pancreatic”`, the result includes “panet” (neuro-endocrine) and “paad” (adenocarcinoma)

**Return type** list[str]

`indra.databases.cbio_client.get_case_lists` (*study\_id*)

Return a list of the case set ids for a particular study.

TAKE NOTE the “case\_list\_id” are the same thing as “case\_set\_id” Within the data, this string is referred to as a “case\_list\_id”. Within API calls it is referred to as a ‘case\_set\_id’. The documentation does not make this explicitly clear.

**Parameters** **study\_id** (*str*) – The ID of the cBio study. Example: ‘cellline\_ccle\_broad’ or ‘paad\_icgc’

**Returns** **case\_set\_ids** – A dict keyed to cases containing a dict keyed to genes containing int

**Return type** dict[dict[int]]

`indra.databases.cbio_client.get_ccle_cna` (*gene\_list, cell\_lines*)

Return a dict of CNAs in given genes and cell lines from CCLE.

CNA values correspond to the following alterations -2 = homozygous deletion -1 = hemizygous deletion

0 = neutral / no change 1 = gain 2 = high level amplification

**Parameters**

- **gene\_list** (*list[str]*) – A list of HGNC gene symbols to get mutations in
- **cell\_lines** (*list[str]*) – A list of CCLE cell line names to get mutations for.

**Returns** **profile\_data** – A dict keyed to cases containing a dict keyed to genes containing int

**Return type** dict[dict[int]]

```
indra.databases.cbio_client.get_ccle_lines_for_mutation(gene,
                                                    amino_acid_change)
```

Return cell lines with a given point mutation in a given gene.

Checks which cell lines in CCLE have a particular point mutation in a given gene and return their names in a list.

**Parameters**

- **gene** (*str*) – The HGNC symbol of the mutated gene in whose product the amino acid change occurs. Example: “BRAF”
- **amino\_acid\_change** (*str*) – The amino acid change of interest. Example: “V600E”

**Returns** **cell\_lines** – A list of CCLE cell lines in which the given mutation occurs.

**Return type** list

```
indra.databases.cbio_client.get_ccle_mrna(gene_list, cell_lines)
```

Return a dict of mRNA amounts in given genes and cell lines from CCLE.

**Parameters**

- **gene\_list** (*list[str]*) – A list of HGNC gene symbols to get mRNA amounts for.
- **cell\_lines** (*list[str]*) – A list of CCLE cell line names to get mRNA amounts for.

**Returns** **mrna\_amounts** – A dict keyed to cell lines containing a dict keyed to genes containing float

**Return type** dict[dict[float]]

```
indra.databases.cbio_client.get_ccle_mutations(gene_list, cell_lines, mutation_type=None)
```

Return a dict of mutations in given genes and cell lines from CCLE.

This is a specialized call to get\_mutations tailored to CCLE cell lines.

**Parameters**

- **gene\_list** (*list[str]*) – A list of HGNC gene symbols to get mutations in
- **cell\_lines** (*list[str]*) – A list of CCLE cell line names to get mutations for.
- **mutation\_type** (*Optional[str]*) – The type of mutation to filter to. *mutation\_type* can be one of: *missense*, *nonsense*, *frame\_shift\_ins*, *frame\_shift\_del*, *splice\_site*

**Returns**

**mutations** – The result from cBioPortal as a dict in the format {*cell\_line* : {*gene* : [*mutation1*, *mutation2*, ...] }}

Example: {‘LOXIMVI\_SKIN’: {‘BRAF’: [‘V600E’, ‘I208V’]}, ‘SKMEL30\_SKIN’: {‘BRAF’: [‘D287H’, ‘E275K’]}}

**Return type** dict

```
indra.databases.cbio_client.get_genetic_profiles(study_id, profile_filter=None)
```

Return all the genetic profiles (data sets) for a given study.

Genetic profiles are different types of data for a given study. For instance the study ‘*cellline\_ccle\_broad*’ has profiles such as ‘*cellline\_ccle\_broad\_mutations*’ for mutations, ‘*cellline\_ccle\_broad\_CNA*’ for copy number alterations, etc.

**Parameters**

- **study\_id** (*str*) – The ID of the cBio study. Example: ‘paad\_icgc’
- **profile\_filter** (*Optional[str]*) – A string used to filter the profiles to return. Will be one of: - MUTATION - MUTATION\_EXTENDED - COPY\_NUMBER\_ALTERATION - MRNA\_EXPRESSION - METHYLATION The genetic profiles can include “mutation”, “CNA”, “rppa”, “methylation”, etc.

**Returns** **genetic\_profiles** – A list of genetic profiles available for the given study.

**Return type** list[str]

```
indra.databases.cbio_client.get_mutations(study_id, gene_list, mutation_type=None,
                                         case_id=None)
```

Return mutations as a list of genes and list of amino acid changes.

**Parameters**

- **study\_id** (*str*) – The ID of the cBio study. Example: ‘cellline\_ccle\_broad’ or ‘paad\_icgc’
- **gene\_list** (*list[str]*) – A list of genes with their HGNC symbols. Example: [‘BRAF’, ‘KRAS’]
- **mutation\_type** (*Optional[str]*) – The type of mutation to filter to. mutation\_type can be one of: missense, nonsense, frame\_shift\_ins, frame\_shift\_del, splice\_site
- **case\_id** (*Optional[str]*) – The case ID within the study to filter to.

**Returns** **mutations** – A tuple of two lists, the first one containing a list of genes, and the second one a list of amino acid changes in those genes.

**Return type** tuple[list]

```
indra.databases.cbio_client.get_num_sequenced(study_id)
```

Return number of sequenced tumors for given study.

This is useful for calculating mutation statistics in terms of the prevalence of certain mutations within a type of cancer.

**Parameters** **study\_id** (*str*) – The ID of the cBio study. Example: ‘paad\_icgc’

**Returns** **num\_case** – The number of sequenced tumors in the given study

**Return type** int

```
indra.databases.cbio_client.get_profile_data(study_id, gene_list, profile_filter,
                                             case_set_filter=None)
```

Return dict of cases and genes and their respective values.

**Parameters**

- **study\_id** (*str*) – The ID of the cBio study. Example: ‘cellline\_ccle\_broad’ or ‘paad\_icgc’
- **gene\_list** (*list[str]*) – A list of genes with their HGNC symbols. Example: [‘BRAF’, ‘KRAS’]
- **profile\_filter** (*str*) – A string used to filter the profiles to return. Will be one of: - MUTATION - MUTATION\_EXTENDED - COPY\_NUMBER\_ALTERATION - MRNA\_EXPRESSION - METHYLATION
- **case\_set\_filter** (*Optional[str]*) – A string that specifies which case\_set\_id to use, based on a complete or partial match. If not provided, will look for study\_id + ‘\_all’

**Returns** **profile\_data** – A dict keyed to cases containing a dict keyed to genes containing int



**Return type** dict[dict[int]]

`indra.databases.cbio_client.send_request`

Return a data frame from a web service request to cBio portal.

Sends a web service request to the cBio portal with arguments given in the dictionary data and returns a Pandas data frame on success.

More information about the service here: [http://www.cbioportal.org/web\\_api.jsp](http://www.cbioportal.org/web_api.jsp)

**Parameters** **kwargs** (*dict*) – A dict of parameters for the query. Entries map directly to web service calls with the exception of the optional ‘skiprows’ entry, whose value is used as the number of rows to skip when reading the result data frame.

**Returns** **df** – Response from cBioPortal as a Pandas DataFrame.

**Return type** pandas.DataFrame

## 4.4 Literature clients (`indra.literature`)

`indra.literature.get_full_text` (*paper\_id, idtype, preferred\_content\_type='text/xml'*)

Return the content and the content type of an article.

This function retrieves the content of an article by its PubMed ID, PubMed Central ID, or DOI. It prioritizes full text content when available and returns an abstract from PubMed as a fallback.

### Parameters

- **paper\_id** (*string*) – ID of the article.
- **idtype** (*'pmid', 'pmcid', or 'doi'*) – Type of the ID.
- **preferred\_content\_type** (*Optional[str]*) – Preference for full-text format, if available. Can be one of ‘text/xml’, ‘text/plain’, ‘application/pdf’. Default: ‘text/xml’

### Returns

- **content** (*str*) – The content of the article.
- **content\_type** (*str*) – The content type of the article

`indra.literature.id_lookup` (*paper\_id, idtype*)

Take an ID of type PMID, PMCID, or DOI and lookup the other IDs.

If the DOI is not found in Pubmed, try to obtain the DOI by doing a reverse-lookup of the DOI in CrossRef using article metadata.

### Parameters

- **paper\_id** (*string*) – ID of the article.
- **idtype** (*'pmid', 'pmcid', or 'doi'*) – Type of the ID.

**Returns** **ids** – A dictionary with the following keys: pmid, pmcid and doi.

**Return type** dict

### 4.4.1 Pubmed client (`indra.literature.pubmed_client`)

Search and get metadata for articles in Pubmed.

`indra.literature.pubmed_client.expand_pagination(pages)`

Convert a page number to long form, e.g., from 456-7 to 456-457.

`indra.literature.pubmed_client.get_abstract(pubmed_id, prepend_title=True)`

Get the abstract of an article in the Pubmed database.

`indra.literature.pubmed_client.get_article_xml`

Get the XML metadata for a single article from the Pubmed database.

`indra.literature.pubmed_client.get_ids`

Search Pubmed for paper IDs given a search term.

The options are passed as named arguments. For details on parameters that can be used, see <https://www.ncbi.nlm.nih.gov/books/NBK25499/#chapter4.ESearch> Some useful parameters to pass are `db='pmc'` to search PMC instead of pubmed `reldate=2` to search for papers within the last 2 days `mindate='2016/03/01'`, `max-date='2016/03/31'` to search for papers in March 2016.

`indra.literature.pubmed_client.get_ids_for_gene`

Get the curated set of articles for a gene in the Entrez database.

Search parameters for the Gene database query can be passed in as keyword arguments.

**Parameters** `hgnc_name` (*string*) – The HGNC name of the gene. This is used to obtain the HGNC ID (using the `hgnc_client` module) and in turn used to obtain the Entrez ID associated with the gene. Entrez is then queried for that ID.

`indra.literature.pubmed_client.get_issns_for_journal`

Get a list of the ISSN numbers for a journal given its NLM ID.

Information on NLM XML DTDs is available at <https://www.nlm.nih.gov/databases/dtd/>

`indra.literature.pubmed_client.get_metadata_for_ids(pmid_list, get_issns_from_nlm=False, get_abstracts=False, prepend_title=False)`

Get article metadata for up to 200 PMIDs from the Pubmed database.

**Parameters**

- `pmid_list` (*list of PMIDs as strings*) – Can contain 1-200 PMIDs.
- `get_issns_from_nlm` (*boolean*) – Look up the full list of ISSN number for the journal associated with the article, which helps to match articles to CrossRef search results. Defaults to False, since it slows down performance.
- `get_abstracts` (*boolean*) – Indicates whether to include the Pubmed abstract in the results.
- `prepend_title` (*boolean*) – If `get_abstracts` is True, specifies whether the article title should be prepended to the abstract text.

**Returns** Dictionary indexed by PMID. Each value is a dict containing the following fields: 'doi', 'title', 'authors', 'journal\_title', 'journal\_abbrev', 'journal\_nlm\_id', 'issn\_list', 'page'.

**Return type** dict of dicts

`indra.literature.pubmed_client.get_metadata_from_xml_tree(tree, get_issns_from_nlm=False, get_abstracts=False, prepend_title=False)`

Get metadata for an XML tree containing PubmedArticle elements.

**Documentation on the XML structure can be found at:**

- [https://www.nlm.nih.gov/bsd/licensee/elements\\_descriptions.html](https://www.nlm.nih.gov/bsd/licensee/elements_descriptions.html)
- [https://www.nlm.nih.gov/bsd/licensee/elements\\_alphabetical.html](https://www.nlm.nih.gov/bsd/licensee/elements_alphabetical.html)

#### Parameters

- **tree** (*xml.etree.ElementTree*) – ElementTree containing one or more PubmedArticle elements.
- **get\_issns\_from\_nlm** (*boolean*) – Look up the full list of ISSN number for the journal associated with the article, which helps to match articles to CrossRef search results. Defaults to False, since it slows down performance.
- **get\_abstracts** (*boolean*) – Indicates whether to include the Pubmed abstract in the results.
- **prepend\_title** (*boolean*) – If get\_abstracts is True, specifies whether the article title should be prepended to the abstract text.

**Returns** Dictionary indexed by PMID. Each value is a dict containing the following fields: ‘doi’, ‘title’, ‘authors’, ‘journal\_title’, ‘journal\_abbrev’, ‘journal\_nlm\_id’, ‘issn\_list’, ‘page’.

**Return type** dict of dicts

```
indra.literature.pubmed_client.get_title(pubmed_id)
    Get the title of an article in the Pubmed database.
```

#### 4.4.2 Pubmed Central client (`indra.literature.pmc_client`)

```
indra.literature.pmc_client.filter_pmids(pmid_list, source_type)
    Filter a list of PMIDs for ones with full text from PMC.
```

##### Parameters

- **pmid\_list** (*list*) – List of PMIDs to filter.
- **source\_type** (*string*) – One of ‘fulltext’, ‘oa\_xml’, ‘oa\_txt’, or ‘auth\_xml’.

##### Returns

**Return type** list of PMIDs available in the specified source/format type.

```
indra.literature.pmc_client.id_lookup(paper_id, idtype=None)
    This function takes a Pubmed ID, Pubmed Central ID, or DOI and use the Pubmed ID mapping service and looks up all other IDs from one of these. The IDs are returned in a dictionary.
```

#### 4.4.3 CrossRef client (`indra.literature.crossref_client`)

```
indra.literature.crossref_client.doi_query(pmid, search_limit=10)
    Get the DOI for a PMID by matching CrossRef and Pubmed metadata.
```

Searches CrossRef using the article title and then accepts search hits only if they have a matching journal ISSN and page number with what is obtained from the Pubmed database.

```
indra.literature.crossref_client.get_fulltext_links(doi)
    Return a list of links to the full text of an article given its DOI. Each list entry is a dictionary with keys: - URL: the URL to the full text - content-type: e.g. text/xml or text/plain - content-version - intended-application: e.g. text-mining
```

`indra.literature.crossref_client.get_metadata`

Returns the metadata of an article given its DOI from CrossRef as a JSON dict

## 4.4.4 Elsevier client (`indra.literature.elsevier_client`)

For information on the Elsevier API, see:

- API Specification: [http://dev.elsevier.com/api\\_docs.html](http://dev.elsevier.com/api_docs.html)
- Authentication: [https://dev.elsevier.com/tecdoc\\_api\\_authentication.html](https://dev.elsevier.com/tecdoc_api_authentication.html)

`indra.literature.elsevier_client.download_article` (*doi*)

Download an article in XML format from Elsevier.

`indra.literature.elsevier_client.get_abstract` (*doi*)

Get the abstract of an article from Elsevier.

`indra.literature.elsevier_client.get_article` (*doi*, *output='txt'*)

Get the full body of an article from Elsevier. There are two output modes: 'txt' strips all xml tags and joins the pieces of text in the main text, while 'xml' simply takes the tag containing the body of the article and returns it as is. In the latter case, downstream code needs to be able to interpret Elsevier's XML format.

`indra.literature.elsevier_client.get_dois`

Search ScienceDirect through the API for articles.

See <http://api.elsevier.com/content/search/fields/scidir> for constructing a query string to pass here. Example: 'abstract(BRAF) AND all("colorectal cancer")'

## 4.5 Preassembly (`indra.preassembler`)

### 4.5.1 Preassembler (`indra.preassembler`)

`class indra.preassembler.Preassembler` (*hierarchies*, *stmts=None*)

De-duplicates statements and arranges them in a specificity hierarchy.

#### Parameters

- **hierarchies** (dict[`indra.preassembler.hierarchy_manager`]) – A dictionary of hierarchies with keys such as 'entity' (hierarchy of entities, primarily specifying relationships between genes and their families) and 'modification' pointing to Hierarchy-Managers
- **stmts** (list of `indra.statements.Statement` or None) – A set of statements to perform pre-assembly on. If None, statements should be added using the `add_statements()` method.

#### **stmts**

list of `indra.statements.Statement` – Starting set of statements for preassembly.

#### **unique\_stmts**

list of `indra.statements.Statement` – Statements resulting from combining duplicates.

#### **related\_stmts**

list of `indra.statements.Statement` – Top-level statements after building the refinement hierarchy.

**hierarchies**

dict[*indra.preassembler.hierarchy\_manager*] – A dictionary of hierarchies with keys such as ‘entity’ and ‘modification’ pointing to HierarchyManagers

**add\_statements** (*stmts*)

Add to the current list of statements.

**Parameters** *stmts* (list of *indra.statements.Statement*) – Statements to add to the current list.

**static combine\_duplicate\_stmts** (*stmts*)

Combine evidence from duplicate Statements.

Statements are deemed to be duplicates if they have the same key returned by the *matches\_key()* method of the Statement class. This generally means that statements must be identical in terms of their arguments and can differ only in their associated *Evidence* objects.

This function keeps the first instance of each set of duplicate statements and merges the lists of Evidence from all of the other statements.

**Parameters** *stmts* (list of *indra.statements.Statement*) – Set of statements to de-duplicate.

**Returns** Unique statements with accumulated evidence across duplicates.

**Return type** list of *indra.statements.Statement*

**Examples**

De-duplicate and combine evidence for two statements differing only in their evidence lists:

```
>>> map2k1 = Agent('MAP2K1')
>>> mapk1 = Agent('MAPK1')
>>> stmt1 = Phosphorylation(map2k1, mapk1, 'T', '185',
... evidence=[Evidence(text='evidence 1')])
>>> stmt2 = Phosphorylation(map2k1, mapk1, 'T', '185',
... evidence=[Evidence(text='evidence 2')])
>>> uniq_stmts = Preassembler.combine_duplicate_stmts([stmt1, stmt2])
>>> uniq_stmts
[Phosphorylation(MAP2K1(), MAPK1(), T, 185)]
>>> sorted([e.text for e in uniq_stmts[0].evidence])
['evidence 1', 'evidence 2']
```

**combine\_duplicates** ()

Combine duplicates among *stmts* and save result in *unique\_stmts*.

A wrapper around the static method *combine\_duplicate\_stmts()*.

**combine\_related** (*return\_toplevel=True, poolsize=None, size\_cutoff=100*)

Connect related statements based on their refinement relationships.

This function takes as a starting point the unique statements (with duplicates removed) and returns a modified flat list of statements containing only those statements which do not represent a refinement of other existing statements. In other words, the more general versions of a given statement do not appear at the top level, but instead are listed in the *supports* field of the top-level statements.

If *unique\_stmts* has not been initialized with the de-duplicated statements, *combine\_duplicates()* is called internally.

After this function is called the attribute *related\_stmts* is set as a side-effect.

The procedure for combining statements in this way involves a series of steps:

1. The statements are grouped by type (e.g., Phosphorylation) and each type is iterated over independently.
2. Statements of the same type are then grouped according to their Agents' entity hierarchy component identifiers. For instance, ERK, MAPK1 and MAPK3 are all in the same connected component in the entity hierarchy and therefore all Statements of the same type referencing these entities will be grouped. This grouping assures that relations are only possible within Statement groups and not among groups. For two Statements to be in the same group at this step, the Statements must be the same type and the Agents at each position in the Agent lists must either be in the same hierarchy component, or if they are not in the hierarchy, must have identical `entity_matches_keys`. Statements with None in one of the Agent list positions are collected separately at this stage.
3. Statements with None at either the first or second position are iterated over. For a statement with a None as the first Agent, the second Agent is examined; then the Statement with None is added to all Statement groups with a corresponding component or `entity_matches_key` in the second position. The same procedure is performed for Statements with None at the second Agent position.
4. The statements within each group are then compared; if one statement represents a refinement of the other (as defined by the `refinement_off()` method implemented for the Statement), then the more refined statement is added to the `supports` field of the more general statement, and the more general statement is added to the `supported_by` field of the more refined statement.
5. A new flat list of statements is created that contains only those statements that have no `supports` entries (statements containing such entries are not eliminated, because they will be retrievable from the `supported_by` fields of other statements). This list is returned to the caller.

On multi-core machines, the algorithm can be parallelized by setting the `poolsize` argument to the desired number of worker processes. This feature is only available in Python > 3.4.

---

**Note:** Subfamily relationships must be consistent across arguments

For now, we require that merges can only occur if the *isa* relationships are all in the *same direction for all the agents* in a Statement. For example, the two statement groups: *RAF\_family -> MEK1* and *BRAF -> MEK\_family* would not be merged, since BRAF *isa* RAF\_family, but MEK\_family is not a MEK1. In the future this restriction could be revisited.

---

### Parameters

- **return\_toplevel** (*Optional[bool]*) – If True only the top level statements are returned. If False, all statements are returned. Default: True
- **poolsize** (*Optional[int]*) – The number of worker processes to use to parallelize the comparisons performed by the function. If None (default), no parallelization is performed. NOTE: Parallelization is only available on Python 3.4 and above.
- **size\_cutoff** (*Optional[int]*) – Groups with `size_cutoff` or more statements are sent to worker processes, while smaller groups are compared in the parent process. Default value is 100. Not relevant when parallelization is not used.

