

Data Model Glossary

Overview

Reactome is [a slot-and-frame knowledge representation](#) of a network of molecular transformations mediated by gene products. Classes (frames) are categories like reaction and protein. Related classes are grouped into superclasses to allow inheritance of shared attributes. For example, the "PhysicalEntity" superclass has child classes such as Complex and GenomeEncodedEntity. Classes have slots (attributes) whose values define properties of instances (e.g., the names and numbers of copies of the molecules that make up a complex). "Defining" attributes are used in the data model to distinguish instances of a class, to ensure that identical information, e.g., a water molecule localized to the cytosol, or a reaction by which a particular molecule from extracellular space is translocated into the cytosol mediated by a particular transport protein, is represented exactly once in Reactome. Many attributes of a class instance are instances of other classes. For example, a **Person** instance has a *surname* attribute that is a text string not taken from anyplace else but can also have an *affiliation* attribute that is an instance of the **affiliation** class, a class of *names* and *addresses* of institutions. In this glossary, **boldface** type indicates use of a word or phrase as a class name (the **CatalystActivity** of a particular enzyme) and *italic* type indicates use as an attribute (the *catalystActivity* that mediates a chemical transformation reaction). Words that can be either class names or attributes are defined under "Reactome Classes"; ones that are only attributes are defined under "Reactome attributes".

Glossary of Key Reactome Classes

AbstractModifiedResidue [superclass]

Used to maintain logical integrity of data model, not used for manual annotation. The superclass in the Reactome data model from which the GeneticallyModifiedResidue, TranscriptionalModification, and TranslationalModification classes inherit their shared attributes. Describes a modification to reference sequence. The three subclasses of modifications are GeneticallyModifiedResidue, TranscriptionalModification, TranslationalModification.

GeneticallyModifiedResidue [superclass]

Used to maintain logical integrity of data model, not used for manual annotation. Subclass of abstractModifiedResidue from which the FragmentModification and ReplacedResidue classes inherit their shared attributes. Describes a protein sequence change resulting from a genetic modification

FragmentModification [superclass]

Used to maintain logical integrity of data model, not used for manual annotation. Children of this class (FragmentDeletionModification, FragmentInsertionModification and FragmentReplacedModification) are used to capture the deletion, insertion or replacement of consecutive amino acid residues in an EWAS in a disease variant of a gene product.

FragmentDeletionModification

The deletion of a continuous fragment of an EntityWithAccessionedSequence, e.g. deletion of amino acid residues 30 to 297 in the EGFR mutant protein EGFRvIII in glioblastoma. Attributes of the FragmentDeletionModification instance are the *referenceSequence* (UniProt P00533 EGFR in the case of EGFRvIII), the *startPositionInReferenceSequence*, which is the first amino acid of the deleted fragment (30 in the case of EGFRvIII), and the *endPositionInReferenceSequence*, which equals the last amino acid of the deleted fragment (297 in the case of EGFRvIII).

FragmentInsertionModification

Used to describe the modification of an EntityWithAccessionedSequence through insertion of a continuous fragment. This class is used to annotate variants EWASs containing duplications or fusions. A duplication is defined as the insertion of a stretch of WT amino-acid(s) that is the direct repeat of the amino-acid(s) immediately 5' to the inserted sequence. For duplications, the *referenceSequence* specified in the FragmentInsertionModification record is the same as the *referenceEntity* specified in the EWAS record.

Fusions are the result of the insertion of a continuous stretch of amino-acids derived from a second EWAS at a specified residue in a first EWAS. For fusions, the *referenceSequence* specified in the FragmentInsertionModification record is the *referenceEntity* for the C-terminal fusion partner, and is different from the *referenceEntity* specified in the EWAS record, which represents the N-terminal partner.

FragmentReplacedModification

The modification of an EntityWithAccessionedSequence through the replacement of the WT amino-acid sequence with a specified alternate sequence of amino-acids as the result of an insertion, fusion, indel, frameshift or extension of the WT sequence. Details of these annotations vary depending on the type of replacement, and are outlined in the Curator Guide. Unlike FragmentDeletionModification and FragmentInsertionModification, FragmentReplacementModification requires that the curator explicitly annotate the single letter amino-acid sequence that arises in the variant protein. Note also that, by Reactome convention, insertions are captured by FragmentReplacedModification, not FragmentInsertionModification. In the FragmentReplaceModification record for an insertion, the *alteredAminoAcidFragment* includes both the flanking WT 5'- and 3' amino-acids as well as the inserted amino-acids. For example, in the variant EGFR D770_N771insNPG, amino acids NPG are inserted between

D770 and N771, and the specified *alteredAminoAcidFragment* in the *FragmentReplacedModification* is DNPGN.

ReplacedResidue

A subclass of *GeneticallyModifiedResidue*, *ReplacedResidue* describes the replacement of an amino acid residue of a UniProt *referenceSequence* with a different amino acid residue or a stop codon as the result of a mutation, germline or somatic, in the corresponding gene. Class attributes identify the *coordinate* of the affected residue in the protein sequence, the identity of the residue specified in the UniProt sequence (first *psiMOD* attribute value), and the identity of the replacement residue (second *psiMOD* attribute value).

NonsenseMutation

A subclass of the **ReplacedResidue** class, **NonsenseMutation** describes the *coordinate* at which a protein *referenceSequence* is truncated due to a nonsense mutation. In the *NonsenseMutation* record, a single *psiMod* is specified, indicating as an amino acid removal the amino acid that is altered to generate a stop codon.

TranscriptionalModification [superclass]

Used to maintain logical integrity of data model, not used for manual annotation. Subclass of *AbstractModifiedResidue*. Describes an RNA sequence modification.

ModifiedNucleotide

Describes the location and chemical nature of a nucleotide modification in an RNA sequence.

TranslationalModification [superclass]

Used to maintain logical integrity of data model, not used for manual annotation. Subclass of **AbstractModifiedResidue**. Includes children **CrosslinkedResidue**, **GroupModifiedResidue** and **ModifiedResidue**. Describes a post translational chemical modification to a protein. Modifications are cross referenced to the psi-MOD ontology of terms that describe protein chemical modifications.