**Returns** The returned list contains Statements representing the more concrete/refined versions of the Statements involving particular entities. The attribute `related_stmts` is also set to this list. However, if `return_toplevel` is False then all statements are returned, irrespective of level of specificity. In this case the relationships between statements can be accessed via the `supports/supported_by` attributes.

**Return type** list of `indra.statement.Statement`

## Examples

A more general statement with no information about a Phosphorylation site is identified as supporting a more specific statement:

```
>>> from indra.preassembler.hierarchy_manager import hierarchies
>>> braf = Agent('BRAF')
>>> map2k1 = Agent('MAP2K1')
>>> st1 = Phosphorylation(braf, map2k1)
>>> st2 = Phosphorylation(braf, map2k1, residue='S')
>>> pa = Preassembler(hierarchies, [st1, st2])
>>> combined_stmts = pa.combine_related()
>>> combined_stmts
[Phosphorylation(BRAF(), MAP2K1(), S)]
>>> combined_stmts[0].supported_by
[Phosphorylation(BRAF(), MAP2K1())]
>>> combined_stmts[0].supported_by[0].supports
[Phosphorylation(BRAF(), MAP2K1(), S)]
```

`indra.preassembler.flatten_evidence(stmts)`

Add evidence from *supporting* stmts to evidence for *supported* stmts.

**Parameters** `stmts` (list of `indra.statements.Statement`) – A list of top-level statements with associated supporting statements resulting from building a statement hierarchy with `combine_related()`.

**Returns** `stmts` – Statement hierarchy identical to the one passed, but with the evidence lists for each statement now containing all of the evidence associated with the statements they are supported by.

**Return type** list of `indra.statements.Statement`

## Examples

Flattening evidence adds the two pieces of evidence from the supporting statement to the evidence list of the top-level statement:

```
>>> from indra.preassembler.hierarchy_manager import hierarchies
>>> braf = Agent('BRAF')
>>> map2k1 = Agent('MAP2K1')
>>> st1 = Phosphorylation(braf, map2k1,
... evidence=[Evidence(text='foo'), Evidence(text='bar')])
>>> st2 = Phosphorylation(braf, map2k1, residue='S',
... evidence=[Evidence(text='baz'), Evidence(text='bak')])
>>> pa = Preassembler(hierarchies, [st1, st2])
>>> pa.combine_related()
[Phosphorylation(BRAF(), MAP2K1(), S)]
>>> [e.text for e in pa.related_stmts[0].evidence]
['baz', 'bak']
>>> flattened = flatten_evidence(pa.related_stmts)
>>> sorted([e.text for e in flattened[0].evidence])
['bak', 'bar', 'baz', 'foo']
```

`indra.preassembler.flatten_stmts(stmts)`

Return the full set of unique stmts in a pre-assembled stmt graph.

The flattened list of statements returned by this function can be compared to the original set of unique statements to make sure no statements have been lost during the preassembly process.

**Parameters** `stmts` (list of `indra.statements.Statement`) – A list of top-level statements with associated supporting statements resulting from building a statement hierarchy with `combine_related()`.

**Returns** `stmts` – List of all statements contained in the hierarchical statement graph.

**Return type** list of `indra.statements.Statement`

## Examples

Calling `combine_related()` on two statements results in one top-level statement; calling `flatten_stmts()` recovers both:

```
>>> from indra.preassembler.hierarchy_manager import hierarchies
>>> braf = Agent('BRAF')
>>> map2k1 = Agent('MAP2K1')
>>> st1 = Phosphorylation(braf, map2k1)
>>> st2 = Phosphorylation(braf, map2k1, residue='S')
>>> pa = Preassembler(hierarchies, [st1, st2])
>>> pa.combine_related()
[Phosphorylation(BRAF(), MAP2K1(), S)]
>>> flattened = flatten_stmts(pa.related_stmts)
>>> flattened.sort(key=lambda x: x.matches_key())
>>> flattened
[Phosphorylation(BRAF(), MAP2K1()), Phosphorylation(BRAF(), MAP2K1(), S)]
```

`indra.preassembler.render_stmt_graph` (*statements*, *agent\_style=None*)

Render the statement hierarchy as a pygraphviz graph.

### Parameters

- **stmts** (list of `indra.statements.Statement`) – A list of top-level statements with associated supporting statements resulting from building a statement hierarchy with `combine_related()`.
- **agent\_style** (*dict or None*) – Dict of attributes specifying the visual properties of nodes. If None, the following default attributes are used:

```
agent_style = {'color': 'lightgray', 'style': 'filled',
               'fontname': 'arial'}
```

**Returns** Pygraphviz graph with nodes representing statements and edges pointing from supported statements to supported\_by statements.

**Return type** `pygraphviz.AGraph`

## Examples

Pattern for getting statements and rendering as a Graphviz graph:

```
>>> from indra.preassembler.hierarchy_manager import hierarchies
>>> braf = Agent('BRAF')
>>> map2k1 = Agent('MAP2K1')
>>> st1 = Phosphorylation(braf, map2k1)
```



```

>>> st2 = Phosphorylation(braf, map2k1, residue='S')
>>> pa = Preassembler(hierarchies, [st1, st2])
>>> pa.combine_related()
[Phosphorylation(BRAF(), MAP2K1(), S)]
>>> graph = render_stmt_graph(pa.related_stmts)
>>> graph.write('example_graph.dot') # To make the DOT file
>>> graph.draw('example_graph.png', prog='dot') # To make an image

```

Resulting graph:



## 4.5.2 Entity grounding curation and mapping (`indra.preassembler.grounding_mapper`)

**class** `indra.preassembler.grounding_mapper.GroundingMapper` (*gm, agent\_map=None*)  
 Maps grounding of INDRA Agents based on a given grounding map.

**gm**

*dict* – The grounding map, a dictionary mapping strings (entity names) to a dictionary of database identifiers.

**agent\_map**

*Optional[dict]* – A dictionary mapping strings to grounded INDRA Agents with given state.

`indra.preassembler.grounding_mapper.protein_map_from_twg` (*twg*)

Build map of entity texts to validated protein grounding.

Looks at the grounding of the entity texts extracted from the statements and finds proteins where there is grounding to a human protein that maps to an HGNC name that is an exact match to the entity text. Returns a dict that can be used to update/expand the grounding map.

## 4.5.3 Site curation and mapping (`indra.preassembler.sitemapper`)

**class** `indra.preassembler.sitemapper.MappedStatement` (*original\_stmt, mapped\_mods, mapped\_stmt*)

Information about a Statement found to have invalid sites.

**Parameters**

- **original\_stmt** (*indra.statements.Statement*) – The statement prior to mapping.
- **mapped\_mods** (*list of tuples*) – A list of invalid sites, where each entry in the list has two elements: ((gene\_name, residue, position), mapped\_site). If the invalid position was not found in the site map, mapped\_site is None; otherwise it is a tuple consisting of (residue, position, comment).
- **mapped\_stmt** (*indra.statements.Statement*) – The statement after mapping. Note that if no information was found in the site map, it will be identical to the original statement.

**class** `indra.preassembler.sitemapper.SiteMapper` (*site\_map*)

Use curated site information to standardize modification sites in stmts.

**Parameters** `site_map` (dict (as returned by `load_site_map()`)) – A dict mapping tuples of the form (`gene`, `orig_res`, `orig_pos`) to a tuple of the form (`correct_res`, `correct_pos`, `comment`), where `gene` is the string name of the gene (canonicalized to HGNC); `orig_res` and `orig_pos` are the residue and position to be mapped; `correct_res` and `correct_pos` are the corrected residue and position, and `comment` is a string describing the reason for the mapping (species error, isoform error, wrong residue name, etc.).

## Examples

Fixing site errors on both the modification state of an agent (MAP2K1) and the target of a Phosphorylation statement (MAPK1):

```
>>> map2k1_phos = Agent('MAP2K1', db_refs={'UP':'Q02750'}, mods=[
... ModCondition('phosphorylation', 'S', '217'),
... ModCondition('phosphorylation', 'S', '221')])
>>> mapk1 = Agent('MAPK1', db_refs={'UP':'P28482'})
>>> stmt = Phosphorylation(map2k1_phos, mapk1, 'T', '183')
>>> (valid, mapped) = default_mapper.map_sites([stmt])
>>> valid
[]
>>> mapped
[
MappedStatement:
  original_stmt: Phosphorylation(MAP2K1(mods: (phosphorylation, S, 217),
↳(phosphorylation, S, 221)), MAPK1(), T, 183)
  mapped_mods: (('MAP2K1', 'S', '217'), ('S', '218', 'off by one'))
               (('MAP2K1', 'S', '221'), ('S', '222', 'off by one'))
               (('MAPK1', 'T', '183'), ('T', '185', 'off by two; mouse sequence
↳'))
  mapped_stmt: Phosphorylation(MAP2K1(mods: (phosphorylation, S, 218),
↳(phosphorylation, S, 222)), MAPK1(), T, 185)
]
>>> ms = mapped[0]
>>> ms.original_stmt
Phosphorylation(MAP2K1(mods: (phosphorylation, S, 217), (phosphorylation, S,
↳221)), MAPK1(), T, 183)
>>> ms.mapped_mods
[ (('MAP2K1', 'S', '217'), ('S', '218', 'off by one')), (('MAP2K1', 'S', '221'), (
↳'S', '222', 'off by one')), (('MAPK1', 'T', '183'), ('T', '185', 'off by two;
↳mouse sequence')) ]
>>> ms.mapped_stmt
Phosphorylation(MAP2K1(mods: (phosphorylation, S, 218), (phosphorylation, S,
↳222)), MAPK1(), T, 185)
```

**map\_sites** (`stmts`, `do_methionine_offset=True`, `do_orthology_mapping=True`,  
`do_isoform_mapping=True`)

Check a set of statements for invalid modification sites.

Statements are checked against Uniprot reference sequences to determine if residues referred to by post-translational modifications exist at the given positions.

If there is nothing amiss with a statement (modifications on any of the agents, modifications made in the statement, etc.), then the statement goes into the list of valid statements. If there is a problem with the statement, the offending modifications are looked up in the site map (`site_map`), and an instance of `MappedStatement` is added to the list of mapped statements.

### Parameters

- **stmts** (list of `indra.statement.Statement`) – The statements to check for site errors.
- **do\_methionine\_offset** (*boolean*) – Whether to check for off-by-one errors in site position (possibly) attributable to site numbering from mature proteins after cleavage of the initial methionine. If True, checks the reference sequence for a known modification at 1 site position greater than the given one; if there exists such a site, creates the mapping. Default is True.
- **do\_orthology\_mapping** (*boolean*) – Whether to check sequence positions for known modification sites in mouse or rat sequences (based on PhosphoSitePlus data). If a mouse/rat site is found that is linked to a site in the human reference sequence, a mapping is created. Default is True.
- **do\_isoform\_mapping** (*boolean*) – Whether to check sequence positions for known modifications in other human isoforms of the protein (based on PhosphoSitePlus data). If a site is found that is linked to a site in the human reference sequence, a mapping is created. Default is True.

**Returns** 2-tuple containing (valid\_statements, mapped\_statements). The first element of the tuple is a list valid statements (`indra.statement.Statement`) that were not found to contain any site errors. The second element of the tuple is a list of mapped statements (`MappedStatement`) with information on the incorrect sites and corresponding statements with correctly mapped sites.

**Return type** tuple

`indra.preassembler.sitemapper.default_mapper = <indra.preassembler.sitemapper.SiteMapper object>`  
A default instance of `SiteMapper` that contains the site information found in `resources/curated_site_map.csv`.

`indra.preassembler.sitemapper.load_site_map(path)`  
Load the modification site map from a file.

The site map file should be a comma-separated file with six columns:

```
Gene: HGNC gene name
OrigRes: Original (incorrect) residue
OrigPos: Original (incorrect) residue position
CorrectRes: The correct residue for the modification
CorrectPos: The correct residue position
Comment: Description of the reason for the error.
```

**Parameters** `path` (*string*) – Path to the tab-separated site map file.

**Returns** A dict mapping tuples of the form (`gene`, `orig_res`, `orig_pos`) to a tuple of the form (`correct_res`, `correct_pos`, `comment`), where `gene` is the string name of the gene (canonicalized to HGNC); `orig_res` and `orig_pos` are the residue and position to be mapped; `correct_res` and `correct_pos` are the corrected residue and position, and `comment` is a string describing the reason for the mapping (species error, isoform error, wrong residue name, etc.).

**Return type** dict

#### 4.5.4 Hierarchy manager (`indra.preassembler.hierarchy_manager`)

```
class indra.preassembler.hierarchy_manager.HierarchyManager(rdf_file,
                                                         build_closure=True,
                                                         uri_as_name=True)
```

Store hierarchical relationships between different types of entities.

Used to store, e.g., entity hierarchies (proteins and protein families) and modification hierarchies (serine phosphorylation vs. phosphorylation).

**Parameters**

- **rdf\_file** (*string*) – Path to the RDF file containing the hierarchy.
- **build\_closure** (*Optional[bool]*) – If True, the transitive closure of the hierarchy is generated up from to speed up processing. Default: True
- **uri\_as\_name** (*Optional[bool]*) – If True, entries are accessed directly by their URIs. If False entries are accessed by finding their name through the `hasName` relationship. Default: True

**graph**

instance of *rdflib.Graph* – The RDF graph containing the hierarchy.

**build\_transitive\_closures** ()

Build the transitive closures of the hierarchy.

This method constructs dictionaries which contain terms in the hierarchy as keys and either all the “isa+” or “partof+” related terms as values.

**find\_entity**

Get the entity that has the specified name (or synonym).

**Parameters** **x** (*string*) – Name or synonym for the target entity.

**get\_children** (*uri*)

Return all (not just immediate) children of a given entry.

**Parameters** **uri** (*str*) – The URI of the entry whose children are to be returned. See the `get_uri` method to construct this URI from a name space and id.

**get\_parents** (*uri, type='all'*)

Return parents of a given entry.

**Parameters**

- **uri** (*str*) – The URI of the entry whose parents are to be returned. See the `get_uri` method to construct this URI from a name space and id.
- **type** (*str*) – ‘all’: return all parents irrespective of level; ‘immediate’: return only the immediate parents; ‘top’: return only the highest level parents

**isa** (*ns1, id1, ns2, id2*)

Indicate whether one entity has an “isa” relationship to another.

**Parameters**

- **ns1** (*string*) – Namespace code for an entity.
- **id1** (*string*) – URI for an entity.
- **ns2** (*string*) – Namespace code for an entity.
- **id2** (*string*) – URI for an entity.

**Returns** True if t1 has an “isa” relationship with t2, either directly or through a series of intermediates; False otherwise.

**Return type** bool

**partof** (*ns1, id1, ns2, id2*)

Indicate whether one entity is physically part of another.

**Parameters**

- **ns1** (*string*) – Namespace code for an entity.
- **id1** (*string*) – URI for an entity.
- **ns2** (*string*) – Namespace code for an entity.
- **id2** (*string*) – URI for an entity.

**Returns** True if t1 has a “partof” relationship with t2, either directly or through a series of intermediates; False otherwise.

**Return type** bool

## 4.6 Belief Engine (`indra.belief`)

**class** `indra.belief.BeliefEngine` (*prior\_probs=None*)

Assigns beliefs to INDRA Statements based on supporting evidence.

**Parameters** **prior\_probs** (*Optional[dict[dict]]*) – A dictionary of prior probabilities used to override/extend the default ones. There are two types of prior probabilities: `rand` and `syst` corresponding to random error and systematic error rate for each knowledge source. The `prior_probs` dictionary has the general structure `{‘rand’: {‘s1’: pr1, ..., ‘sn’: prn}, ‘syst’: {‘s1’: ps1, ..., ‘sn’: psn}}` where ‘s1’ ... ‘sn’ are names of input sources and `pr1 ... prn` and `ps1 ... psn` are error probabilities. Examples: `{‘rand’: {‘some_source’: 0.1}}` sets the random error rate for `some_source` to 0.1; `{‘rand’: {‘’}}`

**prior\_probs**

*dict[dict]* – A dictionary of prior systematic and random error probabilities for each knowledge source.

**set\_hierarchy\_probs** (*statements*)

Sets hierarchical belief probabilities for a list of INDRA Statements.

The Statements are assumed to be in a hierarchical relation graph with the `supports` and `supported_by` attribute of each Statement object having been set. The hierarchical belief probability of each Statement is calculated based on its prior probability and the probabilities propagated from Statements supporting it in the hierarchy graph.

**Parameters** **statements** (*list[indra.statements.Statement]*) – A list of INDRA Statements whose belief scores are to be calculated. Each Statement object’s belief attribute is updated by this function.

**set\_linked\_probs** (*linked\_statements*)

Sets the belief probabilities for a list of linked INDRA Statements.

The list of `LinkedStatement` objects is assumed to come from the `MechanismLinker`. The belief probability of the inferred Statement is assigned the joint probability of its source Statements.

**Parameters** **linked\_statements** (*list[indra.mechlinker.LinkedStatement]*) – A list of INDRA `LinkedStatements` whose belief scores are to be calculated. The belief attribute of the inferred Statement in the `LinkedStatement` object is updated by this function.

**set\_prior\_probs** (*statements*)

Sets the prior belief probabilities for a list of INDRA Statements.

The Statements are assumed to be de-duplicated. In other words, each Statement in the list passed to this function is assumed to have a list of Evidence objects that support it. The prior probability of each Statement is calculated based on the number of Evidences it has and their sources.

**Parameters** `statements` (`list[indra.statements.Statement]`) – A list of INDRA Statements whose belief scores are to be calculated. Each Statement object’s belief attribute is updated by this function.

`indra.belief.sample_statements` (`stmts`, `seed=None`)

Return statements sampled according to belief.

Statements are sampled independently according to their belief scores. For instance, a Statement with a belief score of 0.7 will end up in the returned Statement list with probability 0.7.

#### Parameters

- **stmts** (`list[indra.statements.Statement]`) – A list of INDRA Statements to sample.
- **seed** (`Optional[int]`) – A seed for the random number generator used for sampling.

**Returns** `new_stmts` – A list of INDRA Statements that were chosen by random sampling according to their respective belief scores.

**Return type** `list[indra.statements.Statement]`

## 4.7 Mechanism Linker (`indra.mechlinker`)

**class** `indra.mechlinker.AgentState` (`agent`, `evidence=None`)

A class representing Agent state without identifying a specific Agent.

#### `bound_conditions`

`list[indra.statements.BoundCondition]`

#### `mods`

`list[indra.statements.ModCondition]`

#### `mutations`

`list[indra.statements.Mutation]`

#### `location`

`indra.statements.location`

**apply\_to** (`agent`)

Apply this object’s state to an Agent.

**Parameters** `agent` (`indra.statements.Agent`) – The agent to which the state should be applied

**class** `indra.mechlinker.BaseAgent` (`name`)

Represents all activity types and active forms of an Agent.

#### Parameters

- **name** (`str`) – The name of the BaseAgent
- **activity\_types** (`list[str]`) – A list of activity types that the Agent has
- **active\_states** (`dict`) – A dict of activity types and their associated Agent states
- **activity\_reductions** (`dict`) – A dict of activity types and the type they are reduced to by inference.

**class** `indra.mechlinker.BaseAgentSet`

Container for a set of BaseAgents.

This class wraps a dict of BaseAgent instance and can be used to get and set BaseAgents.

**get\_create\_base\_agent** (*agent*)

Return BaseAgent from an Agent, creating it if needed.

**Parameters** *agent* (`indra.statements.Agent`) –

**Returns** *base\_agent*

**Return type** `indra.mechlinker.BaseAgent`

**class** `indra.mechlinker.LinkedStatement` (*source\_stmts*, *inferred\_stmt*)

A tuple containing a list of source Statements and an inferred Statement.

The list of source Statements are the basis for the inferred Statement.

**Parameters**

- **source\_stmts** (`list[indra.statements.Statement]`) – A list of source Statements
- **inferred\_stmts** (`indra.statements.Statement`) – A Statement that was inferred from the source Statements.

**class** `indra.mechlinker.MechLinker` (*stmts=None*)

Rewrite the activation pattern of Statements and derive new Statements.

The mechanism linker (MechLinker) traverses a corpus of Statements and uses various inference steps to make the activity types and active forms consistent among Statements.

**add\_statements** (*stmts*)

Add statements to the MechLinker.

**Parameters** *stmts* (`list[indra.statements.Statement]`) – A list of Statements to add.

**gather\_explicit\_activities** ()

Aggregate all explicit activities and active forms of Agents.

This function iterates over `self.statements` and extracts explicitly stated activity types and active forms for Agents.

**gather\_implicit\_activities** ()

Aggregate all implicit activities and active forms of Agents.

Iterate over `self.statements` and collect the implied activities and active forms of Agents that appear in the Statements.

Note that using this function to collect implied Agent activities can be risky. Assume, for instance, that a Statement from a reading system states that EGF bound to EGFR phosphorylates ERK. This would be interpreted as implicit evidence for the EGFR-bound form of EGF to have ‘kinase’ activity, which is clearly incorrect.

In contrast the alternative pair of this function: `gather_explicit_activities` collects only explicitly stated activities.

**static infer\_activations** (*stmts*)

Return inferred RegulateActivity from Modification + ActiveForm.

This function looks for combinations of Modification and ActiveForm Statements and infers Activation/Inhibition Statements from them. For example, if we know that A phosphorylates B, and the phosphorylated form of B is active, then we can infer that A activates B. This can also be viewed as having “explained” a given Activation/Inhibition Statement with a combination of more mechanistic Modification + ActiveForm Statements.

**Parameters** `stmts` (`list[indra.statements.Statement]`) – A list of Statements to infer RegulateActivity from.

**Returns** `linked_stmts` – A list of LinkedStatements representing the inferred Statements.

**Return type** `list[indra.mechlinker.LinkedStatement]`

**static infer\_active\_forms** (`stmts`)

Return inferred ActiveForm from RegulateActivity + Modification.

This function looks for combinations of Activation/Inhibition Statements and Modification Statements, and infers an ActiveForm from them. For example, if we know that A activates B and A phosphorylates B, then we can infer that the phosphorylated form of B is active.

**Parameters** `stmts` (`list[indra.statements.Statement]`) – A list of Statements to infer ActiveForms from.

**Returns** `linked_stmts` – A list of LinkedStatements representing the inferred Statements.

**Return type** `list[indra.mechlinker.LinkedStatement]`

**static infer\_complexes** (`stmts`)

Return inferred Complex from Statements implying physical interaction.

**Parameters** `stmts` (`list[indra.statements.Statement]`) – A list of Statements to infer Complexes from.

**Returns** `linked_stmts` – A list of LinkedStatements representing the inferred Statements.

**Return type** `list[indra.mechlinker.LinkedStatement]`

**static infer\_modifications** (`stmts`)

Return inferred Modification from RegulateActivity + ActiveForm.

This function looks for combinations of Activation/Inhibition Statements and ActiveForm Statements that imply a Modification Statement. For example, if we know that A activates B, and phosphorylated B is active, then we can infer that A leads to the phosphorylation of B. An additional requirement when making this assumption is that the activity of B should only be dependent on the modified state and not other context - otherwise the inferred Modification is not necessarily warranted.

**Parameters** `stmts` (`list[indra.statements.Statement]`) – A list of Statements to infer Modifications from.

**Returns** `linked_stmts` – A list of LinkedStatements representing the inferred Statements.

**Return type** `list[indra.mechlinker.LinkedStatement]`

**reduce\_activities** ()

Rewrite the activity types referenced in Statements for consistency.

Activity types are reduced to the most specific form whenever possible. For instance, if 'kinase' is the only specific activity type known for the BaseAgent of BRAF, its generic 'activity' forms are rewritten to 'kinase'.

**replace\_activations** (`linked_stmts=None`)

Remove RegulateActivity Statements that can be inferred out.

This function iterates over self.statements and looks for RegulateActivity Statements that either match or are refined by inferred RegulateActivity Statements that were linked (provided as the linked\_stmts argument). It removes RegulateActivity Statements from self.statements that can be explained by the linked statements.

**Parameters** `linked_stmts` (`Optional[list[indra.mechlinker.LinkedStatement]]`) – A list of linked statements, optionally passed from outside. If



None is passed, the MechLinker runs `self.infer_activations` to infer `RegulateActivities` and obtain a list of `LinkedStatements` that are then used for removing existing Complexes in `self.statements`.

**replace\_complexes** (*linked\_stmts=None*)

Remove Complex Statements that can be inferred out.

This function iterates over `self.statements` and looks for Complex Statements that either match or are refined by inferred Complex Statements that were linked (provided as the `linked_stmts` argument). It removes Complex Statements from `self.statements` that can be explained by the linked statements.

**Parameters linked\_stmts** (*Optional[list[indra.mechlinker.LinkedStatement]]*) – A list of linked statements, optionally passed from outside. If None is passed, the MechLinker runs `self.infer_complexes` to infer Complexes and obtain a list of `LinkedStatements` that are then used for removing existing Complexes in `self.statements`.

**require\_active\_forms** ()

Rewrites Statements with Agents' active forms in active positions.

As an example, the enzyme in a Modification Statement can be expected to be in an active state. Similarly, subjects of `RegulateAmount` and `RegulateActivity` Statements can be expected to be in an active form. This function takes the collected active states of Agents in their corresponding `BaseAgents` and then rewrites other Statements to apply the active Agent states to them.

**Returns new\_stmts** – A list of Statements which includes the newly rewritten Statements. This list is also set as the internal Statement list of the MechLinker.

**Return type** `list[indra.statements.Statement]`

## 4.8 Assemblers of model output (`indra.assemblers`)

### 4.8.1 Executable PySB models (`indra.assemblers.pysb_assembler`)

**class** `indra.assemblers.pysb_assembler.PysbAssembler` (*policies=None*)

Assembler creating a PySB model from a set of INDRA Statements.

**Parameters policies** (*Optional[Union[str, dict]]*) – A string or dictionary that defines one or more assembly policies.

If `policies` is a string, it defines a global assembly policy that applies to all Statement types. Example: `contact_only, one_step`

A dictionary of policies has keys corresponding to Statement types and values to the policy to be applied to that type of Statement. For Statement types whose policy is undefined, the 'default' policy is applied. Example: `{'Phosphorylation': 'two_step'}`

**policies**

*dict* – A dictionary of policies that defines assembly policies for Statement types. It is assigned in the constructor.

**statements**

*list* – A list of INDRA statements to be assembled.

**model**

*pysb.Model* – A PySB model object that is assembled by this class.

**agent\_set**

*\_BaseAgentSet* – A set of `BaseAgents` used during the assembly process.

**add\_default\_initial\_conditions** (*value=None*)

Set default initial conditions in the PySB model.

**Parameters** **value** (*Optional[float]*) – Optionally a value can be supplied which will be the initial amount applied. Otherwise a built-in default is used.

**add\_statements** (*stmts*)

Add INDRA Statements to the assembler’s list of statements.

**Parameters** **stmts** (*list[indra.statements.Statement]*) – A list of *indra.statements.Statement* to be added to the statement list of the assembler.

**export\_model** (*format, file\_name=None*)

Save the assembled model in a modeling formalism other than PySB.

For more details on exporting PySB models, see <http://pysb.readthedocs.io/en/latest/modules/export/index.html>

#### Parameters

- **format** (*str*) – The format to export into, for instance “kappa”, “bngl”, “sbml”, “matlab”, “mathematica”, “potterswheel”. See <http://pysb.readthedocs.io/en/latest/modules/export/index.html> for a list of supported formats. In addition to the formats supported by PySB itself, this method also provides “sbn” output.
- **file\_name** (*Optional[str]*) – An optional file name to save the exported model into.

**Returns** **exp\_str** – The exported model string

**Return type** **str**

**make\_model** (*policies=None, initial\_conditions=True, reverse\_effects=False*)

Assemble the PySB model from the collected INDRA Statements.

This method assembles a PySB model from the set of INDRA Statements. The assembled model is both returned and set as the assembler’s model argument.

#### Parameters

- **policies** (*Optional[Union[str, dict]]*) – A string or dictionary of policies, as defined in *indra.assemblers.PysbAssembler*. This set of policies locally supersedes the default setting in the assembler. This is useful when this function is called multiple times with different policies.
- **initial\_conditions** (*Optional[bool]*) – If True, default initial conditions are generated for the Monomers in the model.

**Returns** **model** – The assembled PySB model object.

**Return type** **pysb.Model**

**print\_model** ()

Print the assembled model as a PySB program string.

This function is useful when the model needs to be passed as a string to another component.

**save\_model** (*file\_name='pysb\_model.py'*)

Save the assembled model as a PySB program file.

**Parameters** **file\_name** (*Optional[str]*) – The name of the file to save the model program code in. Default: *pysb-model.py*

**save\_rst** (*file\_name='pysb\_model.rst', module\_name='pysb\_module'*)

Save the assembled model as an RST file for literate modeling.

**Parameters**

- **file\_name** (*Optional[str]*) – The name of the file to save the RST in. Default: `pysb_model.rst`
- **module\_name** (*Optional[str]*) – The name of the python function defining the module. Default: `pysb_module`

**set\_context** (*cell\_type*)

Set protein expression data as initial conditions.

This method uses `indra.databases.context_client` to get protein expression levels for a given cell type and set initial conditions for Monomers in the model accordingly.

**Parameters**

- **cell\_type** (*str*) – Cell type name for which expression levels are queried. The cell type name follows the CCLE database conventions.
- **Example** (`LOXIMVI_SKIN`, `BT20_BREAST`) –

`indra.assemblers.pysb_assembler.add_rule_to_model(model, rule, annotations=None)`

Add a Rule to a PySB model and handle duplicate component errors.

`indra.assemblers.pysb_assembler.complex_monomers_default(stmt, agent_set)`

In this (very simple) implementation, proteins in a complex are each given site names corresponding to each of the other members of the complex (lower case). So the resulting complex can be “fully connected” in that each member can be bound to all the others.

`indra.assemblers.pysb_assembler.complex_monomers_one_step(stmt, agent_set)`

In this (very simple) implementation, proteins in a complex are each given site names corresponding to each of the other members of the complex (lower case). So the resulting complex can be “fully connected” in that each member can be bound to all the others.

`indra.assemblers.pysb_assembler.export_sbgn(model)`

Return an SBGN model string corresponding to the PySB model.

This function first calls `generate_equations` on the PySB model to obtain a reaction network (i.e. individual species, reactions). It then iterates over each reaction and and instantiates its reactants, products, and the process itself as SBGN glyphs and arcs.

**Parameters** `model` (*pysb.core.Model*) – A PySB model to be exported into SBGN

**Returns** `sbgn_str` – An SBGN model as string

**Return type** `str`

`indra.assemblers.pysb_assembler.get_agent_rule_str(agent)`

Construct a string from an Agent as part of a PySB rule name.

`indra.assemblers.pysb_assembler.get_annotation(component, db_name, db_ref)`

Construct model Annotations for each component.

Annotation formats follow guidelines at <http://identifiers.org/>.

`indra.assemblers.pysb_assembler.get_binding_site_name(agent)`

Return a binding site name from a given agent.

`indra.assemblers.pysb_assembler.get_create_parameter(model, name, value, unique=True)`

Return parameter with given name, creating it if needed.

If `unique` is false and the parameter exists, the value is not changed; if it does not exist, it will be created. If `unique` is true then upon conflict a number is added to the end of the parameter name.

`indra.assemblers.pysb_assembler.get_mod_site_name(mod_type, residue, position)`  
 Return site names for a modification.

`indra.assemblers.pysb_assembler.get_monomer_pattern(model, agent, extra_fields=None)`  
 Construct a PySB MonomerPattern from an Agent.

`indra.assemblers.pysb_assembler.get_site_pattern(agent)`  
 Construct a dictionary of Monomer site states from an Agent.

This crates the mapping to the associated PySB monomer from an INDRA Agent object.

`indra.assemblers.pysb_assembler.get_uncond_agent(agent)`  
 Construct the unconditional state of an Agent.

The unconditional Agent is a copy of the original agent but without any bound conditions and modification conditions. Mutation conditions, however, are preserved since they are static.

`indra.assemblers.pysb_assembler.grounded_monomer_patterns(model, agent)`  
 Get monomer patterns for the agent accounting for grounding information.

`indra.assemblers.pysb_assembler.parse_identifiers_url(url)`  
 Parse an identifiers.org URL into (namespace, ID) tuple.

`indra.assemblers.pysb_assembler.set_base_initial_condition(model, monomer, value)`  
 Set an initial condition for a monomer in its ‘default’ state.

`indra.assemblers.pysb_assembler.set_extended_initial_condition(model, monomer=None, value=0)`  
 Set an initial condition for monomers in “modified” state.

This is useful when using downstream analysis that relies on reactions being active in the model. One example is BioNetGen-based reaction network diagram generation.

## 4.8.2 Cytoscape networks (`indra.assemblers.cx_assembler`)

**class** `indra.assemblers.cx_assembler.CxAssembler(stmts=None, network_name=None)`  
 This class assembles a CX network from a set of INDRA Statements.

The CX format is an aspect oriented data mode for networks. The format is defined at <http://www.home.ndexbio.org/data-model/>. The CX format is the standard for NDEx and is compatible with CytoScape via the CyNDEx plugin.

### Parameters

- **stmts** (*Optional[list[indra.statements.Statement]]*) – A list of INDRA Statements to be assembled.
- **network\_name** (*Optional[str]*) – The name of the network to be assembled. Default: `indra_assembled`

### statements

*list[indra.statements.Statement]* – A list of INDRA Statements to be assembled.

### network\_name

*str* – The name of the network to be assembled.

### cx

*dict* – The structure of the CX network that is assembled.

**add\_statements** (*stmts*)

Add INDRA Statements to the assembler's list of statements.

**Parameters** *stmts* (*list[indra.statements.Statement]*) – A list of *indra.statements.Statement* to be added to the statement list of the assembler.

**make\_model** (*add\_indra\_json=True*)

Assemble the CX network from the collected INDRA Statements.

This method assembles a CX network from the set of INDRA Statements. The assembled network is set as the assembler's *cx* argument.

**Parameters** *add\_indra\_json* (*Optional[bool]*) – If True, the INDRA Statement JSON annotation is added to each edge in the network. Default: True

**Returns** *cx\_str* – The json serialized CX model.

**Return type** str

**print\_cx** (*pretty=True*)

Return the assembled CX network as a json string.

**Parameters** *pretty* (*bool*) – If True, the CX string is formatted with indentation (for human viewing) otherwise no indentation is used.

**Returns** *json\_str* – A json formatted string representation of the CX network.

**Return type** str

**save\_model** (*file\_name='model.cx'*)

Save the assembled CX network in a file.

**Parameters** *file\_name* (*Optional[str]*) – The name of the file to save the CX network to. Default: model.cx

**set\_context** (*cell\_type*)

Set protein expression data and mutational status as node attribute

This method uses *indra.databases.context\_client* to get protein expression levels and mutational status for a given cell type and set a node attribute for proteins accordingly.

**Parameters** *cell\_type* (*str*) – Cell type name for which expression levels are queried. The cell type name follows the CCLE database conventions. Example: LOXIMVI\_SKIN, BT20\_BREAST

**upload\_model** (*ndex\_cred*)

Creates a new NDEx network of the assembled CX model.

To upload the assembled CX model to NDEx, you need to have a registered account on NDEx (<http://ndexbio.org/>) and have the *ndex* python package installed. The uploaded network is private by default.

**Parameters** *ndex\_cred* (*dict*) – A dictionary with the following entries: 'user': NDEx user name 'password': NDEx password

**Returns** *network\_id* – The UUID of the NDEx network that was created by uploading the assembled CX model.

**Return type** str

### 4.8.3 Natural language (*indra.assemblers.english\_assembler*)

**class** *indra.assemblers.english\_assembler.EnglishAssembler* (*stmts=None*)

This assembler generates English sentences from INDRA Statements.

**Parameters** `stmts` (*Optional[list[indra.statements.Statement]]*) – A list of INDRA Statements to be added to the assembler.

**statements**

*list[indra.statements.Statement]* – A list of INDRA Statements to assemble.

**model**

*str* – The assembled sentences as a single string.

**add\_statements** (*stmts*)

Add INDRA Statements to the assembler’s list of statements.

**Parameters** `stmts` (*list[indra.statements.Statement]*) – A list of *indra.statements.Statement* to be added to the statement list of the assembler.

**make\_model** ()

Assemble text from the set of collected INDRA Statements.

**Returns** `stmt_strs` – Return the assembled text as unicode string. By default, the text is a single string consisting of one or more sentences with periods at the end.

**Return type** `str`

#### 4.8.4 Node-edge graphs (`indra.assemblers.graph_assembler`)

```
class indra.assemblers.graph_assembler.GraphAssembler (stmts=None,  
                                                         graph_properties=None,  
                                                         node_properties=None,  
                                                         edge_properties=None)
```

The Graph assembler assembles INDRA Statements into a Graphviz node-edge graph.

**Parameters**

- **stmts** (*Optional[list[indra.statements.Statement]]*) – A list of INDRA Statements to be added to the assembler’s list of Statements.
- **graph\_properties** (*Optional[dict[str: str]]*) – A dictionary of graphviz graph properties overriding the default ones.
- **node\_properties** (*Optional[dict[str: str]]*) – A dictionary of graphviz node properties overriding the default ones.
- **edge\_properties** (*Optional[dict[str: str]]*) – A dictionary of graphviz edge properties overriding the default ones.

**statements**

*list[indra.statements.Statement]* – A list of INDRA Statements to be assembled.

**graph**

*pygraphviz.AGraph* – A pygraphviz graph that is assembled by this assembler.

**existing\_nodes**

*list[tuple]* – The list of nodes (identified by node key tuples) that are already in the graph.

**existing\_edges**

*list[tuple]* – The list of edges (identified by edge key tuples) that are already in the graph.

**graph\_properties**

*dict[str: str]* – A dictionary of graphviz graph properties used for assembly.

**node\_properties**

*dict[str: str]* – A dictionary of graphviz node properties used for assembly.

**edge\_properties**

*dict[str: str]* – A dictionary of graphviz edge properties used for assembly. Note that most edge properties are determined based on the type of the edge by the assembler (e.g. color, arrowhead). These settings cannot be directly controlled through the API.

**add\_statements** (*stmts*)

Add a list of statements to be assembled.

**Parameters** *stmts* (*list[indra.statements.Statement]*) – A list of INDRA Statements to be appended to the assembler’s list.

**get\_string** ()

Return the assembled graph as a string.

**Returns** *graph\_string* – The assembled graph as a string.

**Return type** *str*

**make\_model** ()

Assemble the graph from the assembler’s list of INDRA Statements.

**save\_dot** (*file\_name='graph.dot'*)

Save the graph in a graphviz dot file.

**Parameters** *file\_name* (*Optional[str]*) – The name of the file to save the graph dot string to.

**save\_pdf** (*file\_name='graph.pdf', prog='dot'*)

Draw the graph and save as an image or pdf file.

**Parameters**

- **file\_name** (*Optional[str]*) – The name of the file to save the graph as. Default: graph.pdf
- **prog** (*Optional[str]*) – The graphviz program to use for graph layout. Default: dot

## 4.8.5 SIF / Boolean networks (*indra.assemblers.sif\_assembler*)

**class** *indra.assemblers.sif\_assembler.SifAssembler* (*stmts=None*)

The SIF assembler assembles INDRA Statements into a networkx graph.

This graph can then be exported into SIF (simple interaction format) or a Boolean network.

**Parameters** *stmts* (*Optional[list[indra.statements.Statement]]*) – A list of INDRA Statements to be added to the assembler’s list of Statements.

**graph**

*networkx.DiGraph* – A networkx graph that is assembled by this assembler.

**make\_model** (*use\_name\_as\_key=False, include\_mods=False, include\_complexes=False*)

Assemble the graph from the assembler’s list of INDRA Statements.

**Parameters**

- **use\_name\_as\_key** (*boolean*) – If True, uses the name of the agent as the key to the nodes in the network. If False (default) uses the matches\_key() of the agent.
- **include\_mods** (*boolean*) – If True, adds Modification statements into the graph as directed edges. Default is False.
- **include\_complexes** (*boolean*) – If True, creates two edges (in both directions) between all pairs of nodes in Complex statements. Default is False.

**print\_boolean\_net** (*out\_file=None*)

Return a Boolean network from the assembled graph.

See <https://github.com/ialbert/booleannet> for details about the format used to encode the Boolean rules.

**Parameters** **out\_file** (*Optional[str]*) – A file name in which the Boolean network is saved.

**Returns** **full\_str** – The string representing the Boolean network.

**Return type** str

**print\_loopy** (*as\_url=True*)

Return

**Parameters** **out\_file** (*Optional[str]*) – A file name in which the Loopy network is saved.

**Returns** **full\_str** – The string representing the Loopy network.

**Return type** str

**print\_model** (*include\_unsigned\_edges=False*)

Return a SIF string of the assembled model.

**Parameters** **include\_unsigned\_edges** (*bool*) – If True, includes edges with an unknown activating/inactivating relationship (e.g., most PTMs). Default is False.

**save\_model** (*fname*)

Save the assembled model's SIF string into a file.

**Parameters** **fname** (*str*) – The name of the file to save the SIF into.

## 4.8.6 MITRE “index cards” (`indra.assemblers.index_card_assembler`)

**class** `indra.assemblers.index_card_assembler.IndexCardAssembler` (*statements=None, pmc\_override=None*)

Assembler creating index cards from a set of INDRA Statements.

**Parameters**

- **statements** (*list*) – A list of INDRA statements to be assembled.
- **pmc\_override** (*Optional[str]*) – A PMC ID to assign to the index card.

**statements**

*list* – A list of INDRA statements to be assembled.

**add\_statements** (*statements*)

Add statements to the assembler.

**Parameters** **statements** (*list[indra.statement.Statements]*) – The list of Statements to add to the assembler.

**make\_model** ()

Assemble statements into index cards.

**print\_model** ()

Return the assembled cards as a JSON string.

**Returns** **cards\_json** – The JSON string representing the assembled cards.

**Return type** str



**save\_model** (*file\_name='index\_cards.json'*)

Save the assembled cards into a file.

**Parameters** **file\_name** (*Optional[str]*) – The name of the file to save the cards into.

Default: index\_cards.json

#### 4.8.7 SBGN output (`indra.assemblers.sbg_n_assembler`)

**class** `indra.assemblers.sbg_n_assembler.SBG_NAssembler` (*statements=None*)

This class assembles an SBGN model from a set of INDRA Statements.