CrosslinkedResidue [superclass]

Used to maintain logical integrity of data model, not used for manual annotation. Subclass of **TranslationalModification**. Two subclasses, **Inter-** and **IntraChainCrosslinkedResidue**, are used to annotate disulfide, thioester or other protein crosslinks, including to ie sumoyl or ubiquitin groups.

InterChainCrosslinkedResidue

Subclass of **CrosslinkedResidue**. Used to annotate cross-linking between two different peptide chains.

IntraChainCrosslinkedResidue

Subclass of **CrosslinkedResidue**. Used to annotate cross-linking between amino acid residues within the same peptide chain. **IntraChainCrosslinkedResidue** instances are made in pairs, one associated with the EWAS record of each of the cross-linked proteins, and the **IntraChainCrosslinkedResidue** records are associated with each other through their *equivalentTo* attribute.

GroupModifiedResidue

Subclass of **TranslationalModification**. Describes the modification of an amino acid residue in protein with a chemical entity that cannot be specified in atomic detail, e.g., the attachment of a dextrin or glycogen moiety to a tyrosine side chain in the protein glycogenin. Such incompletely specified chemical entities are beyond the scope of the psi-MOD ontology but are available in the ChEBI ontology. The psiMod attribute would take a value such as MOD:00166 O4'-glucosyl-L-tyrosine, which describes the linkage between the protein and the modifying group. The modifying group is then specified with a ReferenceGroup instance (ChEBI terms) such as CHEBI:28912 "limit dextrin", as in the example: limit dextrans on L-Tyrosine [ChEBI:17895] 194 of UniProt:P46976 GYG1.

ModifiedResidue

Subclass of TranslationalModification. Used to describe the translational or post-translational modification of an amino-acid, such as phosphorylation, gamma-carboxylation etc. The **ModifiedResidue** record defines the associated EWAS reference sequence along with the coordinate of the modified amino-acid and the nature of the modification, taking terms from psiMOD. **ModifiedResidue** class is used when the chemical natures of the modifying group can be specified in atomic detail with a psiMod term. This is in contrast to **GroupModifiedResidue**, which is used when the modification can not be completely described by a term from psi-MOD.

Affiliation

The name and address of an institution, e.g. [Cold Spring Harbor Laboratory / 1 Bungtown Road, Cold Spring Harbor NY USA](#). Any **Person** instance can have an *affiliation* attribute, but at present this information is recorded only for Reactome authors, reviewers, and curators.

CatalystActivity

Describes the catalyst in a **ReactionLikeEvent**: associates one *physicalEntity* with one *GO_molecularFunction*., e.g., [cysteine-type endopeptidase activity of active caspase-8](#). If the

physicalEntity is a **Complex** and a component of the complex mediates the molecularFunction, that component should be identified as the *activeUnit* of the **CatalystActivity**.

If a **PhysicalEntity** can enable multiple **MolecularFunctions**, a separate **CatalystActivity** instance is created for each..

ControlReference [superclass]

Used to maintain logical integrity of data model, not used for manual annotation. Associates one or more **Publications** with an instance (e.g., *catalystActivity*) of another class.

CatalystActivityReference

Associates a **CatalystActivity** instance with one or more **Publications**.

MarkerReference

Associates an RNA or protein **EntityWithAccessionedSequence** that has been identified as a *proteinMarker* or *RNAMarker* of a **Cell** with one or more **Publications**. *MarkerReference* is an attribute of the Cell class of PhysicalEntity.

RegulationReference

Associates any kind of **Regulation** instance with one or more **Publications**.

ControlledVocabulary [superclass]

Used to maintain logical integrity of data model, not used for manual annotation. Lists of words that can be used in attributes (slot values) in specific classes, optionally connected to one another with *is_a* relationships, created by Reactome (as opposed to ones imported from external reference resources like the Gene Ontology, ChEBI, or NCBI Taxonomy) Has three subclasses **DeletedControlledVocabulary** (for use in “reason” attribute of *_Deleted* instances), **NegativePrecedingEventReason** (for use in “reason” attribute of *NegativePrecedingEvent* instances) and *ReactionType* (for use in *reactionType* attribute of **ReactionLikeEvent**).

NegativePrecedingEvent

Used to flag an **Event** that should not be annotated as a value of another **Event**'s *precedingEvent* attribute.

NegativePrecedingEventReason

A subclass of *ControlledVocabulary* used to indicate why a computationally inferred causal connection (*precedingEvent*) between two **ReactionLikeEvents** was manually rejected by a curator (e.g. Abundant Molecules).

ReactionType

Terms used as attributes to classify **ReactionLikeEvents** in ways that support data mining and classification, e.g., in connection with the conversion of Reactome events into GO-CAM models (under development).

DrugActionType

Terms used as attributes to classify **ReactionLikeEvent** involving **drugs** to indicate the mode of action of the drug in the **ReactionLikeEvent**.

DatabaseObject [superclass]

Used to maintain logical integrity of data model, not used for manual annotation. The root term of the Reactome data model; parent to all classes of instances; not edited manually.

Databaselfentifier

Refers to an external database identifier. An instance of this class is used in for the *databaselfentifier* attribute as an optional *crossReference* attribute of a Reactome instance to associate it with the representation of that instance in an external *ReferenceDatabase*, in addition to the required *crossReferences* to external referenceDatabases. . E.g., the Reactome **SimpleEntity** instance [PQQ \(pyroquinoline quinone\)](#), *crossReferenced* to [ReferenceMolecule:29568 pyrroloquinoline quinone \[ChEBI:18315\]](#) has a *crossReference* attribute to the *databaselfentifier* [COMPOUND:C00113](#) that links the Reactome instance to the record for this molecule maintained by the [KEGG “compound” database](#).

_Deleted

Created when an instance is deleted from the database. Includes attributes for the *deletedInstanceDB_ID*, *replacementInstanceDB_IDs* (if relevant), reason that the instance was deleted (see *DeletedControlledVocabulary* below for reason definitions), as well as optional *curatorComment* about the deletion. Additional information about the deleted Instance is provided in the *_DeletedInstance* attribute.