The Systems Biology Graphical Notation (SBGN) is a widely used graphical notation standard for systems biology models. This assembler creates SBGN models following the Process Description (PD) standard, documented at: [https://github.com/sbgn/process-descriptions/blob/master/UserManual/sbgn\\_PD-level1-user-public.pdf](https://github.com/sbgn/process-descriptions/blob/master/UserManual/sbgn_PD-level1-user-public.pdf). For more information on SBGN, see: <http://sbgn.github.io/sbgn/>

**Parameters** **stmts** (*Optional[list[indra.statements.Statement]]*) – A list of INDRA Statements to be assembled.

**statements**

*list[indra.statements.Statement]* – A list of INDRA Statements to be assembled.

**sbg\_n**

*xml.etree.ElementTree* – The structure of the SBGN model that is assembled, represented as an XML ElementTree.

**add\_statements** (*stmts*)

Add INDRA Statements to the assembler's list of statements.

**Parameters** **stmts** (*list[indra.statements.Statement]*) – A list of *indra.statements.Statement* to be added to the statement list of the assembler.

**make\_model** ()

Assemble the SBGN model from the collected INDRA Statements.

This method assembles an SBGN model from the set of INDRA Statements. The assembled model is set as the assembler's `sbg_n` attribute (it is represented as an XML ElementTree internally). The model is returned as a serialized XML string.

**Returns** **sbg\_n\_str** – The XML serialized SBGN model.

**Return type** str

**print\_model** (*pretty=True, encoding='utf8'*)

Return the assembled SBGN model as an XML string.

**Parameters** **pretty** (*Optional[bool]*) – If True, the SBGN string is formatted with indentation (for human viewing) otherwise no indentation is used. Default: True

**Returns** **sbg\_n\_str** – An XML string representation of the SBGN model.

**Return type** bytes (str in Python 2)

**save\_model** (*file\_name='model.sbg\_n'*)

Save the assembled SBGN model in a file.

**Parameters** **file\_name** (*Optional[str]*) – The name of the file to save the SBGN network to. Default: model.sbg\_n

## 4.9 Explanation (`indra.explanation`)

### 4.9.1 Check whether a rule-based model satisfies a property (`indra.explanation.model_checker`)

`class indra.explanation.model_checker.ModelChecker` (*model*, *statements=None*,  
*agent\_obs=None*)

Check a PySB model against a set of INDRA statements.

#### Parameters

- **model** (*pysb.Model*) – A PySB model to check.
- **statements** (*Optional[list[indra.statements.Statement]]*) – A list of INDRA Statements to check the model against.
- **agent\_obs** (*Optional[list[indra.statements.Agent]]*) – A list of INDRA Agents in a given state to be observed.

`add_statements` (*stmts*)

Add to the list of statements to check against the model.

**Parameters** *stmts* (*list[indra.statements.Statement]*) – The list of Statements to be added for checking.

`check_model` (*max\_paths=1*, *max\_path\_length=5*)

Check all the statements added to the ModelChecker.

#### Parameters

- **max\_paths** (*Optional[int]*) – The maximum number of specific paths to return for each Statement to be explained. Default: 1
- **max\_path\_length** (*Optional[int]*) – The maximum length of specific paths to return. Default: 5

**Returns** Each tuple contains the Statement checked against the model and a PathResult object describing the results of model checking.

**Return type** list of (*Statement*, *PathResult*)

`check_statement` (*stmt*, *max\_paths=1*, *max\_path\_length=5*)

Check a single Statement against the model.

#### Parameters

- **stmt** (*indra.statements.Statement*) – The Statement to check.
- **max\_paths** (*Optional[int]*) – The maximum number of specific paths to return for each Statement to be explained. Default: 1
- **max\_path\_length** (*Optional[int]*) – The maximum length of specific paths to return. Default: 5

**Returns** True if the model satisfies the Statement.

**Return type** boolean

`get_im` (*force\_update=False*)

Get the influence map for the model, generating it if necessary.

**Parameters** `force_update` (*bool*) – Whether to generate the influence map when the function is called. If False, returns the previously generated influence map if available. Defaults to True.

### Returns

The influence map can be rendered as a pdf using the dot layout program as follows:

```
influence_map.draw('influence_map.pdf', prog='dot')
```

**Return type** pygraphviz AGraph object containing the influence map.

### `prune_influence_map()`

Remove edges between rules causing problematic non-transitivity.

First, all self-loops are removed. After this initial step, edges are removed between rules when they share *all* child nodes except for each other; that is, they have a mutual relationship with each other and share all of the same children.

Note that edges must be removed in batch at the end to prevent edge removal from affecting the lists of rule children during the comparison process.

### `score_paths(paths, agents_values, loss_of_function=False, sigma=0.15)`

Return scores associated with a given set of paths.

#### Parameters

- **paths** (*list[list[tuple[str, int]]]*) – A list of paths obtained from path finding. Each path is a list of tuples (which are edges in the path), with the first element of the tuple the name of a rule, and the second element its polarity in the path.
- **agents\_values** (*dict[indra.statements.Agent, float]*) – A dictionary of INDRA Agents and their corresponding measured value in a given experimental condition.
- **loss\_of\_function** (*boolean*) – If True, flip the polarity of the path. For instance, if the effect of an inhibitory drug is explained, set this to True. Default: False

**class** `indra.explanation.model_checker.PathMetric` (*source\_node, target\_node, polarity, length*)

Describes results of simple path search (path existence).

**class** `indra.explanation.model_checker.PathResult` (*path\_found, result\_code, max\_paths, max\_path\_length*)

Describes results of running the ModelChecker on a single Statement.

#### Parameters

- **path\_found** (*bool*) –
- **result\_code** (*string*) – STATEMENT\_TYPE\_NOT\_HANDLED  
SUBJECT\_MONOMERS\_NOT\_FOUND OBSERVABLES\_NOT\_FOUND  
NO\_PATHS\_FOUND MAX\_PATH\_LENGTH\_EXCEEDED PATHS\_FOUND IN-  
PUT\_RULES\_NOT\_FOUND MAX\_PATHS\_ZERO

**path\_found**

*boolean*

**result\_code**

*string*

**path\_metrics**

*list of PathMetric*

**paths**

*list of paths*

**max\_paths**

**max\_path\_length**

`indra.explanation.model_checker.remove_im_params(model, im)`

Remove parameter nodes from the influence map.

**Parameters**

- **model** (*pysb.core.Model*) – PySB model.
- **im** (*pygraphviz AGraph*) – Influence map.

**Returns** Influence map with the parameter nodes removed.

**Return type** *pygraphviz AGraph*

`indra.explanation.model_checker.stmt_from_rule(rule_name, model, stmts)`

Return the source INDRA Statement corresponding to a rule in a model.

**Parameters**

- **rule\_name** (*str*) – The name of a rule in the given PySB model.
- **model** (*pysb.core.Model*) – A PySB model which contains the given rule.
- **stmts** (*list[indra.statements.Statement]*) – A list of INDRA Statements from which the model was assembled.

**Returns** **stmt** – The Statement from which the given rule in the model was obtained.

**Return type** *indra.statements.Statement*

## 4.10 Tools (`indra.tools`)

### 4.10.1 Run assembly components in a pipeline (`indra.tools.assemble_corpus`)

`indra.tools.assemble_corpus.dump_statements(stmts, fname)`

Dump a list of statements into a pickle file.

**Parameters** **fname** (*str*) – The name of the pickle file to dump statements into.

`indra.tools.assemble_corpus.dump_stmt_strings(stmts, fname)`

Save printed statements in a file.

**Parameters**

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to save in a text file.
- **fname** (*Optional[str]*) – The name of a text file to save the printed statements into.

`indra.tools.assemble_corpus.expand_families(stmts_in, **kwargs)`

Expand Bioentities Agents to individual genes.

**Parameters**

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to expand.
- **save** (*Optional[str]*) – The name of a pickle file to save the results (`stmts_out`) into.

**Returns** `stmts_out` – A list of expanded statements.

**Return type** `list[indra.statements.Statement]`

`indra.tools.assemble_corpus.filter_belief(stmts_in, belief_cutoff, **kwargs)`

Filter to statements with belief above a given cutoff.

**Parameters**

- **stmts\_in** (`list[indra.statements.Statement]`) – A list of statements to filter.
- **belief\_cutoff** (`float`) – Only statements with belief above the `belief_cutoff` will be returned. Here  $0 < \text{belief\_cutoff} < 1$ .
- **save** (`Optional[str]`) – The name of a pickle file to save the results (`stmts_out`) into.

**Returns** `stmts_out` – A list of filtered statements.

**Return type** `list[indra.statements.Statement]`

`indra.tools.assemble_corpus.filter_by_type(stmts_in, stmt_type, **kwargs)`

Filter to a given statement type.

**Parameters**

- **stmts\_in** (`list[indra.statements.Statement]`) – A list of statements to filter.
- **stmt\_type** (`indra.statements.Statement`) – The class of the statement type to filter for. Example: `indra.statements.Modification`
- **invert** (`Optional[bool]`) – If True, the statements that are not of the given type are returned. Default: False
- **save** (`Optional[str]`) – The name of a pickle file to save the results (`stmts_out`) into.

**Returns** `stmts_out` – A list of filtered statements.

**Return type** `list[indra.statements.Statement]`

`indra.tools.assemble_corpus.filter_direct(stmts_in, **kwargs)`

Filter to statements that are direct interactions

**Parameters**

- **stmts\_in** (`list[indra.statements.Statement]`) – A list of statements to filter.
- **save** (`Optional[str]`) – The name of a pickle file to save the results (`stmts_out`) into.

**Returns** `stmts_out` – A list of filtered statements.

**Return type** `list[indra.statements.Statement]`

`indra.tools.assemble_corpus.filter_enzyme_kinase(stmts_in, **kwargs)`

Filter Phosphorylations to ones where the enzyme is a known kinase.

**Parameters**

- **stmts\_in** (`list[indra.statements.Statement]`) – A list of statements to filter.
- **save** (`Optional[str]`) – The name of a pickle file to save the results (`stmts_out`) into.

**Returns** `stmts_out` – A list of filtered statements.

**Return type** `list[indra.statements.Statement]`

```
indra.tools.assemble_corpus.filter_evidence_source(stmts_in, source_apis, policy='one', **kwargs)
```

Filter to statements that have evidence from a given set of sources.

#### Parameters

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to filter.
- **source\_apis** (*list[str]*) – A list of sources to filter for. Examples: biopax, bel, reach
- **policy** (*Optional[str]*) – If ‘one’, a statement that has evidence from any of the sources is kept. If ‘all’, only those statements are kept which have evidence from all the input sources specified in source\_apis. If ‘none’, only those statements are kept that don’t have evidence from any of the sources specified in source\_apis.
- **save** (*Optional[str]*) – The name of a pickle file to save the results (stmts\_out) into.

**Returns** **stmts\_out** – A list of filtered statements.

**Return type** *list[indra.statements.Statement]*

```
indra.tools.assemble_corpus.filter_gene_list(stmts_in, gene_list, policy, allow_families=False, **kwargs)
```

Return statements that contain genes given in a list.

#### Parameters

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to filter.
- **gene\_list** (*list[str]*) – A list of gene symbols to filter for.
- **policy** (*str*) – The policy to apply when filtering for the list of genes. “one”: keep statements that contain at least one of the list of genes and possibly others not in the list “all”: keep statements that only contain genes given in the list
- **allow\_families** (*Optional[bool]*) – Will include statements involving Bioentities families containing one of the genes in the gene list. Default: False
- **save** (*Optional[str]*) – The name of a pickle file to save the results (stmts\_out) into.

**Returns** **stmts\_out** – A list of filtered statements.

**Return type** *list[indra.statements.Statement]*

```
indra.tools.assemble_corpus.filter_genes_only(stmts_in, **kwargs)
```

Filter to statements containing genes only.

#### Parameters

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to filter.
- **specific\_only** (*Optional[bool]*) – If True, only elementary genes/proteins will be kept and families will be filtered out. If False, families are also included in the output. Default: False
- **save** (*Optional[str]*) – The name of a pickle file to save the results (stmts\_out) into.

**Returns** **stmts\_out** – A list of filtered statements.

**Return type** *list[indra.statements.Statement]*

`indra.tools.assemble_corpus.filter_grounded_only` (*stmts\_in*, *\*\*kwargs*)

Filter to statements that have grounded agents.

**Parameters**

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to filter.
- **save** (*Optional[str]*) – The name of a pickle file to save the results (*stmts\_out*) into.

**Returns** *stmts\_out* – A list of filtered statements.

**Return type** *list[indra.statements.Statement]*

`indra.tools.assemble_corpus.filter_human_only` (*stmts\_in*, *\*\*kwargs*)

Filter out statements that are not grounded to human genes.

**Parameters**

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to filter.
- **save** (*Optional[str]*) – The name of a pickle file to save the results (*stmts\_out*) into.

**Returns** *stmts\_out* – A list of filtered statements.

**Return type** *list[indra.statements.Statement]*

`indra.tools.assemble_corpus.filter_inconsequential_acts` (*stmts\_in*, *whitelist=None*, *\*\*kwargs*)

Filter out Activations that modify inconsequential activities

Inconsequential here means that the site is not mentioned / tested in any other statement. In some cases specific activity types should be preserved, for instance, to be used as readouts in a model. In this case, the given activities can be passed in a whitelist.

**Parameters**

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to filter.
- **whitelist** (*Optional[dict]*) – A whitelist containing agent activity types which should be preserved even if no other statement refers to them. The whitelist parameter is a dictionary in which the key is a gene name and the value is a list of activity types. Example: `whitelist = {'MAP2K1': ['kinase']}`
- **save** (*Optional[str]*) – The name of a pickle file to save the results (*stmts\_out*) into.

**Returns** *stmts\_out* – A list of filtered statements.

**Return type** *list[indra.statements.Statement]*

`indra.tools.assemble_corpus.filter_inconsequential_mods` (*stmts\_in*, *whitelist=None*, *\*\*kwargs*)

Filter out Modifications that modify inconsequential sites

Inconsequential here means that the site is not mentioned / tested in any other statement. In some cases specific sites should be preserved, for instance, to be used as readouts in a model. In this case, the given sites can be passed in a whitelist.

**Parameters**

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to filter.

- **whitelist** (*Optional[dict]*) – A whitelist containing agent modification sites whose modifications should be preserved even if no other statement refers to them. The whitelist parameter is a dictionary in which the key is a gene name and the value is a list of tuples of (modification\_type, residue, position). Example: `whitelist = {'MAP2K1': [('phosphorylation', 'S', '222')]}`
- **save** (*Optional[str]*) – The name of a pickle file to save the results (stmts\_out) into.

**Returns** `stmts_out` – A list of filtered statements.

**Return type** `list[indra.statements.Statement]`

`indra.tools.assemble_corpus.filter_mod_nokinase(stmts_in, **kwargs)`

Filter non-phospho Modifications to ones with a non-kinase enzyme.

#### Parameters

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to filter.
- **save** (*Optional[str]*) – The name of a pickle file to save the results (stmts\_out) into.

**Returns** `stmts_out` – A list of filtered statements.

**Return type** `list[indra.statements.Statement]`

`indra.tools.assemble_corpus.filter_mutation_status(stmts_in, mutations, deletions, **kwargs)`

Filter statements based on existing mutations/deletions

This filter helps to contextualize a set of statements to a given cell type. Given a list of deleted genes, it removes statements that refer to these genes. It also takes a list of mutations and removes statements that refer to mutations not relevant for the given context.

#### Parameters

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to filter.
- **mutations** (*dict*) – A dictionary whose keys are gene names, and the values are lists of tuples of the form (residue\_from, position, residue\_to). Example: `mutations = {'BRAF': [('V', '600', 'E')]}`
- **deletions** (*list*) – A list of gene names that are deleted.
- **save** (*Optional[str]*) – The name of a pickle file to save the results (stmts\_out) into.

**Returns** `stmts_out` – A list of filtered statements.

**Return type** `list[indra.statements.Statement]`

`indra.tools.assemble_corpus.filter_no_hypothesis(stmts_in, **kwargs)`

Filter to statements that are not marked as hypothesis in epistemics.

#### Parameters

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to filter.
- **save** (*Optional[str]*) – The name of a pickle file to save the results (stmts\_out) into.

**Returns** `stmts_out` – A list of filtered statements.

**Return type** `list[indra.statements.Statement]`



`indra.tools.assemble_corpus.filter_top_level` (*stmts\_in*, *\*\*kwargs*)

Filter to statements that are at the top-level of the hierarchy.

Here top-level statements correspond to most specific ones.

#### Parameters

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to filter.
- **save** (*Optional[str]*) – The name of a pickle file to save the results (*stmts\_out*) into.

**Returns** *stmts\_out* – A list of filtered statements.

**Return type** *list[indra.statements.Statement]*

`indra.tools.assemble_corpus.filter_transcription_factor` (*stmts\_in*, *\*\*kwargs*)

Filter out RegulateAmounts where subject is not a transcription factor.

#### Parameters

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to filter.
- **save** (*Optional[str]*) – The name of a pickle file to save the results (*stmts\_out*) into.

**Returns** *stmts\_out* – A list of filtered statements.

**Return type** *list[indra.statements.Statement]*

`indra.tools.assemble_corpus.filter_uuid_list` (*stmts\_in*, *uuids*, *\*\*kwargs*)

Filter to Statements corresponding to given UUIDs

#### Parameters

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to filter.
- **uuids** (*list[str]*) – A list of UUIDs to filter for.
- **save** (*Optional[str]*) – The name of a pickle file to save the results (*stmts\_out*) into.

**Returns** *stmts\_out* – A list of filtered statements.

**Return type** *list[indra.statements.Statement]*

`indra.tools.assemble_corpus.load_statements` (*fname*, *as\_dict=False*)

Load statements from a pickle file.

#### Parameters

- **fname** (*str*) – The name of the pickle file to load statements from.
- **as\_dict** (*Optional[bool]*) – If True and the pickle file contains a dictionary of statements, it is returned as a dictionary. If False, the statements are always returned in a list. Default: False

**Returns** *stmts* – A list or dict of statements that were loaded.

**Return type** *list*

`indra.tools.assemble_corpus.map_grounding` (*stmts\_in*, *\*\*kwargs*)

Map grounding using the GroundingMapper.

#### Parameters

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to map.

- **do\_rename** (*Optional[bool]*) – If True, Agents are renamed based on their mapped grounding.
- **save** (*Optional[str]*) – The name of a pickle file to save the results (stmts\_out) into.

**Returns** `stmts_out` – A list of mapped statements.

**Return type** `list[indra.statements.Statement]`

`indra.tools.assemble_corpus.map_sequence(stmts_in, **kwargs)`

Map sequences using the SiteMapper.

#### Parameters

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to map.
- **do\_methionine\_offset** (*boolean*) – Whether to check for off-by-one errors in site position (possibly) attributable to site numbering from mature proteins after cleavage of the initial methionine. If True, checks the reference sequence for a known modification at 1 site position greater than the given one; if there exists such a site, creates the mapping. Default is True.
- **do\_orthology\_mapping** (*boolean*) – Whether to check sequence positions for known modification sites in mouse or rat sequences (based on PhosphoSitePlus data). If a mouse/rat site is found that is linked to a site in the human reference sequence, a mapping is created. Default is True.
- **do\_isoform\_mapping** (*boolean*) – Whether to check sequence positions for known modifications in other human isoforms of the protein (based on PhosphoSitePlus data). If a site is found that is linked to a site in the human reference sequence, a mapping is created. Default is True.
- **save** (*Optional[str]*) – The name of a pickle file to save the results (stmts\_out) into.

**Returns** `stmts_out` – A list of mapped statements.

**Return type** `list[indra.statements.Statement]`

`indra.tools.assemble_corpus.reduce_activities(stmts_in, **kwargs)`

Reduce the activity types in a list of statements

#### Parameters

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to reduce activity types in.
- **save** (*Optional[str]*) – The name of a pickle file to save the results (stmts\_out) into.

**Returns** `stmts_out` – A list of reduced activity statements.

**Return type** `list[indra.statements.Statement]`

`indra.tools.assemble_corpus.run_preassembly(stmts_in, **kwargs)`

Run preassembly on a list of statements.

#### Parameters

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to pre-assemble.
- **return\_toplevel** (*Optional[bool]*) – If True, only the top-level statements are returned. If False, all statements are returned irrespective of level of specificity. Default: True

- **poolsize** (*Optional[int]*) – The number of worker processes to use to parallelize the comparisons performed by the function. If None (default), no parallelization is performed. NOTE: Parallelization is only available on Python 3.4 and above.
- **size\_cutoff** (*Optional[int]*) – Groups with size\_cutoff or more statements are sent to worker processes, while smaller groups are compared in the parent process. Default value is 100. Not relevant when parallelization is not used.
- **save** (*Optional[str]*) – The name of a pickle file to save the results (stmts\_out) into.
- **save\_unique** (*Optional[str]*) – The name of a pickle file to save the unique statements into.

**Returns** `stmts_out` – A list of preassembled top-level statements.

**Return type** `list[indra.statements.Statement]`

```
indra.tools.assemble_corpus.run_preassembly_duplicate(preassembler, beliefengine,
                                                    **kwargs)
```

Run deduplication stage of preassembly on a list of statements.

#### Parameters

- **preassembler** (`indra.preassembler.Preassembler`) – A Preassembler instance
- **beliefengine** (`indra.belief.BeliefEngine`) – A BeliefEngine instance
- **save** (*Optional[str]*) – The name of a pickle file to save the results (stmts\_out) into.

**Returns** `stmts_out` – A list of unique statements.

**Return type** `list[indra.statements.Statement]`

```
indra.tools.assemble_corpus.run_preassembly_related(preassembler, beliefengine,
                                                    **kwargs)
```

Run related stage of preassembly on a list of statements.

#### Parameters

- **preassembler** (`indra.preassembler.Preassembler`) – A Preassembler instance which already has a set of unique statements internally.
- **beliefengine** (`indra.belief.BeliefEngine`) – A BeliefEngine instance
- **return\_toplevel** (*Optional[bool]*) – If True, only the top-level statements are returned. If False, all statements are returned irrespective of level of specificity. Default: True
- **poolsize** (*Optional[int]*) – The number of worker processes to use to parallelize the comparisons performed by the function. If None (default), no parallelization is performed. NOTE: Parallelization is only available on Python 3.4 and above.
- **size\_cutoff** (*Optional[int]*) – Groups with size\_cutoff or more statements are sent to worker processes, while smaller groups are compared in the parent process. Default value is 100. Not relevant when parallelization is not used.
- **save** (*Optional[str]*) – The name of a pickle file to save the results (stmts\_out) into.

**Returns** `stmts_out` – A list of preassembled top-level statements.

**Return type** `list[indra.statements.Statement]`

```
indra.tools.assemble_corpus.strip_agent_context(stmts_in, **kwargs)
```

Strip any context on agents within each statement.

**Parameters**

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements whose agent context should be stripped.
- **save** (*Optional[str]*) – The name of a pickle file to save the results (stmts\_out) into.

**Returns** **stmts\_out** – A list of stripped statements.

**Return type** *list[indra.statements.Statement]*

## 4.10.2 Build a network from a gene list (`indra.tools.gene_network`)

**class** `indra.tools.gene_network.GeneNetwork` (*gene\_list, basename=None*)

Build a set of INDRA statements for a given gene list from databases.

**Parameters**

- **gene\_list** (*string*) – List of gene names.
- **basename** (*string or None (default)*) – Filename prefix to be used for caching of intermediates (Biopax OWL file, pickled statement lists, etc.). If None, no results are cached and no cached files are used.

**gene\_list**

*string* – List of gene names

**basename**

*string or None* – Filename prefix for cached intermediates, or None if no cached used.

**results**

*dict* – Dict containing results of preassembly (see return type for `run_preassembly()`).

**get\_bel\_stmts** (*filter=False*)

Get relevant statements from the BEL large corpus.

Performs a series of neighborhood queries and then takes the union of all the statements. Because the query process can take a long time for large gene lists, the resulting list of statements are cached in a pickle file with the filename `<basename>_bel_stmts.pkl`. If the pickle file is present, it is used by default; if not present, the queries are performed and the results are cached.

**Parameters** **filter** (*bool*) – If True, includes only those statements that exclusively mention genes in *gene\_list*. Default is False. Note that the full (unfiltered) set of statements are cached.

**Returns** List of INDRA statements extracted from the BEL large corpus.

**Return type** list of *indra.statements.Statement*

**get\_biopax\_stmts** (*filter=False, query='pathsbetween', database\_filter=None*)

Get relevant statements from Pathway Commons.

Performs a “paths between” query for the genes in *gene\_list* and uses the results to build statements. This function caches two files: the list of statements built from the query, which is cached in `<basename>_biopax_stmts.pkl`, and the OWL file returned by the Pathway Commons Web API, which is cached in `<basename>_pc_pathsbetween.owl`. If these cached files are found, then the results are returned based on the cached file and Pathway Commons is not queried again.

**Parameters**

- **filter** (*Optional[bool]*) – If True, includes only those statements that exclusively mention genes in *gene\_list*. Default is False.

- **query** (*Optional[str]*) – Defined what type of query is executed. The two options are ‘pathsbetween’ which finds paths between the given list of genes and only works if more than 1 gene is given, and ‘neighborhood’ which searches the immediate neighborhood of each given gene. Note that for pathsbetween queries with more than 60 genes, the query will be executed in multiple blocks for scalability.
- **database\_filter** (*Optional[list[str]]*) – A list of PathwayCommons databases to include in the query.

**Returns** List of INDRA statements extracted from Pathway Commons.

**Return type** list of *indra.statements.Statement*

**get\_statements** (*filter=False*)

Return the combined list of statements from BEL and Pathway Commons.

Internally calls *get\_biopax\_stmts()* and *get\_bel\_stmts()*.

**Parameters** **filter** (*bool*) – If True, includes only those statements that exclusively mention genes in *gene\_list*. Default is False.

**Returns** List of INDRA statements extracted the BEL large corpus and Pathway Commons.

**Return type** list of *indra.statements.Statement*

**run\_preassembly** (*stmts, print\_summary=True*)

Run complete preassembly procedure on the given statements.

Results are returned as a dict and stored in the attribute *results*. They are also saved in the pickle file *<basename>\_results.pkl*.

**Parameters**

- **stmts** (list of *indra.statements.Statement*) – Statements to preassemble.
- **print\_summary** (*bool*) – If True (default), prints a summary of the preassembly process to the console.

**Returns**

A dict containing the following entries:

- *raw*: the starting set of statements before preassembly.
- *duplicates1*: statements after initial de-duplication.
- *valid*: statements found to have valid modification sites.
- *mapped*: mapped statements (list of *indra.preassembler.sitemapper.MappedStatement*).
- *mapped\_stmts*: combined list of valid statements and statements after mapping.
- *duplicates2*: statements resulting from de-duplication of the statements in *mapped\_stmts*.
- *related2*: top-level statements after combining the statements in *duplicates2*.

**Return type** dict

### 4.10.3 Build an executable model from a fragment of a large network (`indra.tools.executable_subnetwork`)

```
indra.tools.executable_subnetwork.get_subnetwork(statements, nodes, relevance_network=None, relevance_node_lim=10)
```

Return a PySB model based on a subset of given INDRA Statements.

Statements are first filtered for nodes in the given list and other nodes are optionally added based on relevance in a given network. The filtered statements are then assembled into an executable model using INDRA's PySB Assembler.

#### Parameters

- **statements** (*list[indra.statements.Statement]*) – A list of INDRA Statements to extract a subnetwork from.
- **nodes** (*list[str]*) – The names of the nodes to extract the subnetwork for.
- **relevance\_network** (*Optional[str]*) – The UUID of the NDEX network in which nodes relevant to the given nodes are found.
- **relevance\_node\_lim** (*Optional[int]*) – The maximal number of additional nodes to add to the subnetwork based on relevance.

**Returns** **model** – A PySB model object assembled using INDRA's PySB Assembler from the INDRA Statements corresponding to the subnetwork.

**Return type** `pysb.Model`

### 4.10.4 Build a model incrementally over time (`indra.tools.incremental_model`)

```
class indra.tools.incremental_model.IncrementalModel(model_fname=None)
```

Assemble a model incrementally by iteratively adding new Statements.

**Parameters** **model\_fname** (*Optional[str]*) – The name of the pickle file in which a set of INDRA Statements are stored in a dict keyed by PubMed IDs. This is the state of an IncrementalModel that is loaded upon instantiation.

#### **stmts**

*dict[str, list[indra.statements.Statement]]* – A dictionary of INDRA Statements keyed by PMIDs that stores the current state of the IncrementalModel.

#### **assembled\_stmts**

*list[indra.statements.Statement]* – A list of INDRA Statements after assembly.

#### **add\_statements** (*pmid, stmts*)

Add INDRA Statements to the incremental model indexed by PMID.

#### Parameters

- **pmid** (*str*) – The PMID of the paper from which statements were extracted.
- **stmts** (*list[indra.statements.Statement]*) – A list of INDRA Statements to be added to the model.

#### **get\_model\_agents** ()

Return a list of all Agents from all Statements.

**Returns** **agents** – A list of Agents that are in the model.

**Return type** `list[indra.statements.Agent]`

**get\_statements** ()

Return a list of all Statements in a single list.

**Returns** `stmts` – A list of all the INDRA Statements in the model.

**Return type** `list[indra.statements.Statement]`

**get\_statements\_noprior** ()

Return a list of all non-prior Statements in a single list.

**Returns** `stmts` – A list of all the INDRA Statements in the model (excluding the prior).

**Return type** `list[indra.statements.Statement]`

**get\_statements\_prior** ()

Return a list of all prior Statements in a single list.

**Returns** `stmts` – A list of all the INDRA Statements in the prior.

**Return type** `list[indra.statements.Statement]`

**load\_prior** (*prior\_fname*)

Load a set of prior statements from a pickle file.

The prior statements have a special key in the `stmts` dictionary called “prior”.

**Parameters** `prior_fname` (*str*) – The name of the pickle file containing the prior Statements.

**preassemble** (*filters=None*)

Preassemble the Statements collected in the model.

Use INDRA’s GroundingMapper, Preassembler and BeliefEngine on the IncrementalModel and save the unique statements and the top level statements in class attributes.

Currently the following filter options are implemented: - `grounding`: require that all Agents in statements are grounded - `human_only`: require that all proteins are human proteins - `prior_one`: require that at least one Agent is in the prior model - `prior_all`: require that all Agents are in the prior model

**Parameters** `filters` (*Optional[list[str]]*) – A list of filter options to apply when choosing the statements. See description above for more details. Default: None

**save** (*model\_fname='model.pkl'*)

Save the state of the IncrementalModel in a pickle file.

**Parameters** `model_fname` (*Optional[str]*) – The name of the pickle file to save the state of the IncrementalModel in. Default: `model.pkl`

#### 4.10.5 The RAS Machine (`indra.tools.machine`)

##### Prerequisites

First, install the machine-specific dependencies:

```
pip install indra[machine]
```

## Starting a New Model

To start a new model, run

```
python -m indra.tools.machine make model_name
```

Alternatively, the command line interface can be invoked with

```
indra-machine make model_name
```

where `model_name` corresponds to the name of the model to initialize.

This script generates the following folders and files

- `model_name`
- `model_name/log.txt`
- `model_name/config.yaml`
- `model_name/jsons/`

You should the edit `model_name/config.yaml` to set up the search terms and optionally the credentials to use Twitter, Gmail or NDEx bindings.

## Setting Up Search Terms

The `config.yaml` file is a standard YAML configuration file. A template is available in `model_name/config.yaml` after having created the machine.

Two important fields in `config.yaml` are `search_terms` and `search_genes` both of which are YAML lists. The entries of `search_terms` are used directly as queries in PubMed search (for more information on PubMed search strings, read [https://www.ncbi.nlm.nih.gov/books/NBK3827/#pubmedhelp.Searching\\_PubMed](https://www.ncbi.nlm.nih.gov/books/NBK3827/#pubmedhelp.Searching_PubMed)).

Example:

```
search_terms:  
- breast cancer  
- proteasome  
- apoptosis
```

The entries of `search_genes` is a special list in which only standard HGNC gene symbols are allowed. Entries in this list are also used to search PubMed but also serve as a list of *prior* genes that are known to be relevant for the model.

#Entries in this can be used to search #PubMed specifically for articles that are tagged with the gene's unique #identifier rather than its string name. This mode of searching for articles #on specific genes is much more reliable than searching for them using #string names.

Example:

```
search_genes:  
- AKT1  
- MAPK3  
- EGFR
```



## Extending a Model

To extend a model, run

```
python -m indra.tools.machine run_with_search model_name
```

Alternatively, the command line interface can be invoked with

```
indra-machine run_with_search model_name
```

Extending a model involves extracting PMIDs from emails (if Gmail credentials are given), and searching using INDRA's PubMed client with each entry of `search_terms` in `config.yaml` as a search term. INDRA's literature client is then used to find the full text corresponding to each PMID or its abstract when the full text is not available. The REACH parser is then used to read each new paper. INDRA uses the REACH output to construct Statements corresponding to mechanisms. It then adds them to an incremental model through a process of assembly involving duplication and overlap resolution and the application of filters.

`indra.tools.machine.copy_default_config(destination)`

Copies the default configuration to the given destination

**Parameters** `destination` (*str*) – The location to which a default RAS Machine config file is placed.

## 4.11 INDRA Database tools (`indra.db`)

**class** `indra.db.DatabaseManager` (*host*, *sqltype*='postgresql', *label*=None)

An object used to access INDRA's database.

This object can be used to access and manage indra's database. It includes both basic methods and some useful, more high-level methods. It is designed to be used with postgresql, or sqlite.

This object is primarily built around sqlalchemy, which is a required package for its use. It also optionally makes use of the pgcopy package for large data transfers.

If you wish to access the primary database, you can simply use the `get_primary_db` to get an instance of this object using the default settings.

### Parameters

- **host** (*str*) – The database to which you want to interface.
- **sqltype** (*OPTIONAL* [*str*]) – The type of sql library used. Use one of the sql types provided by *sqltypes*. Default is *sqltypes.POSTGRESQL*
- **label** (*OPTIONAL* [*str*]) – A short string to indicate the purpose of the db instance. Set as primary when initialized by `get_primary_db`.

### Example

If you wish to access the primary database and find the metadata for a particular pmid, 1234567:

```
>> from indra.db import get_primary_db() >> db = get_primary_db() >> res = db.select_all(db.TextRef,
db.TextRef.pmid == '1234567')
```

You will get a list of objects whose attributes give the metadata contained in the columns of the table.

For more sophisticated examples, several use cases can be found in `indra.tests.test_db`.

**commit** (*err\_msg*)

Commit, and give useful info if there is an exception.

**copy** (*tbl\_name, data, cols=None*)

Use `pg_copy` to copy over a large amount of data.

**create\_tables** ()

Create the tables for INDRA database.

**delete\_all** (*entry\_list*)

Remove the given records from the given table.

**drop\_tables** ()

Drop all the tables for INDRA database

**filter\_query** (*tbls, \*args*)

Query a table and filter results.

**get\_abstracts\_by\_pmids** (*pmid\_list, unzip=True*)

Get abstracts using the pmids in `pmid_list`.

**get\_active\_tables** ()

Get the tables currently active in the database.

**get\_all\_pmids** ()

Get a list of all the pmids on record.

**get\_columns** (*tbl\_name*)

Get a list of the column labels for a table.

**get\_tables** ()

Get a list of available tables.

**get\_values** (*entry\_list, col\_names=None, keyed=False*)

Get the column values from the entries in `entry_list`

**grab\_session** ()

Get an active session with the database.

**has\_entry** (*tbls, \*args*)

Check whether an entry/entries matching given specs live in the db.

**insert** (*tbl\_name, ret\_info='id', \*\*input\_dict*)

Insert a an entry into specified table, and return id.

**insert\_db\_stmts** (*stmts, db\_ref\_id*)

Insert statement, their database, and any affiliated agents.

**insert\_many** (*tbl\_name, input\_dict\_list, ret\_info='id'*)

Insert many records into the table given by `tbl_name`.

**select\_all** (*tbls, \*args*)

Select any and all entries from table given by `tbl_name`.

The results will be filtered by your keyword arguments. For example if you want to get a text ref with pmid '10532205', you would call:

```
db.select_all('text_ref', db.TextRef.pmid == '10532205')
```

Note that double equals are required, not a single equal. Equivalently you could call:

```
db.select_all(db.TextRef, db.TextRef.pmid == '10532205')
```

For a more complicated example, suppose you want to get all text refs that have full text from pmc oa, you could select:

```
db.select_all(
    [db.TextRef, db.TextContent],
    db.TextContent.text_ref_id == db.TextRef.id,
    db.TextContent.source == 'pmc_oa',
    db.TextContent.text_type == 'fulltext'
)
```

**select\_one** (*tbls*, \**args*)

Select the first value that matches requirements.

Requirements are given in kwargs from table indicated by *tbl\_name*. See *select\_all*.

Note that if your specification yields multiple results, this method will just return the first result without exception.

**show\_tables** ()

Print a list of all the available tables.

`indra.db.get_defaults` ()

Get the default database hosts provided in the specified *DEFAULTS\_FILE*.

`indra.db.get_primary_db` ()

Get an instance to the primary database host.

#### 4.11.1 Content Manager (`indra.db.manage_content`)

**class** `indra.db.manage_content.IntegrityError`

Using this in a try-except will catch nothing. (That's the point.)

**class** `indra.db.manage_content.Manager`

Abstract class for all upload/update managers.

This abstract class provides the api required for any object that is used to manage content between the database and the content.

**copy\_into\_db** (*db*, \**args*, \*\**kwargs*)

Wrapper around the `db.copy` feature, pickles args upon exception.

**populate** (*db*)

A stub for the method used to initially populate the database.

**update** (*db*)

A stub for the method used to update the content on the database.

**class** `indra.db.manage_content.Manuscripts` (\**args*, \*\**kwargs*)

Manager for the pmc manuscripts.

**get\_tarname\_from\_filename** (*fname*)

Get the name of the tar file based on the file name (or a pmcid).

**class** `indra.db.manage_content.Medline` (\**args*, \*\**kwargs*)

Manager for the medline content.

**fix\_doi** (*doi*)

Sometimes the doi is doubled (no idea why). Fix it.

**populate** (*db*, *n\_procs=1*, *continuing=False*)

Perform the initial input of the pubmed content into the database.

### Parameters

- **db** (*indra.db.DatabaseManager instance*) – The database to which the data will be uploaded.
- **n\_procs** (*int*) – The number of processes to use when parsing xmls.
- **continuing** (*bool*) – If true, assume that we are picking up after an error, or otherwise continuing from an earlier process. This means we will skip over source files contained in the database. If false, all files will be read and parsed.

**class** `indra.db.manage_content.NihFtpClient` (*my\_path*, *ftp\_url='ftp.ncbi.nlm.nih.gov'*, *local=False*)

High level access to the NIH FTP repositories.

### Parameters

- **my\_path** (*str*) – The path to the subdirectory around which this client operates.
- **ftp\_url** (*str*) – The url to the ftp site. May be a local directory (see *local*). By default this is *'ftp.ncbi.nlm.nih.gov'*.
- **local** (*bool*) – These methods may be run on a local directory (intended for testing). (default is *False*).

**download\_file** (*f\_path*)

Download a file into a file given by *f\_path*.

**ftp\_ls** (*ftp\_path=None*)

Get a list of the contents in the ftp directory.

**get\_csv\_as\_dict** (*csv\_file*)

Get the content from a csv file as a list of dicts.

**get\_file** (*f\_path*, *force\_str=True*)

Get the contents of a file as a string.

**get\_uncompressed\_bytes** (*f\_path*, *force\_str=True*)

Get a file that is gzipped, and return the unzipped string.

**get\_xml\_file** (*xml\_file*)

Get the content from an xml file as an ElementTree.

**ret\_file** (*f\_path*, *buf*)

Load the content of a file into the given buffer.

**class** `indra.db.manage_content.NihManager` (*\*args*, *\*\*kwargs*)

Abstract class for all the managers that use the NIH FTP service.

See *NihFtpClient* for parameters.

**class** `indra.db.manage_content.PmcManager` (*\*args*, *\*\*kwargs*)

Abstract class for uploaders of PMC content.

For Parameters, see *NihManager*.

**filter\_text\_content** (*db*, *tc\_data*)

Filter the text content to identify pre-existing records.

**filter\_text\_refs** (*db*, *tr\_data*)

Try to reconcile the data we have with what's already on the db.

**get\_data\_from\_xml\_str** (*xml\_str*, *filename*)

Get the data out of the xml string.

**get\_missing\_pmids** (*tr\_data*)

Try to get missing pmids.

**populate** (*db, n\_procs=1, continuing=False*)

Perform the initial population of the pmc content into the database.

**Parameters**

- **db** (*indra.db.DatabaseManager instance*) – The database to which the data will be uploaded.
- **n\_procs** (*int*) – The number of processes to use when parsing xmls.
- **continuing** (*bool*) – If true, assume that we are picking up after an error, or otherwise continuing from an earlier process. This means we will skip over source files contained in the database. If false, all files will be read and parsed.

**upload\_archive** (*archive, q=None, db=None*)

Process a single tar gzipped article archive.

**upload\_batch** (*db, tr\_data, tc\_data*)

Add a batch of text refs and text content to the database.

**class** *indra.db.manage\_content.PmcOA* (*\*args, \*\*kwargs*)

Manager for the pmc open access content.



## 5.1 Using natural language to build models

In this tutorial we build a simple model using natural language, and export it into different formats.