DeletedControlledVocabulary

An attribute of the `_Deleted` instance (which is created when a curator opts to delete an instance). Describes the reason an instance was deleted from the database. Reasons are: Duplicated (instance was a duplicate of an existing instance), Killed not replaced (instance was found to be invalid and removed without replacement), Merged (an instances was merged with another instance), Obsoleted (no longer in use: see specified replacement or comment), Split (the instance was split into multiple instances).

`_DeletedInstance`

Attribute of `_Deleted` that includes the class of the deleted instance, `deletedInstanceDB_ID`, `deletedInstanceStableIdentifier`, name of the deleted instance, and species of the deleted instance (if relevant).

Disease

Disease encompasses terms that describe impairments of normal functions caused by pathogenic genetic mutations or environmental factors, such as microbial pathogens from an ExternalOntology, the Disease Ontology (DO) (<https://disease-ontology.org/>). Terms from the Disease class are used to populate the disease attribute which identifies any abnormal physicalEntity (**GenomeEncodedEntitiy**, **Complex**, EntitySets) and any abnormal event (**ReactionLikeEvent**, **Pathway**), and specifies the disease(s) with which these abnormal instances are associated. **Disease** instances include *identifier* and *name* attributes taken from DO.

EntityFunctionalStatus

EntityFunctionalStatus is used to mark a disease ReactionLikeEvent as either a gain- or a loss-of-function, and to describe the underlying reason for the change in behavior. The EntityFunctionalStatus has three attributes, *diseaseEntity*, the variant **EWAS** or **Complex** or *set* of variants, responsible for the phenotype, *normalEntity*, the wild-type instance corresponding to the disease **PhysicalEntity** and *functionalStatus*, that relates the molecular change to the functional one.

FunctionalStatus

Class used to describe the molecular basis of a change in function; combines a term from the *functionalStatusType* controlled vocabulary maintained by Reactome that identifies alterations in gene function (e.g., gain-of-function) and a *structuralVariant* term from the externalOntology Sequence Ontology that describes the nature of the genetic change (e.g., copy number gain).

Class **FunctionalStatus** is used to populate the *functionalStatus* attribute of **EntityFunctionalStatus**

Event [superclass]

Used to maintain logical integrity of data model; not used for manual annotation. Any transformation of input entities to output entities in one step (**ReactionlikeEvent**) or two or more steps (**Pathway**).

Pathway

A sequence of two or more causally connected **ReactionlikeEvents** and/or other **Pathways**, identified as *hasEvent* attributes of the pathway being annotated.

CellLineagePath

A **CellLineagePath**, organizationally similar to the **Pathway** subclass, is composed of other **CellLineagePath** instances or **CellDevelopmentSteps** as subevents that are connected through shared input/output Cells.

Both Cell Development Step and Cell Lineage Path are associated with a Gene Ontology (GO) biological process term and with a “tissue” term from UBERON, an ExternalOntology of Anatomy terms.

ReactionLikeEvent [superclass]

Used to maintain logical integrity of data model; not used for manual annotation.

ReactionLikeEvent is a superclass that is used to organize other concrete reaction types, such as **Reaction**, **Polymerization** and **BlackBoxEvent**. A molecular process in which one or more input *physicalEntities* are transformed in a single step into output *physicalEntities*, optionally mediated by a *catalystActivity* and subject to *regulation*. Kinds of transformation include chemical reactions, translocation of molecules from one **compartment** to another, association of molecules to form **complexes** and **polymers**, dissociation of **complexes** and polymers, and developmental/differentiation relationships among Cell instances.

BlackBoxEvent

This class can be used to describe an incompletely specified **ReactionLikeEvent** or series of RLEs. A BlackBoxEvent for a reaction is for one that is known to happen but whose molecular details are incompletely known, e.g., a partially purified protein catalyzes a reaction so no EWAS can be identified, or degradation of a protein yields unspecified polypeptide products. BBEs may also be used to represent a multistep process (such as the transcription and

translation of a gene into a protein) when the details don't need to be specified in full. A fully annotated sequence of events for a BBE may be depicted in full elsewhere in Reactome.

CellDevelopmentStep

A **CellDevelopmentStep** is a **ReactionlikeEvent** that contains **Cell** instances as its inputs and outputs. The input attribute represents the cell of origin, the starting point of this development event, and the output attribute represents the destination cell type, the outcome of the event. Additional CellDevelopmentStep attributes include regulators (molecules promoting or inhibiting the step) and required input components (input cell biomarkers required for the action of regulators). **CellDevelopmentStep** is also associated with a relevant Gene Ontology (GO) biological process term and with a “tissue” term from the UBERON, an ExternalOntology of Anatomy terms.

Depolymerisation

A subclass of **ReactionlikeEvent** that is used to annotate depolymerisation and follows the pattern: Polymer -> Polymer + Unit (reverse situation of Polymerisation).

FailedReaction

A **FailedReaction** instance is a step in a disease process that is directly affected by a loss-of-function (LoF) mutation (germline or somatic). This type of disease event has its normal **ReactionLikeEvent** counterpart (the reaction mediated by the un-mutated gene product) specified at its normalReaction attribute and is represented as having *inputs* (an abnormal physicalEntity plus any wild-type entities that are inputs in the normal reaction), but no *outputs*.

FailedReaction instances are labeled with disease term(s) that are used to populate the disease attribute of the associated abnormal **PhysicalEntity** (**GenomeEncodedEntity**, **Complex** or **EntitySet**).

.

Polymerisation

A subclass of reactionlikeEvent in which two or more identical molecules or complexes are assembled into a polymer.

Reaction

The transformation of input physical entities into output ones in a single step. Transformations include chemical changes, as in metabolism, the transport of an entity between locations, the association of entities to form a complex, and the dissociation of complexes. These transformations can be catalyzed and regulated.

EvidenceType

EvidenceType is a subclass of ExternalOntology, manually populated with [Gene Ontology evidence code](#) terms. Used only to flag inferred orthologous events created computationally as part of the release process (IEA, inferred by electronic annotation) and not for human manual annotation by curators.

ExternalOntology

An ExternalOntology is one developed and maintained outside of Reactome and used as-is by Reactome to align our annotation with the community vocabularies and logical relationships incorporated into these ontologies. This is a superclass and used for generalizing sharing attributes among multiple types of external ontologies. This class should not be used for instance creation.

Anatomy

Anatomy encompasses terms representing body parts, organs and tissues from an ExternalOntology, Uberon (<https://obophenotype.github.io/uberon/>). Anatomy terms are used to populate organ, tissue, and tissueLayer attributes of the Cell subclass of PhysicalEntity and tissue attribute of CellDevelopmentStep and CellLineagePath subclasses of Event.

CellType

CellType terms represent different categories of animal cells, characterized by their specific structure, function, and biochemistry, derived from an ExternalOntology, Cell Ontology (<https://obofoundry.org/ontology/cl.html>). CellType terms are used to populate the *cellType* attribute of Cell and EntityWithAccessionedSequence subclasses of PhysicalEntity.

Disease

Disease encompasses terms that describe impairments of normal functions caused by pathogenic genetic mutations or environmental factors, such as microbial pathogens. Terms are derived from EBI Disease Ontology (DO) (<https://disease-ontology.org/>) and are used to populate the disease attribute of abnormal physicalEntities (genomeEncodedEntity, complex, entitySet) and abnormal events (reactionLikeEvents, and pathways) to specify the disease(s) with which these abnormal instances are associated. Disease instances include *identifier* and *name* attributes taken from DO.

PsiMod

PsiMod encompasses terms representing post-translational chemical modifications of proteins derived from an ExternalOntology Protein Modification Ontology (MOD):

<https://ontobee.org/ontology/MOD>). The `psiMod` is an attribute of all the subclasses of `TranslationalModification`.

SequenceOntology

`SequenceOntology` is a local copy of portions of the ExternalOntology `Sequence Ontology` (<http://www.sequenceontology.org/>). It encompasses terms representing the variant features and attributes of biological sequence, used to populate the `structuralVariant` attribute of the subclasses of `FunctionalStatus`.

Figure

Figures for Reactome events are created in collaboration with Reactome's graphic illustrator. Curators needing to add a figure should consult with the illustrator and plans entered into to the illustration calendar. The `url` attribute of the **figure** instance holds the address of the image file in the Reactome CVS website repository, and the `figure` attribute of the data object to be illustrated holds the name of the **figure** instance. Example: `/figures/apoptotic_factor_responses.jpg` is this file location for the figure used to illustrate the pathway "[Apoptotic factor-mediated response](#)".

FrontPage

A list of the top-level pathways of the Reactome event hierarchy (autophagy, cell cycle, cell-cell communication, ...) listed on the [Reactome pathway browser page](#). The single instance of this class is maintained by the Editor-in-Chief and Managing Editor.

GO_BiologicalProcess

A local copy of the [Gene Ontology Biological Process ontology](#). Instances of this class provide *goBiologicalProcess* attributes for Reactome **Events** and **Regulation** instances.

GO_CellularComponent

A local copy of the [Gene Ontology Cellular Component ontology](#). Instances of this class provide the subcellular location or locations of Reactome **Events** and **PhysicalEntities** .

Compartment

A subset of `GO_CellularComponent` terms manually curated by Reactome to specify non-overlapping subcellular locations (*compartment* attribute values) of **Events** or **physicalEntities**. If a **physicalEntity** occurs in more than one compartment, different instances of the entity are created for each location, distinguished from one another by their *compartment* attributes.

GO_MolecularFunction

A local copy of the [Gene Ontology Molecular Function ontology](#). Instances of this class provide the *activity* of an instance of **catalystActivity**.

InstanceEdit

Records the date and time when a Reactome instance was *created* or *modified* (and, in the case of events, when they were authored, reviewed, or edited). It identifies the *person* responsible for the creation or modification and a timestamp. InstanceEdit instances are automatically generated when the central database is updated from the curator tool, and should not be manually edited.

PathwayDiagram

An SBML-compliant graphical representation of one or more **pathways** (*representedPathway* attribute values). Every **ReactionlikeEvent** in the included pathways is represented in the diagram, with icons that represent PhysicalEntities and edges that represent causal connections (provides input / output, catalyzes, regulates) among the physicalEntities. These are superimposed on icons that represent **compartments** (subcellular locations).

Person

The name of a person. Used to identify curators, authors, reviewers, and **literatureReference** authors. Any **Person** instance can have an *affiliation* attribute, but at present this information is recorded only for Reactome contributors, i.e., authors, reviewers, and curators. A contributor identifier such as ORCID can be associated with a person instance via the CrossReference attribute allowing to claim their contributions to Reactome in ORCID.

PhysicalEntity [superclass]

Used to maintain logical integrity of data model, not used for manual annotation. **PhysicalEntity** contains the classes **Cell**, **Complex**, **Drug**, **EntitySet**, **GenomeEncodedEntity**, **OtherEntity**, **Polymer**, and **SimpleEntity**. A PhysicalEntity is a physical substance that can interact with other substances. PhysicalEntities include all kinds of small molecules, proteins, nucleic acids,

chemical compounds, complexes, larger macromolecular assemblies, atoms (including ionized atoms), electrons, and photons.

Cell

A subclass of **PhysicalEntity**, represents a type of cell in a particular state of development/differentiation. A Cell instance has *cellType*, *tissueLayer*, *tissue*, and *organ* as single-valued attributes which are populated using terms from the Cell Ontology. Anatomy terms from the UBERON ExternalOntology are used to populate *tissueLayer*, *tissue*, and *organ* attributes. Cell instances also have multi-valued *ProteinMarker* and *RNAMarker* attributes, which are specific to the cell type and/or the cell state or are differentially upregulated in the Cell. The markers (protein or RNA) are manually curated and associated with literature references for evidence.