### 5.1.1 Read INDRA Statements from a natural language string

First we import INDRA's API to the TRIPS reading system. We then define a block of text which serves as the description of the mechanism to be modeled in the *model\_text* variable. Finally, *indra.sources.trips.process\_text* is called which sends a request to the TRIPS web service, gets a response and processes the extraction knowledge base to obtain a list of INDRA Statements

```
In [1]: from indra.sources import trips
In [2]: model_text = 'MAP2K1 phosphorylates MAPK1 and DUSP6 dephosphorylates MAPK1.'
In [3]: tp = trips.process_text(model_text)
```

At this point *tp.statements* should contain 2 INDRA Statements: a Phosphorylation Statement and a Dephosphorylation Statement. Note that the evidence sentence for each Statement is propagated:

```
In [4]: for st in tp.statements:
...:     print('%s with evidence "%s"' % (st, st.evidence[0].text))
...:
Phosphorylation(MAP2K1(), MAPK1()) with evidence "MAP2K1 phosphorylates MAPK1 and_
↪DUSP6 dephosphorylates MAPK1."
Dephosphorylation(DUSP6(), MAPK1()) with evidence "MAP2K1 phosphorylates MAPK1 and_
↪DUSP6 dephosphorylates MAPK1."
```

## 5.1.2 Assemble the INDRA Statements into a rule-based executable model

We next use INDRA's PySB Assembler to automatically assemble a rule-based model representing the biochemical mechanisms described in *model\_text*. First a PysbAssembler object is instantiated, then the list of INDRA Statements is added to the assembler. Finally, the assembler's *make\_model* method is called which assembles the model and returns it, while also storing it in *pa.model*. Notice that we are using *policies='two\_step'* as an argument of *make\_model*. This directs the assemble to use rules in which enzymatic catalysis is modeled as a two-step process in which enzyme and substrate first reversibly bind and the enzyme-substrate complex produces and releases a product irreversibly.

```
In [5]: from indra.assemblers.pysb_assembler import PysbAssembler

In [6]: pa = PysbAssembler()

In [7]: pa.add_statements(tp.statements)

In [8]: pa.make_model(policies='two_step')
Out[8]: <Model 'None' (monomers: 3, rules: 6, parameters: 9, expressions: 0,
↪compartments: 0) at 0x7f8de2765860>
```

At this point *pa.model* contains a PySB model object with 3 monomers,

```
In [9]: for monomer in pa.model.monomers:
...:     print(monomer)
...:
Monomer('MAP2K1', ['mapk1'])
Monomer('DUSP6', ['mapk1'])
Monomer('MAPK1', ['phospho', 'map2k1', 'dusp6'], {'phospho': ['u', 'p']})
```

6 rules,

```
In [10]: for rule in pa.model.rules:
...:     print(rule)
...:
Rule('MAP2K1_phosphorylation_bind_MAPK1_phospho', MAP2K1(mapk1=None) + MAPK1(phospho=
↪'u', map2k1=None) >> MAP2K1(mapk1=1) % MAPK1(phospho='u', map2k1=1), kf_mm_bind_1)
Rule('MAP2K1_phosphorylation_MAPK1_phospho', MAP2K1(mapk1=1) % MAPK1(phospho='u',
↪map2k1=1) >> MAP2K1(mapk1=None) + MAPK1(phospho='p', map2k1=None), kc_mm_
↪phosphorylation_1)
Rule('MAP2K1_dissoc_MAPK1', MAP2K1(mapk1=1) % MAPK1(map2k1=1) >> MAP2K1(mapk1=None) +
↪MAPK1(map2k1=None), kr_mm_bind_1)
Rule('DUSP6_dephosphorylation_bind_MAPK1_phospho', DUSP6(mapk1=None) + MAPK1(phospho=
↪'p', dusp6=None) >> DUSP6(mapk1=1) % MAPK1(phospho='p', dusp6=1), kf_dm_bind_1)
Rule('DUSP6_dephosphorylation_MAPK1_phospho', DUSP6(mapk1=1) % MAPK1(phospho='p',
↪dusp6=1) >> DUSP6(mapk1=None) + MAPK1(phospho='u', dusp6=None), kc_dm_
↪dephosphorylation_1)
Rule('DUSP6_dissoc_MAPK1', DUSP6(mapk1=1) % MAPK1(dusp6=1) >> DUSP6(mapk1=None) +
↪MAPK1(dusp6=None), kr_dm_bind_1)
```

and 9 parameters (6 kinetic rate constants and 3 total protein amounts) that are set to nominal but plausible values,

```
In [11]: for parameter in pa.model.parameters:
...:     print(parameter)
...:
Parameter('kf_mm_bind_1', 1e-06)
Parameter('kr_mm_bind_1', 0.1)
Parameter('kc_mm_phosphorylation_1', 100.0)
Parameter('kf_dm_bind_1', 1e-06)
Parameter('kr_dm_bind_1', 0.1)
```



```
Parameter('kc_dm_dephosphorylation_1', 100.0)
Parameter('MAP2K1_0', 10000.0)
Parameter('DUSP6_0', 10000.0)
Parameter('MAPK1_0', 10000.0)
```

The model also contains extensive annotations that tie the monomers to database identifiers and also annotate the semantics of each component of each rule.

```
In [12]: for annotation in pa.model.annotations:
        ....:     print(annotation)
        ....:
Annotation(MAP2K1, 'http://identifiers.org/uniprot/Q02750', 'is')
Annotation(MAP2K1, 'http://identifiers.org/hgnc/HGNC:6840', 'is')
Annotation(MAP2K1, 'http://identifiers.org/ncit/C17808', 'is')
Annotation(DUSP6, 'http://identifiers.org/uniprot/Q16828', 'is')
Annotation(DUSP6, 'http://identifiers.org/hgnc/HGNC:3072', 'is')
Annotation(DUSP6, 'http://identifiers.org/ncit/C106026', 'is')
Annotation(MAPK1, 'http://identifiers.org/uniprot/P28482', 'is')
Annotation(MAPK1, 'http://identifiers.org/hgnc/HGNC:6871', 'is')
Annotation(MAPK1, 'http://identifiers.org/ncit/C17589', 'is')
Annotation(MAP2K1_phosphorylation_bind_MAPK1_phospho, 'c8ea8ab7-ce03-4767-bcef-
↪ccb84514d01b', 'from_indra_statement')
Annotation(MAP2K1_phosphorylation_MAPK1_phospho, 'MAP2K1', 'rule_has_subject')
Annotation(MAP2K1_phosphorylation_MAPK1_phospho, 'MAPK1', 'rule_has_object')
Annotation(MAP2K1_phosphorylation_MAPK1_phospho, 'c8ea8ab7-ce03-4767-bcef-ccb84514d01b
↪', 'from_indra_statement')
Annotation(MAP2K1_dissoc_MAPK1, 'c8ea8ab7-ce03-4767-bcef-ccb84514d01b', 'from_indra_
↪statement')
Annotation(DUSP6_dephosphorylation_bind_MAPK1_phospho, '2a363008-1dfd-4945-b8cb-
↪66a580ba4ee1', 'from_indra_statement')
Annotation(DUSP6_dephosphorylation_MAPK1_phospho, 'DUSP6', 'rule_has_subject')
Annotation(DUSP6_dephosphorylation_MAPK1_phospho, 'MAPK1', 'rule_has_object')
Annotation(DUSP6_dephosphorylation_MAPK1_phospho, '2a363008-1dfd-4945-b8cb-
↪66a580ba4ee1', 'from_indra_statement')
Annotation(DUSP6_dissoc_MAPK1, '2a363008-1dfd-4945-b8cb-66a580ba4ee1', 'from_indra_
↪statement')
```

### 5.1.3 Exporting the model into other common formats

From the assembled PySB format it is possible to export the model into other common formats such as SBML, BNGL and Kappa. One can also generate a Matlab or Mathematica script with ODEs corresponding to the model.

```
pa.export_model('sbml')
pa.export_model('bngl')
```

One can also pass a file name argument to the *export\_model* function to save the exported model directly into a file:

```
pa.export_model('sbml', 'example_model.sbml')
```

## 5.2 Large-Scale Machine Reading with Starcluster

The following doc describes the steps involved in reading a large numbers of papers in parallel on Amazon EC2 using REACH, caching the JSON output on Amazon S3, then processing the REACH output into INDRA Statements. Prerequisites for doing the following are:

- A cluster of Amazon EC2 nodes configured using Starcluster, with INDRA installed and in the PYTHONPATH
- An Amazon S3 bucket containing full text contents for papers, keyed by Pubmed ID (creation of this S3 repository will be described in another tutorial).

This tutorial goes through the individual steps involved before describing how all of them can be run through the use of a single submission script, `submit_reading_pipeline.py`.

Note also that the prerequisite installation steps can be streamlined by putting them in a setup script that can be re-run upon instantiating a new Amazon cluster or by using them to configure a custom Amazon EC2 AMI.

### 5.2.1 Install REACH

Install SBT. On an EC2 Linux machine, run the following lines (drawn from <http://www.scala-sbt.org/0.13/docs/Installing-sbt-on-Linux.html>):

```
echo "deb https://dl.bintray.com/sbt/debian /" | sudo tee -a /etc/apt/sources.list.d/
↪sbt.list
sudo apt-key adv --keyserver hkp://keyserver.ubuntu.com:80 --recv 642AC823
sudo apt-get update
sudo apt-get install sbt
```

Clone REACH from <https://github.com/clulab/reach>.

Add the following line to `reach/build.sbt`:

```
mainClass in assembly := Some("org.clulab.reach.ReachCLI")
```

This assigns `ReachCLI` as the main class.

Compile and assemble REACH. Note that the path to the `.ivy2` directory must be given. Use the assembly task to assemble a fat JAR containing all of the dependencies with the correct main class. Run the following from the directory containing the REACH `build.sbt` file (e.g., `/pmc/reach`):

```
sbt -Dsbty.ivy.home=/pmc/reach/.ivy2 compile
sbt -Dsbty.ivy.home=/pmc/reach/.ivy2 assembly
```

### 5.2.2 Install Amazon S3 support

Install `boto3`:

```
pip install boto3
```

---

**Note:** If using EC2, make sure to install `boto3`, `jsonpickle`, and Amazon credentials on all nodes, not just the master node.

---

Add Amazon credentials to access the S3 bucket. First create the `.aws` directory on the EC2 instance:

```
mkdir /home/sgeadmin/.aws
```

Then set up Amazon credentials, for example by copying from your local machine using `StarCluster`:

```
starcluster put mycluster ~/.aws/credentials /home/sgeadmin/.aws
```



```
<ipython-input-5-dac6af0328b3> in <module>()
      1 with open('%s_pmids.txt' % gene, 'w') as f:
----> 2     for pmid in pmids_oa_xml:
      3         f.write('%s\n' % pmid)
      4

NameError: name 'pmids_oa_xml' is not defined
```

This creates a file, `SOS2_pmids.txt`, containing the PMIDs that we will read with REACH.

## 5.2.5 Process the papers with REACH

The next step is to read the content of the papers with REACH in a parallelizable, high-throughput way. To do this, run the script `indra/tools/reading/run_reach_on_pmids.py`. If necessary update the lines at the top of the script with the REACH settings, e.g.:

```
cleanup = False
verbose = True
path_to_reach = '/pmc/reach/target/scala-2.11/reach-assembly-1.3.2-SNAPSHOT.jar'
reach_version = '1.3.2'
source_text = 'pmc_oa_xml'
force_read = False
```

The `reach_version` is important because it is used to determine whether the paper has already been read with this version of REACH (in which case it will be skipped), or if the REACH output needs to be updated. Alternatively, if you want to read all the papers regardless of whether they've been read before with the given version of REACH, set the `force_read` variable to `True`.

Next, create a top-level temporary directory to use during reading. This will be used to store the input files and the JSON output:

```
mkdir my_temp_dir
```

Run `run_reach_on_pmids.py`, passing arguments for the PMID list file, the temp directory, the number of cores to use on the machine, the PMID start index (in the PMID list file) and the end index. The start and end indices are used to subdivide the job into parallelizable chunks. If the end index is greater than the total number of PMIDs, it will process up to the last one in the list. For example:

```
python run_reach_on_pmids.py SOS2_pmids.txt my_temp_dir 8 0 10
```

This uses 8 cores to process the first ten papers listed in the file `SOS2_pmids.txt`. REACH will run, output the JSON files in the temporary directory, e.g. in `my_temp_dir/read_0_to_10_MSP6YI/output`, assemble the JSON files together, and upload the results to S3. If you attempt to process the files again with the same version of REACH, the script will detect that the JSON output from that version is already on S3 and skip those papers.

This can be submitted to run offline using the job scheduler on EC2 with, e.g.:

```
qsub -b y -cwd -V -pe orte 8 python run_reach_on_pmids.py SOS2_pmids.txt my_temp_dir_
↪8 0 10
```

**Note:** The number of cores requested in the `qsub` call (`'-pe orte 8'`) should match the number of cores passed to the `run_reach_on_pmids.py` script, which determines the number of threads that REACH will attempt to use (the third-to-last argument above). This should also match the total number of nodes on the Amazon EC2 node (e.g., 8 cores for

c3.2xlarge). This way the job scheduler will schedule the job to run on all the cores of a single EC2 node, and REACH will use them all.

## 5.2.6 Extract INDRA Statements from the REACH output on S3

The script `indra/tools/reading/process_reach_from_s3.py` is used to extract INDRA Statements from the REACH output uploaded to S3 in the previous step. This process can also be parallelized by submitting chunks of papers to be processed by different cores. The INDRA statements for each chunk of papers are pickled and can be assembled into a single pickle file in a subsequent step.

Following the example above, run the following to process the REACH output for the SOS2 papers into INDRA statements. We'll do this in two chunks to show how the process can be parallelized and the statements assembled from multiple files:

```
python process_reach_from_s3.py SOS2_pmids.txt 0 5
python process_reach_from_s3.py SOS2_pmids.txt 5 10
```

The two runs create two different files for the results from the seven papers, `reach_stmts_0_5.pkl` (with statements from the first five papers) and `reach_stmts_5_7.pkl` (with statements from the last two). Note that the results are pickled as a dict (rather than a list), with PMIDs as keys and lists of Statements as values.

Of course, what we really want is a single file containing all of the statements for the entire corpus. To get this, run:

```
python assemble_reach_stmts.py reach_stmts_*.pkl
```

The results will be stored in `reach_stmts.pkl`.

## 5.2.7 Running the whole pipeline with one script

If you want to run the whole pipeline in one go, you can run the script `submit_reading_pipeline.py` (in `indra/tools/reading`) on a cluster of Amazon EC2 nodes. The script divides up the jobs evenly among the nodes and cores. Usage:

```
python submit_reading_pipeline.py pmid_list tmp_dir num_nodes num_cores_per_node
```

For example if you have a cluster with 8 c3.8xlarge nodes with 32 VCPUs each, you would call it with:

```
python submit_reading_pipeline.py SOS2_pmids.txt my_tmp_dir 8 32
```

The script submits the jobs to the scheduler with appropriate dependencies such that the REACH reading step completes first, then the INDRA processing step, and then the final assembly into a single pickle file.

## 5.3 Large-Scale Machine Reading with Amazon Batch

The following doc describes the steps involved in reading a large numbers of papers in parallel on Amazon EC2 using REACH, caching the JSON output on Amazon S3, processing the REACH output into INDRA Statements, and then caching the statements also on S3. Prerequisites for doing the following are:

- An Amazon S3 bucket containing full text contents for papers, keyed by Pubmed ID (creation of this S3 repository will be described in another tutorial).
- Amazon AWS credentials for using AWS Batch.

- A corpus of PMIDs (see *Large-Scale Machine Reading with Starcluster* for information on how to assemble this)
- Optional: Elsevier text and data mining API key and institution key for subscriber access to Elsevier full text content.

### 5.3.1 How it Works

- The reading pipeline makes use of a Docker image that contains INDRA and all necessary dependencies, including REACH, Kappa, PySB, etc. The Docker file for this image is available at: [https://github.com/johnbachman/indra\\_docker](https://github.com/johnbachman/indra_docker).
- The INDRA Docker image is built by AWS Codebuild and pushed to Amazon’s EC2 Container Service (ECS), where it is available via the Repository URI:

```
292075781285.dkr.ecr.us-east-1.amazonaws.com/indra
```

- An AWS Batch *Compute Environment* named “run\_reach” is configured to use this Docker image for handling AWS jobs. This compute environment is configured to use only Spot instances with a maximum spot price of 40% of the on-demand price, and 16 vCPUs.
- An AWS *Job Queue*, “run\_reach\_queue”, is configured to use instances of the “run\_reach” Compute Environment.
- An AWS *Job Definition*, “run\_reach\_jobdef”, is configured to run in the “run\_reach\_queue”, and to use 16 vCPUs and 30GiB of RAM.
- Reading jobs are submitted by running the script:

```
python -m indra.tools.reading.submit_reading_pipeline_aws read [args]
```

which, given a list of PMIDs:

- Copies the PMID list to the key `reading_results/[job_name]/pmids` on Amazon S3
- Breaks the list up into chunks (e.g., of 3000 PMIDs) and submits an AWS Batch job for each (using the “run\_reach\_jobdef” definition as a template).
- The ECS instance created by the AWS Batch job runs the script `indra.tools.reading.run_reach_on_pmids_aws`, which:
  - Checks for cached content on Amazon S3
  - If the PMID has not been read by the current version of REACH, checks for content
  - If the content is not available, downloads the content using the INDRA literature client, and caches on S3
  - The content to be read is downloaded to the `/tmp` directory of the instance
  - REACH is run using the command-line interface (RunReachCLI), and configured to read the papers in the `/tmp` directory using all of the vCPUs on the instance
  - When done, the result REACH JSON in the output folder is uploaded to S3
  - The JSON for both the previously and newly read papers is processed in parallel to INDRA Statements
  - The resulting subset of statements for the given range of papers is cached on S3 at `reading_results/[job_name]/stmts/[start_ix]_[end_ix].pkl`. This set of statements takes the form of a pickled (protocol 3) Python dict with PMIDs as keys and lists of INDRA Statements as values.
  - In addition, information about the sources of content available for each PMID is cached for each PMID subset at `reading_results/[job_name]/content_types/[start_ix]_[end_ix].pkl`.

- When the reading jobs for each of the subsets of PMIDs have been completed and cached on S3, the final combined set of statements (and combined information on content sources) can be assembled using:

```
python -m indra.tools.reading.submit_reading_pipeline_aws combine [job_name]
```

- This script submits an AWS batch job for a machine with 1 vCPU but a large amount of memory (60GiB)
  - The job runs the script `indra.tools.reading.assemble_reach_stmts_aws`, which unpickles the results from all of the PMID subsets, combines them, and stores them on S3
  - The resulting files are obtainable from S3 at `reading_results/[job_name]/stmts.pkl` and `reading_results/[job_name]/content_types.pkl`.
- To run the entire pipeline, where the assembly of the combined set of statements is automatically performed after the reading step is completed, run:

```
python -m indra.tools.reading.submit_reading_pipeline_aws full [args]
```

## 5.4 Assembling everything known about a particular gene

Assume you are interested in collecting all mechanisms that a particular gene is involved in. Using INDRA, it is possible to collect everything curated about the gene in pathway databases and then read all the accessible literature discussing the gene of interest. This knowledge is aggregated as a set of INDRA Statements which can then be assembled into several different model and network formats and possibly shared online.

For the sake of example, assume that the gene of interest is TMEM173.

It is important to use the standard HGNC gene symbol of the gene throughout the example (this information is available on <http://www.genenames.org/> or <http://www.uniprot.org/>) - arbitrary synonyms will not work!

### 5.4.1 Collect mechanisms from PathwayCommons and the BEL Large Corpus

We first collect Statements from the PathwayCommons database via INDRA's BioPAX API and then collect Statements from the BEL Large Corpus via INDRA's BEL API.

```
from indra.tools.gene_network import GeneNetwork

gn = GeneNetwork(['TMEM173'])
biopax_stmts = gn.get_biopax_stmts()
bel_stmts = gn.get_bel_stmts()
```

at this point `biopax_stmts` and `bel_stmts` are two lists of INDRA Statements.

### 5.4.2 Collect a list of publications that discuss the gene of interest

We next use INDRA's literature client to find PubMed IDs (PMIDs) that discuss the gene of interest. To find articles that are annotated with the given gene, INDRA first looks up the Entrez ID corresponding to the gene name and then finds associated publications.

```
from indra import literature

pmids = literature.pubmed_client.get_ids_for_gene('TMEM173')
```

The variable `pmid` now contains a list of PMIDs associated with the gene.

### 5.4.3 Get the full text or abstract corresponding to the publications

Next we use INDRA's literature client to fetch the full text (if available) or the abstract corresponding to the PMIDs we have just collected.

```
from indra import literature

paper_contents = {}
for pmid in pmids:
    content, content_type = literature.get_full_text(pmid, 'pmid')
    paper_contents[pmid] = (content, content_type)
```

We now have a dictionary called *paper\_contents* which stores the content and the content type of each PMID we looked up.

### 5.4.4 Read the content of the publications

We next run the REACH reading system on the publications. Depending on the content type, different calls need to be made via INDRA's REACH API.

```
from indra import literature
from indra.sources import reach

read_offline = True

literature_stmts = []
for pmid, (content, content_type) in paper_contents.items():
    rp = None
    print('Reading %s' % pmid)
    if content_type == 'abstract':
        rp = reach.process_text(content, citation=pmid, offline=read_offline)
    elif content_type == 'pmc_oa_xml':
        rp = reach.process_nxml_str(content, offline=read_offline)
    elif content_type == 'elsevier_xml':
        txt = literature.elsevier_client.extract_text(content)
        if txt:
            rp = reach.process_text(txt, citation=pmid, offline=read_offline)
    if rp is not None:
        literature_stmts += rp.statements
```

The list *literature\_stmts* now contains the results of all the statements that were read.