Complex

A **PhysicalEntity** that represents a larger functional unit formed by the covalent or noncovalent association of two or more **PhysicalEntity components** (which can themselves be **Complexes** or **Sets**). Usually, complex components are proteins, but they can also be RNA or DNA molecules, or simpleEntity chemicals, e.g. metal ions, nucleotides, lipids that function as enzyme cofactors. Complexes include homomultimers, e.g., [ESR1 homodimer \(nucleoplasm\)](#), heteromultimers, e.g., [FANCD2:FANCI \[nucleoplasm\]](#), and complexes composed of other complexes, e.g., [PPARG:RXRA:Corepressor Complex \[nucleoplasm\]](#).

Drug

A superclass which is not used to instantiate specific instances, the **Drug** superclass contains subclasses **ChemicalDrug**, **ProteinDrug**, and **RNADrug**. A drug may be defined as any substance that is administered as a medicine to alter metabolic processes in the body, most often processes involved in diseases or abnormal states.

Reactome does not classify as drugs those substances, such as insulin or epinephrine, that are produced naturally by the human body, though these substances may be administered as medicines and may be considered drugs in other databases.

Artificial derivatives of naturally produced human substances, such as glargine (a derivative of insulin), are considered drugs, as are natural non-human substances such as quinine and caffeine.

ChemicalDrug

A subclass of the **Drug** class. A ChemicalDrug is a synthetic substance or non-human natural substance that is not a protein, RNA, or DNA molecule and is administered to the body to modify a metabolic reaction or reactions, usually to treat a disease or abnormal state.

Mandatory attributes of ChemicalDrug include one or more *diseases* treated by a drug, and *referenceEntity*, which is a reference to a corresponding entity in a database such as Guide to Pharmacology.

ProteinDrug

A **ProteinDrug** is a substance that contains an active component that is a protein, whether modified or unmodified, and is administered to the body to modify a metabolic reaction or reactions, usually to treat a disease or abnormal state. A subclass of the **Drug** class. Mandatory attributes of ProteinDrug include *disease*, which is a reference to the disease treated by a drug, and *referenceEntity*, which is a reference to a corresponding entity in a database such as Guide to Pharmacology. Examples of **ProteinDrugs** are semaglutide, a chemically modified form of GLP1, and etanercept, a fusion between the TNFalpha receptor and the Fc portion of IgG1. Naturally occurring human proteins such as insulin are considered instances of the class EntityWithAccessionedSequence rather than instances of the class ProteinDrug. Artificially modified versions of human proteins, such as glargine or lispro insulin, are considered **ProteinDrugs**.

RNADrug

A subclass of the **Drug** class. Mandatory attributes of RNADrug include *disease*, which is a reference to the disease treated by a drug, and *referenceEntity*, which is a reference to a corresponding entity in a database such as Guide to Pharmacology. A RNADrug is a substance that contains an active component that is a ribonucleic acid, whether modified or unmodified, and is administered to the body to modify a metabolic reaction or reactions, usually to treat a disease or abnormal state. **EntityWithAccessionedSequence** is the non-drug class that corresponds to **RNADrug**.

Examples of RNADrugs are inclisiran and asvasiran, which are non-natural small interfering RNAs (siRNAs). Naturally occurring human RNAs are considered instances of the class EntityWithAccessionedSequence rather than instances of the class RNADrug.

EntitySet [superclass]

Used to maintain logical integrity of data model, not used for manual annotation. Two or more **PhysicalEntities** grouped because of a shared molecular feature or function. The superclass for **CandidateSet** and **DefinedSet**. While sets are, by default, homogeneous (members having the same subclass of PhysicalEntity), they are not required to be. For example, the defined set [platelet alpha granule contents](#) contains, as members, EntitiesWithAccessionedSequences, Complexes and Sets.

CandidateSet

A collection of physical entities that are functionally indistinguishable for the purpose of Reactome annotation, some of which are well-characterized (the “members” of the set) and

some of which are incompletely characterized (the “candidates” of the set), as judged by the curator who assembles the set and the outside expert reviewers who evaluate it.

Members of a set are related by an "OR" function. That is, either one member OR another member can participate in a reaction.

Sets may be ordered or unordered as specified in the *isOrdered* attribute. A specific member of an ordered set has a correspondence with a specific member of another ordered set, as specified by their positions within the sets. For example, consider an ordered set containing substrate1 and substrate2 that reacts to yield an ordered set containing product1 and product2. In this case, substrate1 will yield product1 and substrate2 will yield product2. In the case of unordered sets, any member in an input set can correspond to any member in an output set. Sets in Reactome are considered ordered by default.

Example:

RIP2 ubiquitin ligases

One (*member*) has been shown to participate as a ubiquitin ligase (ITCH); the other (*candidates*) are thought to do so. <https://reactome.org/content/detail/R-HSA-1248659>

DefinedSet

A collection of well-characterized physical entities that are functionally indistinguishable for the purpose of Reactome annotation, e.g., a collection of isoforms of a protein that all mediate the identical metabolic reaction. A set is formally a list of instances linked by logical ORs.

Sets may be ordered or unordered as specified in the *isOrdered* attribute. A specific member of an ordered set has a correspondence with a specific member of another ordered set, as specified by their positions within the sets. For example, consider an ordered set containing substrate1 and substrate2 that reacts to yield an ordered set containing product1 and product2. In this case, substrate1 will yield only product1 and substrate2 will yield only product2. In the case of unordered sets, any member in an input set can correspond to any member in an output set. Sets in Reactome are considered ordered by default.

GenomeEncodedEntity

Any informational macromolecule (DNA, RNA, protein) or entity derived from one by post-synthetic modifications, for example, covalently modified forms of proteins.

GenomeEncodedEntity is a subclass of **PhysicalEntity** and has one subclass: **EntityWithAccessionedSequence**. Unlike an **EntityWithAccessionedSequence**, a **GenomeEncodedEntity** is not required to have a reference entity, though a reference to an entity in a database can be specified in the *crossReference* attribute.

EntityWithAccessionedSequence

The subclass of **GenomeEncodedEntities** that can be associated through the *referenceEntity* attribute with reference molecules in UniProt (proteins) or ENSEMBL (DNA, RNA) databases. A

protein that has been partially purified and whose enzymatic properties are known but whose amino acid sequence is not, is a `GenomeEncodedEntity` but not an `EntityWithAccessionedSequence`.

OtherEntity

PhysicalEntities that we are unable or unwilling to describe in chemical detail and which, therefore, cannot be put in any other class. `OtherEntity` can be used to represent complex structures in the cell that take part in a reaction but which we can't or don't want to define molecularly.

Example 1: Cell membrane. In a case in which protein X associates with the membrane, but the actual membrane component(s) with which protein X interacts are unknown, the membrane can be represented as an "OtherEntity".

Example 2: kinesin-1, a microtubule motor protein, is involved in many kinds of movement in the cell, by 'walking' along microtubules, while dragging things like mitochondria, secretory vesicles, parts of the golgi, etc. Kinesin-1 binds to these complicated structures that we would not want to describe molecularly and which we can designate as "otherEntities".

Example 3: Holliday structure `\[nucleoplasm]`

<https://reactome.org/content/detail/R-ALL-75220#Homo%20sapiens>

Polymer

Molecules that consist of indeterminate numbers of repeated units, and complexes that contain repeated units whose stoichiometry is variable or unknown. The repeated unit(s) (identified in the `repeatedUnit` attribute) can be any **PhysicalEntity**. The presence of more than one `repeatedUnit` value implies that the relative numbers of units in the polymer are unknown. If the units are present in known proportions, a **complex** of the appropriate numbers of units is used as the `repeatedUnit`. The size range of a **polymer** can be specified with `minUnitCount` and `maxUnitCount` values. Examples:

-`\'glycogen\'` with `\'glucose\'` as `repeatedUnit`.

-`\'fibrin multimer\'` with `\'fibrin "monomer"\'` (itself a `Complex`) as `repeatedUnit`.

-A microtubule consisting of equal amounts of alpha and beta tubulin would be constructed as a **polymer** containing a `Complex` of alpha and beta tubulins in the `repeatedUnit` attribute.

-Completely hypothetical Example: A complex consisting of 1 "part" of A and "4 "parts" of B (i.e. 1:4 ratio) would be represented as a **polymer** with a **complex** of one A and 4 B as its `repeatedUnit`.

-Another hypothetical Example: a complex where the ratio of individual building blocks A and B is unknown or variable is represented as a **polymer** containing A and B directly in the `repeatedUnit` attribute.

SimpleEntity

A defined chemical species not encoded directly or indirectly in the genome, typically a small molecule such as ATP or ethanol. The detailed structure of a **simpleEntity** is specified by linking it to the information provided for the molecule in the ChEBI external database via the *referenceEntity* slot. Separate **simpleEntity** instances are needed for each subcellular location (*compartment*) in which a molecule is found, e.g., ATP [cytosol] and ATP [nucleoplasm]. SimpleEntities such as ATP that occur in many species are not assigned a species.

Publication [superclass]

Used to maintain logical integrity of data model, not used for manual annotation. The Publication superclass encompasses any document, paper or electronic, that makes information available to the public. The subclasses of Publication are Book (chapters of books or entire books), LiteratureReference (journal articles), and URL (uniform resource locators that allow navigation of the internet to an electronic document).

Book

A book may be used in the literatureReference attribute of the relevant data class instances (ie pathway, reaction, entity). *Author*, *title*, and *year* of publication are required attributes. Chapters of books can also be cited within a book by using the *chapterTitle* attribute of the Book class.

LiteratureReference

A journal article, typically abstracted in PubMed, cited in a **summation** or linked to an **entity** or **event** instance. The attribute *pubMedIdentifier* contains the PubMed identifier (PMID) obtained from PubMed (<https://pubmed.ncbi.nlm.nih.gov/>) and is mandatory. Further information about the literature reference is also obtained from PubMed and is found in the attributes *author*, *year*, *title*, *journal*, *volume*, and *pages*.

URL

A website cited as a literature reference in data class instances such as **summation**, **entity**, or an **event** instance. The URL for the website is listed in the uniformResourceLocator attribute. The URL should contain all information, including transfer protocols such as "http://", "https://", or "ftp://", required to navigate to the resource.

psiMod

A local copy of the [PSI-MOD ontology](#), used to create descriptions of *modifiedResidues* of proteins.

ReferenceDatabase

This class describes the source (external database) of an identifier in the **Databaselfentifier** (and SequenceDatabaselfentifier) instance, e.g., *Databaselfentifier* [COMPOUND:C00113](#) links the Reactome instance to the record maintained by the [KEGG "compound" database](#).

ReferenceEntity [superclass]

Used to maintain logical integrity of data model, not used for manual annotation.

ReferenceEntity captures the invariant features of a molecule such as its names, molecular structure and links to external community reference databases like UniProt, ENSEMBL or ChEBI, and is created by importing content from those databases. A ReferenceEntity is an attribute of the PhysicalEntity subclasses Drug, EWAS and SimpleEntity, and enables grouping of Reactome PhysicalEntities like 'Glc ([alpha-D-glucose](#)), [extracellular region](#)' and 'Glc ([alpha-D-glucose](#)), [cytosol](#)' (all linked to the record of alpha-D-glucose maintained by [ChEBI database](#), indicating their identical chemical nature. ReferenceEntities are not used in Reactions directly; they are attached to the PhysicalEntities involved. ReferenceEntities derived from Uniprot and ChEBI usually don't need to be created by curators; they are imported automatically. In the case of as yet unexisting ChEBI entries, contact the managing Editor.

ReferenceGroup

ReferenceGroup is a chemical group instance (children of [CHEBI:24433](#)) from the ChEBI database, used to populate the *modification* attribute of classes such as **GroupModifiedResidue**. The ReferenceGroup record consists of an *identifier* and a *referenceDatabase*.