### 5.4.5 Combine all statements and run pre-assembly

```
from indra.tools import assemble_corpus

stmts = biopax_stmts + bel_stmts + literature_stmts

stmts = assemble_corpus.map_grounding(stmts)
stmts = assemble_corpus.map_sequence(stmts)
stmts = assemble_corpus.run_preassembly(stmts)
```

At this point *stmts* contains a list of Statements collected with grounding, sequences having been mapped, duplicates combined and less specific variants of statements hidden. It is possible to run other filters on the results such as to keep only human genes, remove Statements with ungrounded genes, or to keep only certain types of interactions.



### 5.4.6 Assemble the statements into a network model

```
from indra.assemblers import CxAssembler
from indra.databases import ndex_client

cxa = CxAssembler(stmts)
cx_str = cxa.make_model()
```

We can now upload this network to the Network Data Exchange (NDEx).

```
ndex_cred = {'user': 'myusername', 'password': 'xxx'}
network_id = ndex_client.create_network(cx_str, ndex_cred)
print(network_id)
```



Many functionalities of INDRA can be used via a REST API. This enables making use of INDRA's knowledge sources and assembly capabilities in a RESTful, platform independent fashion. The REST service is meant to be used locally (on a single machine or local network) and is currently not offered as a public web service by the creators of INDRA.

### 6.1 Installation

The REST service requires the *bottle* package to be installed in addition to all the other requirements of INDRA.

### 6.2 Launching the REST service

The REST service can be launched by running *api.py* in the *rest\_api* folder within *indra*.

### 6.3 Documentation

The specific end-points and input/output parameters offered by the REST API are documented in *rest\_api/docs/index.html*, which is accessible locally within the *indra* folder.



## CHAPTER 7

---

### Indices and tables

---

- `genindex`
- `modindex`
- `search`



**i**

indra.assemblers.cx\_assembler, 70  
indra.assemblers.english\_assembler, 71  
indra.assemblers.graph\_assembler, 72  
indra.assemblers.index\_card\_assembler,  
74  
indra.assemblers.pysb\_assembler, 67  
indra.assemblers.sbgm\_assembler, 75  
indra.assemblers.sif\_assembler, 73  
indra.belief, 63  
indra.databases.biogrid\_client, 46  
indra.databases.cbio\_client, 48  
indra.databases.chebi\_client, 45  
indra.databases.context\_client, 46  
indra.databases.hgnc\_client, 41  
indra.databases.ndex\_client, 47  
indra.databases.relevance\_client, 46  
indra.databases.uniprot\_client, 42  
indra.db, 91  
indra.db.manage\_content, 93  
indra.explanation.model\_checker, 76  
indra.literature, 51  
indra.literature.crossref\_client, 53  
indra.literature.elsevier\_client, 54  
indra.literature.pmc\_client, 53  
indra.literature.pubmed\_client, 51  
indra.mechlinker, 64  
indra.preassembler, 54  
indra.preassembler.grounding\_mapper, 59  
indra.preassembler.hierarchy\_manager,  
61  
indra.preassembler.sitemapper, 59  
indra.sources.bel.bel\_api, 25  
indra.sources.bel.processor, 26  
indra.sources.biopax.biopax\_api, 30  
indra.sources.biopax.pathway\_commons\_client,  
33  
indra.sources.biopax.processor, 31  
indra.sources.ndex\_cx.ndex\_cx\_api, 39  
indra.sources.ndex\_cx.processor, 40  
indra.sources.reach.processor, 36  
indra.sources.reach.reach\_api, 34  
indra.sources.reach.reach\_reader, 36  
indra.sources.trips.processor, 37  
indra.sources.trips.trips\_api, 37  
indra.sources.trips.trips\_client, 39  
indra.statements, 13  
indra.tools.assemble\_corpus, 78  
indra.tools.executable\_subnetwork, 88  
indra.tools.gene\_network, 86  
indra.tools.incremental\_model, 88  
indra.tools.machine, 89





## A

- Acetylation (class in `indra.statements`), 15
- Activation (class in `indra.statements`), 15
- ActiveForm (class in `indra.statements`), 15
- ActivityCondition (class in `indra.statements`), 16
- `add_default_initial_conditions()` (`indra.assemblers.pysb_assembler.PysbAssembler` method), 67
- `add_rule_to_model()` (in module `indra.assemblers.pysb_assembler`), 69
- `add_statements()` (`indra.assemblers.cx_assembler.CxAssembler` method), 70
- `add_statements()` (`indra.assemblers.english_assembler.EnglishAssembler` method), 72
- `add_statements()` (`indra.assemblers.graph_assembler.GraphAssembler` method), 73
- `add_statements()` (`indra.assemblers.index_card_assembler.IndexCardAssembler` method), 74
- `add_statements()` (`indra.assemblers.pysb_assembler.PysbAssembler` method), 68
- `add_statements()` (`indra.assemblers.sbg_n_assembler.SBGNAAssembler` method), 75
- `add_statements()` (`indra.explanation.model_checker.ModelChecker` method), 76
- `add_statements()` (`indra.mechlinker.MechLinker` method), 65
- `add_statements()` (`indra.preassembler.Preassembler` method), 55
- `add_statements()` (`indra.tools.incremental_model.IncrementalModel` method), 88
- Agent (class in `indra.statements`), 16
- `agent_map` (`indra.preassembler.grounding_mapper.GroundingMapper` attribute), 59
- `agent_set` (`indra.assemblers.pysb_assembler.PysbAssembler` attribute), 67
- AgentState (class in `indra.mechlinker`), 64
- `all_direct_stmts` (`indra.sources.bel.processor.BelProcessor` attribute), 26
- `all_events` (`indra.sources.reach.processor.ReachProcessor` attribute), 36
- `all_indirect_stmts` (`indra.sources.bel.processor.BelProcessor` attribute), 26
- `api_ruler` (`indra.sources.reach.reach_reader.ReachReader` attribute), 36
- `apply_to()` (`indra.mechlinker.AgentState` method), 64
- `assembled_stmts` (`indra.tools.incremental_model.IncrementalModel` attribute), 88
- Autophosphorylation (class in `indra.statements`), 17

## B

- BaseAgent (class in `indra.mechlinker`), 64
- BaseAgentSet (class in `indra.mechlinker`), 64
- `basename` (`indra.tools.gene_network.GeneNetwork` attribute), 86
- BeliefEngine (class in `indra.belief`), 63
- BelProcessor (class in `indra.sources.bel.processor`), 26
- BiopaxProcessor (class in `indra.sources.biopax.processor`), 31
- `bound_conditions` (`indra.mechlinker.AgentState` attribute), 64
- BoundCondition (class in `indra.statements`), 17
- `build_transitive_closures()` (`indra.preassembler.hierarchy_manager.HierarchyManager` method), 62

## C

- `check_model()` (`indra.explanation.model_checker.ModelChecker` method), 76
- `check_statement()` (`indra.explanation.model_checker.ModelChecker` method), 76
- `cleanup` (`indra.sources.reach.processor.ReachProcessor` attribute), 36
- `combine_duplicate_stmts()` (`indra.preassembler.Preassembler` static method), 55
- `combine_duplicates()` (`indra.preassembler.Preassembler` method), 55
- `combine_related()` (`indra.preassembler.Preassembler` method), 55

- commit() (indra.db.DatabaseManager method), 91
- Complex (class in indra.statements), 17
- complex\_monomers\_default() (in module indra.assemblers.pysb\_assembler), 69
- complex\_monomers\_one\_step() (in module indra.assemblers.pysb\_assembler), 69
- Conversion (class in indra.statements), 18
- converted\_direct\_stmts (indra.sources.bel.processor.BelProcessor attribute), 26
- converted\_indirect\_stmts (indra.sources.bel.processor.BelProcessor attribute), 26
- copy() (indra.db.DatabaseManager method), 92
- copy\_default\_config() (in module indra.tools.machine), 91
- copy\_into\_db() (indra.db.manage\_content.Manager method), 93
- create\_network() (in module indra.databases.ndex\_client), 47
- create\_tables() (indra.db.DatabaseManager method), 92
- Cx (indra.assemblers.cx\_assembler.CxAssembler attribute), 70
- CxAssembler (class in indra.assemblers.cx\_assembler), 70
- ## D
- DatabaseManager (class in indra.db), 91
- Deacetylation (class in indra.statements), 18
- DecreaseAmount (class in indra.statements), 18
- Defarnesylation (class in indra.statements), 18
- default\_mapper (in module indra.preassembler.sitemapper), 61
- degenerate\_stmts (indra.sources.bel.processor.BelProcessor attribute), 26
- Degeranylgeranylation (class in indra.statements), 18
- Deglycosylation (class in indra.statements), 18
- Dehydroxylation (class in indra.statements), 18
- delete\_all() (indra.db.DatabaseManager method), 92
- Demethylation (class in indra.statements), 18
- Demyristoylation (class in indra.statements), 19
- Depalmitoylation (class in indra.statements), 19
- Dephosphorylation (class in indra.statements), 19
- Deribosylation (class in indra.statements), 19
- Desumoylation (class in indra.statements), 19
- Deubiquitination (class in indra.statements), 19
- doc\_id (indra.sources.trips.processor.TripsProcessor attribute), 38
- doi\_query() (in module indra.literature.crossref\_client), 53
- download\_article() (in module indra.literature.elsevier\_client), 54
- download\_file() (indra.db.manage\_content.NihFtpClient method), 94
- drop\_tables() (indra.db.DatabaseManager method), 92
- dump\_statements() (in module indra.tools.assemble\_corpus), 78
- dump\_stmt\_strings() (in module indra.tools.assemble\_corpus), 78
- ## E
- edge\_properties (indra.assemblers.graph\_assembler.GraphAssembler attribute), 72
- EnglishAssembler (class in indra.assemblers.english\_assembler), 71
- Evidence (class in indra.statements), 19
- existing\_edges (indra.assemblers.graph\_assembler.GraphAssembler attribute), 72
- existing\_nodes (indra.assemblers.graph\_assembler.GraphAssembler attribute), 72
- expand\_families() (in module indra.tools.assemble\_corpus), 78
- expand\_pagination() (in module indra.literature.pubmed\_client), 51
- export\_model() (indra.assemblers.pysb\_assembler.PysbAssembler method), 68
- export\_sbgn() (in module indra.assemblers.pysb\_assembler), 69
- extracted\_events (indra.sources.trips.processor.TripsProcessor attribute), 38
- ## F
- Farnesylation (class in indra.statements), 20
- filter\_belief() (in module indra.tools.assemble\_corpus), 79
- filter\_by\_type() (in module indra.tools.assemble\_corpus), 79
- filter\_direct() (in module indra.tools.assemble\_corpus), 79
- filter\_enzyme\_kinase() (in module indra.tools.assemble\_corpus), 79
- filter\_evidence\_source() (in module indra.tools.assemble\_corpus), 79
- filter\_gene\_list() (in module indra.tools.assemble\_corpus), 80
- filter\_genes\_only() (in module indra.tools.assemble\_corpus), 80
- filter\_grounded\_only() (in module indra.tools.assemble\_corpus), 80
- filter\_human\_only() (in module indra.tools.assemble\_corpus), 81
- filter\_inconsequential\_acts() (in module indra.tools.assemble\_corpus), 81
- filter\_inconsequential\_mods() (in module indra.tools.assemble\_corpus), 81
- filter\_mod\_nokinase() (in module indra.tools.assemble\_corpus), 82

filter\_mutation\_status() (in module indra.tools.assemble\_corpus), 82  
 filter\_no\_hypothesis() (in module indra.tools.assemble\_corpus), 82  
 filter\_pmids() (in module indra.literature.pmc\_client), 53  
 filter\_query() (indra.db.DatabaseManager method), 92  
 filter\_text\_content() (indra.db.manage\_content.PmcManager method), 94  
 filter\_text\_refs() (indra.db.manage\_content.PmcManager method), 94  
 filter\_top\_level() (in module indra.tools.assemble\_corpus), 82  
 filter\_transcription\_factor() (in module indra.tools.assemble\_corpus), 83  
 filter\_uuid\_list() (in module indra.tools.assemble\_corpus), 83  
 find\_entity (indra.preassembler.hierarchy\_manager.HierarchyManagement method), 32  
 attribute), 62  
 fix\_doi() (indra.db.manage\_content.Medline method), 93  
 flatten\_evidence() (in module indra.preassembler), 57  
 flatten\_stmts() (in module indra.preassembler), 57  
 ftp\_ls() (indra.db.manage\_content.NihFtpClient method), 94

## G

g (indra.sources.bel.processor.BelProcessor attribute), 26  
 Gap (class in indra.statements), 20  
 gather\_explicit\_activities() (indra.mechlinker.MechLinker method), 65  
 gather\_implicit\_activities() (indra.mechlinker.MechLinker method), 65  
 Gef (class in indra.statements), 20  
 gene\_list (indra.tools.gene\_network.GeneNetwork attribute), 86  
 GeneNetwork (class in indra.tools.gene\_network), 86  
 Geranylgeranylation (class in indra.statements), 20  
 get\_abstract() (in module indra.literature.elsevier\_client), 54  
 get\_abstract() (in module indra.literature.pubmed\_client), 52  
 get\_abstracts\_by\_pmids() (indra.db.DatabaseManager method), 92  
 get\_activating\_mods() (indra.sources.bel.processor.BelProcessor method), 26  
 get\_activating\_subs() (indra.sources.bel.processor.BelProcessor method), 26  
 get\_activation() (indra.sources.bel.processor.BelProcessor method), 27  
 get\_activation() (indra.sources.reach.processor.ReachProcessor method), 36  
 get\_activations() (indra.sources.trips.processor.TripsProcessor method), 38  
 get\_activations\_causal() (indra.sources.trips.processor.TripsProcessor method), 38  
 get\_activations\_stimulate() (indra.sources.trips.processor.TripsProcessor method), 38  
 get\_active\_forms() (indra.sources.trips.processor.TripsProcessor method), 38  
 get\_active\_forms\_state() (indra.sources.trips.processor.TripsProcessor method), 38  
 get\_active\_tables() (indra.db.DatabaseManager method), 92  
 get\_activity\_modification() (indra.sources.biopax.processor.BiopaxProcessor method), 32  
 get\_agent\_rule\_str() (in module indra.assemblers.pysb\_assembler), 69  
 get\_agents() (indra.sources.ndex\_cx.processor.NdexCxProcessor method), 40  
 get\_agents() (indra.sources.trips.processor.TripsProcessor method), 38  
 get\_all\_direct\_statements() (indra.sources.bel.processor.BelProcessor method), 27  
 get\_all\_events() (indra.sources.reach.processor.ReachProcessor method), 36  
 get\_all\_events() (indra.sources.trips.processor.TripsProcessor method), 38  
 get\_all\_indirect\_statements() (indra.sources.bel.processor.BelProcessor method), 27  
 get\_all\_pmids() (indra.db.DatabaseManager method), 92  
 get\_annotation() (in module indra.assemblers.pysb\_assembler), 69  
 get\_api\_ruler() (indra.sources.reach.reach\_reader.ReachReader method), 37  
 get\_article() (in module indra.literature.elsevier\_client), 54  
 get\_article\_xml (in module indra.literature.pubmed\_client), 52  
 get\_bel\_stmts() (indra.tools.gene\_network.GeneNetwork method), 86  
 get\_binding\_site\_name() (in module indra.assemblers.pysb\_assembler), 69  
 get\_biopax\_stmts() (indra.tools.gene\_network.GeneNetwork method), 86  
 get\_cancer\_studies() (in module indra.databases.cbio\_client), 48  
 get\_cancer\_types() (in module indra.databases.cbio\_client), 48

get\_case\_lists() (in module indra.databases.cbio\_client), 48  
 get\_ccle\_cna() (in module indra.databases.cbio\_client), 48  
 get\_ccle\_lines\_for\_mutation() (in module indra.databases.cbio\_client), 49  
 get\_ccle\_mrna() (in module indra.databases.cbio\_client), 49  
 get\_ccle\_mutations() (in module indra.databases.cbio\_client), 49  
 get\_chebi\_id\_from\_pubchem() (in module indra.databases.chebi\_client), 45  
 get\_children() (indra.preassembler.hierarchy\_manager.HierarchyManager method), 62  
 get\_columns() (indra.db.DatabaseManager method), 92  
 get\_complexes() (indra.sources.bel.processor.BelProcessor method), 27  
 get\_complexes() (indra.sources.biopax.processor.BiopaxProcessor method), 32  
 get\_complexes() (indra.sources.reach.processor.ReachProcessor method), 36  
 get\_complexes() (indra.sources.trips.processor.TripsProcessor method), 38  
 get\_composite\_activating\_mods() (indra.sources.bel.processor.BelProcessor method), 28  
 get\_conversions() (indra.sources.bel.processor.BelProcessor method), 28  
 get\_conversions() (indra.sources.biopax.processor.BiopaxProcessor method), 32  
 get\_create\_base\_agent() (indra.mechlinker.BaseAgentSet method), 64  
 get\_create\_parameter() (in module indra.assemblers.pysb\_assembler), 69  
 get\_csv\_as\_dict() (indra.db.manage\_content.NihFtpClient method), 94  
 get\_data\_from\_xml\_str() (indra.db.manage\_content.PmcManager method), 94  
 get\_defaults() (in module indra.db), 93  
 get\_degenerate\_statements() (indra.sources.bel.processor.BelProcessor method), 28  
 get\_degradations() (indra.sources.trips.processor.TripsProcessor method), 38  
 get\_dois (in module indra.literature.elsevier\_client), 54  
 get\_entrez\_id() (in module indra.databases.hgnc\_client), 41  
 get\_family\_members() (in module indra.databases.uniprot\_client), 42  
 get\_file() (indra.db.manage\_content.NihFtpClient method), 94  
 get\_full\_text() (in module indra.literature), 51  
 get\_fulltext\_links() (in module indra.literature.crossref\_client), 53  
 get\_gap() (indra.sources.biopax.processor.BiopaxProcessor method), 32  
 get\_gef() (indra.sources.biopax.processor.BiopaxProcessor method), 32  
 get\_gene\_name() (in module indra.databases.uniprot\_client), 42  
 get\_genetic\_profiles() (in module indra.databases.cbio\_client), 49  
 get\_heat\_kernel() (in module indra.databases.relevance\_client), 46  
 get\_hgnc\_entry (in module indra.databases.hgnc\_client), 41  
 get\_hgnc\_from\_entrez() (in module indra.databases.hgnc\_client), 41  
 get\_hgnc\_from\_mouse() (in module indra.databases.hgnc\_client), 41  
 get\_hgnc\_from\_rat() (in module indra.databases.hgnc\_client), 41  
 get\_hgnc\_id() (in module indra.databases.hgnc\_client), 41  
 get\_hgnc\_name() (in module indra.databases.hgnc\_client), 42  
 get\_id\_from\_mgi() (in module indra.databases.uniprot\_client), 43  
 get\_id\_from\_mnemonic() (in module indra.databases.uniprot\_client), 43  
 get\_id\_from\_rgd() (in module indra.databases.uniprot\_client), 43  
 get\_ids (in module indra.literature.pubmed\_client), 52  
 get\_ids\_for\_gene (in module indra.literature.pubmed\_client), 52  
 get\_im() (indra.explanation.model\_checker.ModelChecker method), 76  
 get\_issns\_for\_journal (in module indra.literature.pubmed\_client), 52  
 get\_metadata (in module indra.literature.crossref\_client), 53  
 get\_metadata\_for\_ids() (in module indra.literature.pubmed\_client), 52  
 get\_metadata\_from\_xml\_tree() (in module indra.literature.pubmed\_client), 52  
 get\_mgi\_id() (in module indra.databases.uniprot\_client), 43  
 get\_missing\_pmids() (indra.db.manage\_content.PmcManager method), 94  
 get\_mnemonic() (in module indra.databases.uniprot\_client), 43  
 get\_mod\_site\_name() (in module indra.assemblers.pysb\_assembler), 69  
 get\_model\_agents() (indra.tools.incremental\_model.IncrementalModel method), 88