ReferenceMolecule

ReferenceMolecule is used to cross-reference individual chemical molecules to the ChEBI database, and is used to populate the *referenceEntity* attribute of **SimpleEntity**. Consists of an *identifier* and a *referenceDatabase*.

ReferenceSequence [superclass]

Used to maintain logical integrity of data model, not used for manual annotation. Molecules with an accessionedSequence. This class is a subclass of **ReferenceEntity**, and is the superclass of **ReferenceDNASequence**, **ReferenceGeneProduct**, and **ReferenceRNASequence**, so instances of it should not be created manually.

ReferenceDNASequence

Used to populate the *referenceEntity* attribute for EWASs describing DNA molecules (for instance EWAS's of human genes). The **ReferenceDNASequence** record consists of a *referenceDatabase* (generally ENSEMBL), the *identifier*, and a *species*.

ReferenceGeneProduct

Used to populate the *referenceEntity* attribute for EWASs describing proteins. The **ReferenceGeneProduct** record consists of a *referenceDatabase* (generally Uniprot), the *identifier* and a *species*. The canonical form of the protein from UniProt is used as the **ReferenceGeneProduct** unless the literature supports the role of a particular isoform of a protein in the described function. **ReferenceGeneProduct** is also used to populate the *referenceSequence* attribute of **ModifiedResidue**.

ReferenceIsoform

Used to populate the *referenceEntity* attribute for EWASs describing proteins, where the experimental literature supports the role of a specified isoform of a protein in a particular function. The **ReferenceIsoform** record consists of a *referenceDatabase* (generally Uniprot), the *identifier* and a *species*.

ReferenceRNASequence

Used to populate the *referenceEntity* attribute for EWASs describing RNA molecules (for instance EWAS's of mRNAs). The **ReferenceRNASequence** record consists of a *referenceDatabase* (generally ENSEMBL), the *identifier*, and a *species*.

ReferenceTherapeutic

Used to populate the *referenceEntity* attribute for **PhysicalEntities** of the **Drug** class, including **ChemicalDrug**, **ProteinDrug**, and **RNADrug**. The **ReferenceTherapeutic** record consists of a *referenceDatabase* (generally Guide to Pharmacology) and the *identifier*.

Regulation [superclass]

Used to maintain logical integrity of data model, not used for manual annotation. Consists of subclasses **NegativeRegulation** and **PositiveRegulation**, along with their subclasses, **NegativeGeneExpressionRegulation** and **PositiveGeneExpressionRegulation** and **Requirement**, respectively. Instances of Regulation subclasses are added to the *regulatedBy* attribute of **ReactionLikeEvents** to capture the negative or positive effect of a **PhysicalEntity** on the described event.

NegativeRegulation

NegativeRegulation is an optional class used to populate the *regulatedBy* attribute of a **ReactionLikeEvent** that describes an inhibitory effect of a Regulator, a **PhysicalEntity**, on that event. The mechanism of action of a Regulator, if known, can be described through the optional *activity* attribute which is populated by terms from an ExternalOntology GO_MolecularFunction. If the Regulator is a Complex, the specific Complex component(s) that play the regulatory role can be specified as *activeUnit(s)* of the NegativeRegulation instance.

NegativeGeneExpressionRegulation

NegativeGeneExpressionRegulation is a subclass of **NegativeRegulation** that describes a direct inhibitory effect of a Regulator on a **BlackBoxEvent** that represents gene expression. The Regulator is a complex of regulatory component(s) such as transcription factors with the target gene or mRNA molecule. The regulatory component(s) are defined as *activeUnit(s)* of the **NegativeGeneExpressionRegulation** instance.

PositiveRegulation

PositiveRegulation is an optional class used to populate the *regulatedBy* attribute of a **ReactionLikeEvent** that describes a stimulatory effect of a Regulator, a **PhysicalEntity**, on that event. The mechanism of action of a Regulator, if known, can be described through the optional *activity* attribute which is populated by terms from an External Ontology GO_MolecularFunction. If the Regulator is a Complex, the specific Complex component(s) that play the regulatory role can be specified as *activeUnit(s)* of the PositiveRegulation instance.

PositiveGeneExpressionRegulation

PositiveGeneExpressionRegulation is a subclass of PositiveRegulation that describes a direct stimulatory effect of a Regulator on a BlackBoxEvent that represents gene expression, so that the Regulator represents a complex of regulatory component(s) such as transcription factors with the target gene or mRNA molecule. The regulatory component(s) are defined as *activeUnit(s)* of the PositiveGeneExpressionRegulation instance.

sequence-related comments that are provided by the reference database. It is filled in automatically. Curators should not alter this.

Requirement

A subclass of **PositiveRegulation** that denotes a regulatory **PhysicalEntity** without which the regulated event cannot occur.

ReviewStatus

An attribute of event. Describes the level of internal/external review that the event has undergone.

One star (1): Restructured after internal review NOT RELEASED.

Two star (2) Restructured after external review NOT RELEASED.

Three star (3) Internally reviewed, awaiting external review RELEASED.

Four star (4) Restructured after external review, then internally re-reviewed RELEASED.

Five star. 5 Externally reviewed RELEASED. A more detailed description can be found in the user guide [here \(https://reactome.org/userguide/review-status\)](https://reactome.org/userguide/review-status).

StableIdentifier

A unique identifier associated with a database object that does not change over time. Stable Identifiers have the format R-XXX-YYYYYY.Z where R stands for Reactome and invariant. XXX is the three letter species code (see appendix XX for species codes). If the instance applies to all species (e.g. a simple entity such as ATP), the code is "ALL". If the instance applies to none or multiple but not all species, the code is "NUL". YYYYYY is the database identifier of the instance. If the instance is modified between database releases, the version number (Z) in the IdentifierVersion attribute is incremented.

StableIdentifierHistory

This class is used for tracking, and each instance represents a stable identifier with a particular minor version that was present at one time in the history (either still existing currently on at one time in the past and no longer). The class is used with "StableIdentifierReleaseStatus" to indicate which release numbers had this minor version and if/when it was changed (i.e. replaced, removed, incremented, etc.)