`get_modifications()` (in `indra.sources.bel.processor.BelProcessor` method), 28  
`get_modifications()` (in `indra.sources.biopax.processor.BiopaxProcessor` method), 32  
`get_modifications()` (in `indra.sources.reach.processor.ReachProcessor` method), 36  
`get_modifications()` (in `indra.sources.trips.processor.TripsProcessor` method), 38  
`get_modifications_indirect()` (in `indra.sources.trips.processor.TripsProcessor` method), 38  
`get_monomer_pattern()` (in module `indra.assemblers.pysb_assembler`), 70  
`get_mouse_id()` (in module `indra.databases.hgnc_client`), 42  
`get_mouse_id()` (in module `indra.databases.uniprot_client`), 43  
`get_mutations()` (in module `indra.databases.cbio_client`), 50  
`get_mutations()` (in module `indra.databases.context_client`), 46  
`get_node_names()` (`indra.sources.ndex_cx.processor.NdexCxProcessor` method), 40  
`get_num_sequenced()` (in module `indra.databases.cbio_client`), 50  
`get_parents()` (`indra.preassembler.hierarchy_manager.HierarchyManager` method), 62  
`get_pmids()` (`indra.sources.ndex_cx.processor.NdexCxProcessor` method), 40  
`get_primary_db()` (in module `indra.db`), 93  
`get_primary_id()` (in module `indra.databases.uniprot_client`), 43  
`get_profile_data()` (in module `indra.databases.cbio_client`), 50  
`get_protein_expression()` (in module `indra.databases.context_client`), 46  
`get_pubchem_id()` (in module `indra.databases.chebi_client`), 45  
`get_publications()` (in module `indra.databases.biogrid_client`), 46  
`get_rat_id()` (in module `indra.databases.hgnc_client`), 42  
`get_rat_id()` (in module `indra.databases.uniprot_client`), 44  
`get_regulate_activities()` (in `indra.sources.biopax.processor.BiopaxProcessor` method), 32  
`get_regulate_amounts()` (in `indra.sources.biopax.processor.BiopaxProcessor` method), 32  
`get_regulate_amounts()` (in `indra.sources.reach.processor.ReachProcessor` method), 36  
`get_regulate_amounts()` (in `indra.sources.trips.processor.TripsProcessor` method), 38  
`get_relevant_nodes()` (in module `indra.databases.relevance_client`), 47  
`get_rgd_id()` (in module `indra.databases.uniprot_client`), 44  
`get_site_pattern()` (in module `indra.assemblers.pysb_assembler`), 70  
`get_statements()` (`indra.sources.ndex_cx.processor.NdexCxProcessor` method), 40  
`get_statements()` (`indra.tools.gene_network.GeneNetwork` method), 87  
`get_statements()` (`indra.tools.incremental_model.IncrementalModel` method), 89  
`get_statements_noprior()` (in `indra.tools.incremental_model.IncrementalModel` method), 89  
`get_statements_prior()` (in `indra.tools.incremental_model.IncrementalModel` method), 89  
`get_string()` (`indra.assemblers.graph_assembler.GraphAssembler` method), 73  
`get_subnetwork()` (in module `indra.tools.executable_subnetwork`), 88  
`get_syntheses()` (`indra.sources.trips.processor.TripsProcessor` method), 39  
`get_tarnames()` (`indra.db.DatabaseManager` method), 92  
`get_tarname_from_filename()` (in `indra.db.manage_content.Manuscripts` method), 93  
`get_title()` (in module `indra.literature.pubmed_client`), 53  
`get_transcription()` (`indra.sources.bel.processor.BelProcessor` method), 29  
`get_translocation()` (`indra.sources.reach.processor.ReachProcessor` method), 36  
`get_uncompressed_bytes()` (in `indra.db.manage_content.NihFtpClient` method), 94  
`get_uncond_agent()` (in module `indra.assemblers.pysb_assembler`), 70  
`get_uniprot_id()` (in module `indra.databases.hgnc_client`), 42  
`get_valid_location()` (in module `indra.statements`), 25  
`get_valid_residue()` (in module `indra.statements`), 25  
`get_values()` (`indra.db.DatabaseManager` method), 92  
`get_xml()` (in module `indra.sources.trips.trips_client`), 39  
`get_xml_file()` (`indra.db.manage_content.NihFtpClient` method), 94  
Glycosylation (class in `indra.statements`), 21  
gm (`indra.preassembler.grounding_mapper.GroundingMapper` attribute), 59

- grab\_session() (indra.db.DatabaseManager method), 92
- graph (indra.assemblers.graph\_assembler.GraphAssembler attribute), 72
- graph (indra.assemblers.sif\_assembler.SifAssembler attribute), 73
- graph (indra.preassembler.hierarchy\_manager.HierarchyManager attribute), 62
- graph\_properties (indra.assemblers.graph\_assembler.GraphAssembler attribute), 72
- graph\_query() (in module indra.sources.biopax.pathway\_commons\_client), 33
- GraphAssembler (class in indra.assemblers.graph\_assembler), 72
- grounded\_monomer\_patterns() (in module indra.assemblers.pysb\_assembler), 70
- GroundingMapper (class in indra.preassembler.grounding\_mapper), 59
- ## H
- has\_entry() (indra.db.DatabaseManager method), 92
- HasActivity (class in indra.statements), 21
- hierarchies (indra.preassembler.Preassembler attribute), 54
- HierarchyManager (class in indra.preassembler.hierarchy\_manager), 61
- Hydroxylation (class in indra.statements), 21
- ## I
- id\_lookup() (in module indra.literature), 51
- id\_lookup() (in module indra.literature.pmc\_client), 53
- IncreaseAmount (class in indra.statements), 21
- IncrementalModel (class in indra.tools.incremental\_model), 88
- IndexCardAssembler (class in indra.assemblers.index\_card\_assembler), 74
- indirect\_stmts (indra.sources.bel.processor.BelProcessor attribute), 26
- indra.assemblers.cx\_assembler (module), 70
- indra.assemblers.english\_assembler (module), 71
- indra.assemblers.graph\_assembler (module), 72
- indra.assemblers.index\_card\_assembler (module), 74
- indra.assemblers.pysb\_assembler (module), 67
- indra.assemblers.sbg\_n\_assembler (module), 75
- indra.assemblers.sif\_assembler (module), 73
- indra.belief (module), 63
- indra.databases.biogrid\_client (module), 46
- indra.databases.cbio\_client (module), 48
- indra.databases.chebi\_client (module), 45
- indra.databases.context\_client (module), 46
- indra.databases.hgnc\_client (module), 41
- indra.databases.ndex\_client (module), 47
- indra.databases.relevance\_client (module), 46
- indra.databases.uniprot\_client (module), 42
- indra.db (module), 91
- indra.db.manage\_content (module), 93
- indra.explanation.model\_checker (module), 76
- indra.literature (module), 51
- indra.literature.crossref\_client (module), 53
- indra.literature.elsevier\_client (module), 54
- indra.literature.pmc\_client (module), 53
- indra.literature.pubmed\_client (module), 51
- indra.mechlinker (module), 64
- indra.preassembler (module), 54
- indra.preassembler.grounding\_mapper (module), 59
- indra.preassembler.hierarchy\_manager (module), 61
- indra.preassembler.sitemapper (module), 59
- indra.sources.bel.bel\_api (module), 25
- indra.sources.bel.processor (module), 26
- indra.sources.biopax.biopax\_api (module), 30
- indra.sources.biopax.pathway\_commons\_client (module), 33
- indra.sources.biopax.processor (module), 31
- indra.sources.ndex\_cx.ndex\_cx\_api (module), 39
- indra.sources.ndex\_cx.processor (module), 40
- indra.sources.reach.processor (module), 36
- indra.sources.reach.reach\_api (module), 34
- indra.sources.reach.reach\_reader (module), 36
- indra.sources.trips.processor (module), 37
- indra.sources.trips.trips\_api (module), 37
- indra.sources.trips.trips\_client (module), 39
- indra.statements (module), 13
- indra.tools.assemble\_corpus (module), 78
- indra.tools.executable\_subnetwork (module), 88
- indra.tools.gene\_network (module), 86
- indra.tools.incremental\_model (module), 88
- indra.tools.machine (module), 89
- infer\_activations() (indra.mechlinker.MechLinker static method), 65
- infer\_active\_forms() (indra.mechlinker.MechLinker static method), 66
- infer\_complexes() (indra.mechlinker.MechLinker static method), 66
- infer\_modifications() (indra.mechlinker.MechLinker static method), 66
- Inhibition (class in indra.statements), 21
- insert() (indra.db.DatabaseManager method), 92
- insert\_db\_stmts() (indra.db.DatabaseManager method), 92
- insert\_many() (indra.db.DatabaseManager method), 92
- IntegrityError (class in indra.db.manage\_content), 93
- InvalidLocationError, 22
- InvalidResidueError, 22
- is\_human() (in module indra.databases.uniprot\_client), 44
- is\_mouse() (in module indra.databases.uniprot\_client), 44
- is\_rat() (in module indra.databases.uniprot\_client), 44

- is\_secondary() (in module in- dra.databases.uniprot\_client), 44
- isa() (indra.preassembler.hierarchy\_manager.HierarchyManager method), 62
- ## L
- LinkedStatement (class in indra.mechlinker), 65
- load\_prior() (indra.tools.incremental\_model.IncrementalModel method), 89
- load\_site\_map() (in module in- dra.preassembler.sitemapper), 61
- load\_statements() (in module in- dra.tools.assemble\_corpus), 83
- location (indra.mechlinker.AgentState attribute), 64
- ## M
- make\_model() (indra.assemblers.cx\_assembler.CxAssembler method), 71
- make\_model() (indra.assemblers.english\_assembler.EnglishAssembler method), 72
- make\_model() (indra.assemblers.graph\_assembler.GraphAssembler method), 73
- make\_model() (indra.assemblers.index\_card\_assembler.IndexCardAssembler method), 74
- make\_model() (indra.assemblers.pysb\_assembler.PysbAssembler method), 68
- make\_model() (indra.assemblers.sbg\_n\_assembler.SBG\_NAssembler method), 75
- make\_model() (indra.assemblers.sif\_assembler.SifAssembler method), 73
- Manager (class in indra.db.manage\_content), 93
- Manuscripts (class in indra.db.manage\_content), 93
- map\_grounding() (in module in- dra.tools.assemble\_corpus), 83
- map\_sequence() (in module in- dra.tools.assemble\_corpus), 84
- map\_sites() (indra.preassembler.sitemapper.SiteMapper method), 60
- MappedStatement (class in in- dra.preassembler.sitemapper), 59
- max\_path\_length (indra.explanation.model\_checker.PathResult attribute), 78
- max\_paths (indra.explanation.model\_checker.PathResult attribute), 78
- MechLinker (class in indra.mechlinker), 65
- Medline (class in indra.db.manage\_content), 93
- Methylation (class in indra.statements), 22
- ModCondition (class in indra.statements), 22
- model (indra.assemblers.english\_assembler.EnglishAssembler attribute), 72
- model (indra.assemblers.pysb\_assembler.PysbAssembler attribute), 67
- model (indra.sources.biopax.processor.BiopaxProcessor attribute), 31
- model\_to\_owl() (in module in- dra.sources.biopax.pathway\_commons\_client), 33
- ModelChecker (class in in- dra.explanation.model\_checker), 76
- Modification (class in indra.statements), 22
- mods (indra.mechlinker.AgentState attribute), 64
- mutations (indra.mechlinker.AgentState attribute), 64
- MutCondition (class in indra.statements), 23
- Myristoylation (class in indra.statements), 23
- ## N
- namespace\_from\_uri() (in module in- dra.sources.bel.processor), 29
- NdexCxProcessor (class in in- dra.sources.ndex\_cx.processor), 40
- network\_name (indra.assemblers.cx\_assembler.CxAssembler attribute), 70
- NihEnrClient (class in indra.db.manage\_content), 94
- NihManager (class in indra.db.manage\_content), 94
- node\_properties (indra.assemblers.graph\_assembler.GraphAssembler attribute), 72
- ## O
- owl\_to\_model() (in module in- dra.sources.biopax.pathway\_commons\_client), 33
- owl\_to\_model() (in module in- dra.sources.biopax.pathway\_commons\_client), 33
- ## P
- Palmitoylation (class in indra.statements), 23
- par\_to\_sec (indra.sources.trips.processor.TripsProcessor attribute), 38
- paragraphs (indra.sources.trips.processor.TripsProcessor attribute), 38
- parse\_identifiers\_url() (in module in- dra.assemblers.pysb\_assembler), 70
- partof() (indra.preassembler.hierarchy\_manager.HierarchyManager method), 62
- path\_found (indra.explanation.model\_checker.PathResult attribute), 77
- path\_metrics (indra.explanation.model\_checker.PathResult attribute), 77
- PathMetric (class in indra.explanation.model\_checker), 77
- PathResult (class in indra.explanation.model\_checker), 77
- paths (indra.explanation.model\_checker.PathResult attribute), 77
- Phosphorylation (class in indra.statements), 23
- PmcManager (class in indra.db.manage\_content), 94
- PmcOA (class in indra.db.manage\_content), 95

- policies (indra.assemblers.pysb\_assembler.PysbAssembler attribute), 67  
 populate() (indra.db.manage\_content.Manager method), 93  
 populate() (indra.db.manage\_content.Medline method), 93  
 populate() (indra.db.manage\_content.PmcManager method), 95  
 preassemble() (indra.tools.incremental\_model.IncrementalModel method), 89  
 Preassembler (class in indra.preassembler), 54  
 print\_boolean\_net() (indra.assemblers.sif\_assembler.SifAssembler method), 73  
 print\_cx() (indra.assemblers.cx\_assembler.CxAssembler method), 71  
 print\_event\_statistics() (indra.sources.reach.processor.ReachProcessor method), 36  
 print\_loopy() (indra.assemblers.sif\_assembler.SifAssembler method), 74  
 print\_model() (indra.assemblers.index\_card\_assembler.IndexCardAssembler method), 74  
 print\_model() (indra.assemblers.pysb\_assembler.PysbAssembler method), 68  
 print\_model() (indra.assemblers.sbgm\_assembler.SBGmAssembler method), 75  
 print\_model() (indra.assemblers.sif\_assembler.SifAssembler method), 74  
 print\_statement\_coverage() (indra.sources.bel.processor.BelProcessor method), 29  
 print\_statements() (indra.sources.bel.processor.BelProcessor method), 29  
 print\_statements() (indra.sources.biopax.processor.BiopaxProcessor method), 33  
 prior\_probs (indra.belief.BeliefEngine attribute), 63  
 process\_belrdf() (in module indra.sources.bel.bel\_api), 25  
 process\_cx() (in module indra.sources.ndex\_cx.ndex\_cx\_api), 39  
 process\_cx\_file() (in module indra.sources.ndex\_cx.ndex\_cx\_api), 39  
 process\_json\_file() (in module indra.sources.reach.reach\_api), 34  
 process\_json\_str() (in module indra.sources.reach.reach\_api), 34  
 process\_model() (in module indra.sources.biopax.biopax\_api), 30  
 process\_ndex\_neighborhood() (in module indra.sources.bel.bel\_api), 25  
 process\_ndex\_network() (in module indra.sources.ndex\_cx.ndex\_cx\_api), 40  
 process\_nxml\_file() (in module indra.sources.reach.reach\_api), 34  
 process\_nxml\_str() (in module indra.sources.reach.reach\_api), 34  
 process\_owl() (in module indra.sources.biopax.biopax\_api), 30  
 process\_pc\_neighborhood() (in module indra.sources.biopax.biopax\_api), 30  
 process\_pc\_pathsbetween() (in module indra.sources.biopax.biopax\_api), 30  
 process\_pc\_pathsfromto() (in module indra.sources.biopax.biopax\_api), 31  
 process\_pmc() (in module indra.sources.reach.reach\_api), 35  
 process\_pubmed\_abstract() (in module indra.sources.reach.reach\_api), 35  
 process\_text() (in module indra.sources.reach.reach\_api), 35  
 process\_text() (in module indra.sources.trips.trips\_api), 37  
 process\_xml() (in module indra.sources.trips.trips\_api), 37  
 process\_xml\_from\_twg() (in module indra.preassembler.grounding\_mapper), 59  
 prior\_influence\_map() (indra.explanation.model\_checker.ModelChecker method), 77  
 PysbAssembler (class in indra.assemblers.pysb\_assembler), 67
- ## Q
- query\_protein (in module indra.databases.uniprot\_client), 44
- ## R
- ReachProcessor (class in indra.sources.reach.processor), 36  
 ReachReader (class in indra.sources.reach.reach\_reader), 36  
 reduce\_activities() (in module indra.tools.assemble\_corpus), 84  
 reduce\_activities() (indra.mechlinker.MechLinker method), 66  
 RegulateActivity (class in indra.statements), 23  
 RegulateAmount (class in indra.statements), 24  
 related\_stmts (indra.preassembler.Preassembler attribute), 54  
 remove\_im\_params() (in module indra.explanation.model\_checker), 78  
 render\_stmt\_graph() (in module indra.preassembler), 58  
 replace\_activations() (indra.mechlinker.MechLinker method), 66  
 replace\_complexes() (indra.mechlinker.MechLinker method), 67



- require\_active\_forms() (indra.mechlinker.MechLinker method), 67
- result\_code (indra.explanation.model\_checker.PathResult attribute), 77
- results (indra.tools.gene\_network.GeneNetwork attribute), 86
- ret\_file() (indra.db.manage\_content.NihFtpClient method), 94
- Ribosylation (class in indra.statements), 24
- run\_preassembly() (in module indra.tools.assemble\_corpus), 84
- run\_preassembly() (indra.tools.gene\_network.GeneNetwork method), 87
- run\_preassembly\_duplicate() (in module indra.tools.assemble\_corpus), 85
- run\_preassembly\_related() (in module indra.tools.assemble\_corpus), 85
- ## S
- sample\_statements() (in module indra.belief), 64
- save() (indra.tools.incremental\_model.IncrementalModel method), 89
- save\_dot() (indra.assemblers.graph\_assembler.GraphAssembler method), 73
- save\_model() (indra.assemblers.cx\_assembler.CxAssembler method), 71
- save\_model() (indra.assemblers.index\_card\_assembler.IndexCardAssembler method), 74
- save\_model() (indra.assemblers.pysb\_assembler.PysbAssembler method), 68
- save\_model() (indra.assemblers.sbg\_n\_assembler.SBGNAAssembler method), 75
- save\_model() (indra.assemblers.sif\_assembler.SifAssembler method), 74
- save\_model() (indra.sources.biopax.processor.BiopaxProcessor method), 33
- save\_pdf() (indra.assemblers.graph\_assembler.GraphAssembler method), 73
- save\_rst() (indra.assemblers.pysb\_assembler.PysbAssembler method), 68
- save\_xml() (in module indra.sources.trips.trips\_client), 39
- sbg\_n (indra.assemblers.sbg\_n\_assembler.SBGNAAssembler attribute), 75
- SBGNAAssembler (class in indra.assemblers.sbg\_n\_assembler), 75
- score\_paths() (indra.explanation.model\_checker.ModelChecker method), 77
- select\_all() (indra.db.DatabaseManager method), 92
- select\_one() (indra.db.DatabaseManager method), 93
- SelfModification (class in indra.statements), 24
- send\_query() (in module indra.sources.trips.trips\_client), 39
- send\_request (in module indra.databases.cbio\_client), 51
- send\_request() (in module indra.databases.ndex\_client), 47
- sentences (indra.sources.trips.processor.TripsProcessor attribute), 38
- set\_base\_initial\_condition() (in module indra.assemblers.pysb\_assembler), 70
- set\_context() (indra.assemblers.cx\_assembler.CxAssembler method), 71
- set\_context() (indra.assemblers.pysb\_assembler.PysbAssembler method), 69
- set\_extended\_initial\_condition() (in module indra.assemblers.pysb\_assembler), 70
- set\_hierarchy\_probs() (indra.belief.BeliefEngine method), 63
- set\_linked\_probs() (indra.belief.BeliefEngine method), 63
- set\_prior\_probs() (indra.belief.BeliefEngine method), 63
- show\_tables() (indra.db.DatabaseManager method), 93
- SifAssembler (class in indra.assemblers.sif\_assembler), 73
- SiteMapper (class in indra.preassembler.sitemapper), 59
- Statement (class in indra.statements), 24
- statements (indra.assemblers.cx\_assembler.CxAssembler attribute), 70
- statements (indra.assemblers.english\_assembler.EnglishAssembler attribute), 72
- statements (indra.assemblers.graph\_assembler.GraphAssembler attribute), 72
- statements (indra.assemblers.index\_card\_assembler.IndexCardAssembler attribute), 74
- statements (indra.assemblers.pysb\_assembler.PysbAssembler attribute), 67
- statements (indra.assemblers.sbg\_n\_assembler.SBGNAAssembler attribute), 75
- statements (indra.sources.bel.processor.BelProcessor attribute), 26
- statements (indra.sources.biopax.processor.BiopaxProcessor attribute), 32
- statements (indra.sources.ndex\_cx.processor.NdexCxProcessor attribute), 40
- statements (indra.sources.reach.processor.ReachProcessor attribute), 36
- statements (indra.sources.trips.processor.TripsProcessor attribute), 38
- stmt\_from\_rule() (in module indra.explanation.model\_checker), 78
- stmts (indra.preassembler.Preassembler attribute), 54
- stmts (indra.tools.incremental\_model.IncrementalModel attribute), 88
- strip\_agent\_context() (in module indra.tools.assemble\_corpus), 85
- Sumoylation (class in indra.statements), 24

## T

term\_from\_uri() (in module indra.sources.bel.processor),  
29

to\_graph() (indra.statements.Statement method), 24

to\_json() (indra.statements.Statement method), 24

Translocation (class in indra.statements), 24

Transphosphorylation (class in indra.statements), 25

tree (indra.sources.reach.processor.ReachProcessor at-  
tribute), 36

tree (indra.sources.trips.processor.TripsProcessor at-  
tribute), 37

TripsProcessor (class in indra.sources.trips.processor), 37

## U

Ubiquitination (class in indra.statements), 25

unique\_stmts (indra.preassembler.Preassembler at-  
tribute), 54

update() (indra.db.manage\_content.Manager method), 93

update\_network() (in module in-  
dra.databases.ndex\_client), 47

upload\_archive() (indra.db.manage\_content.PmcManager  
method), 95

upload\_batch() (indra.db.manage\_content.PmcManager  
method), 95

upload\_model() (indra.assemblers.cx\_assembler.CxAssembler  
method), 71

## V

verify\_location() (in module in-  
dra.databases.uniprot\_client), 45

verify\_modification() (in module in-  
dra.databases.uniprot\_client), 45