StableIdentifierReleaseStatus

Instances of this class are used as values for StableIdentifierHistory's status attribute to indicate the status (i.e. "EXISTS", "INCREMENTED", "REPLACED", "REMOVED", etc.) for a given release version.

Summation

A free text description of an **Event** or **physicalEntity**. Citations that provide useful background information, but that are not sources of primary data for the event or entity, can be linked to summations. (**LiteratureReferences** providing experimental data for the details of an event should be linked directly to the event). Attributes: *Text*- description of the event

LiteratureReference- paper(s) that support the text description, but not the primary evidence for the event. See Summation example for [FAM83B](#), ([FAM83A](#), [FAM83D](#)) bind EGFR.

Taxon

A controlled vocabulary populated manually by Reactome from the [NCBI Taxonomy Database](#), to provide terms to identify the *species* of events and genome-encoded physical entities.

Taxonomy terms of higher rank than species, and their relationships, are maintained in Reactome to facilitate computational reasoning and data mining. Each taxon instance includes the taxon's *name* as assigned by NCBI, its NCBI identifier (in the form of a Reactome *crossReference*), and its NCBI-assigned *supertaxon*. This class is maintained by the Editor-in-Chief and Managing Editor. Taxon terms of species rank are used for annotation of gene products and other relevant PhysicalEntities, and events.

Species

Gene products and events have **species**-level NCBI-derived **taxon** terms as attributes. This class is maintained by the Editor-in-Chief and Managing Editor.

_Release

A class which indicates a Reactome version number and its date of release. Also, an attribute of the class `_updateTracker` that indicates the release number and date of a specified Reactome database release.

_UpdateTracker

Describes the changes that were made to an instance at a specified Reactome release relative to the previous release. Specifies the DBID and name of the instance that has changed (an event or PhysicalEntity), the release number and date of that release that the change occurred, and a list of changes made to the entity (see appendix R9 for a list of changes that are tracked)

Reactome Attributes

_displayName

This attribute contains the name of the instance that will be displayed on the Reactome website. It is automatically filled in by the curator tool each time a new instance is created, using attribute

values supplied for the instance by the curator. For example, the `_displayName` of an event consists of the event's manually generated *name* and its manually selected *species*.

accession

Holds the numerical portions of identifiers assigned to GO terms in the local copies of the GO ontologies. Should not be manually edited.

accessUrl

Template that is used to form the URL that will be used to link to a particular record in an external database. Generally, this should not need to be touched. Contact developers if you think any changes are necessary.

activity

contains the GO Molecular Function ontology. An Activity term is used to cross-reference a Reactome CatalystActivity. Curators should never create instances of class Activity directly. If curators want to use a GO term that is not in Reactome (yet) contact Managing Editor.

address

This is an attribute for the affiliation class. Enter the address of the research institute (etc).

affiliation

An Affiliation is the name and address of an institution, e.g., Cold Spring Harbor Laboratory 1 Bungtown Road, Cold Spring Harbor NY USA.

author

This attribute must be filled in for instances of LiteratureReference, (where it refers to name of the author of a particular reference), for InstanceEdit (where it refers to the curator that has generated the instanceEdit), and for events.

comment

This attribute is used for SequenceDatabaseIdentifier class instances and contains any cation in the cell. Is an attribute of PhysicalEntities and Events. Note that complexes with components that span multiple cellular locations should specify one primary location as the compartment and the location of the other components will be automatically filled into the "includedLocation" attribute during the database release process. For example, a transmembrane complex with

components in the extracellular space, plasma membrane, and cytosol would have plasma membrane specified as the compartment.

componentOf

An attribute of GO_CellularComponent that is specified by the Gene Ontology and describes the GO Cellular Component instances that are parent terms. Imported from GO.

coordinate

Refers to the amino acid residue location at which a modification occurs within a protein.

created

Automatically filled in with the name of the curator/developer creating the instance and date/time stamp that the instance was created.

crossReference

This attribute holds references to the equivalent instance in other databases. For example, the small molecule 4MAA has the KEGG COMPOUND identifier COMPOUND:C01036 in its *crossReference* attribute.

databaseIdentifier

Holds database identifiers from external databases like UniProt and ChEBI (but not GO – see *Accession* attribute). Has attributes for the *referenceDatabase* and the *identifier* in that database.

dateTime (on InstanceEdit)

A timestamp generated automatically when the central database is updated from a curator tool project. Do not edit it manually.

DB_ID

These are the unique stable identifiers applied to each instance in the Reactome database. They are generated automatically when newly created instances are first submitted to the database. They should never be modified or manually created by a curator.

definition

This attribute is used in Event, PhysicalEntity, Activity, Evidencetype and GO_BiologicalProcess instances. The definition of a Reactome Activity is the official Gene Ontology definition for the equivalent GO Molecular Function term. The GO_BiologicalProcess definition is the official Gene Ontology definition for the corresponding Biological process term. The Evidencetype definition is taken from the GO evidence code definitions.

description

This is found in the class 'SequenceDBI'. This is for the 'Description DE' line from SwissProt.

editor

Choose or create a Person instance/s for the editor/s. Used to determine the information for the frontpage.

containedIn

endCoordinate

The amino acid residue location at which an EntityWithAccessionedSequencepart ends.

figure

Certain events in Reactome are associated with a descriptive illustration. See, for example, [Autophagy](#). When such an illustration exists, the figure attribute of the event is populated with a figure instance containing the location of the image file on the server in its URL attribute. The URL has the format: figures/.../[FigureFileName].svg

firstname

The first name of a person. Optional attribute – can be left blank. Reactome uses only the first letter of the first name (entered separately in the *initial* attribute) to create names of people for display on the web site.

formula

The chemical formula of a **SimpleEntity**. For example, ATP = C₁₀H₁₆N₅O₁₃P₃.

functionalStatusType

Describes the functional consequence of this genetic change (gain-of-function, loss-of-function, decreased_transcript_level etc). With sequenceVariant, functionalStatusType makes up **FunctionalStatus**, which is used to fill the *functionalStatus* attribute of EntityFunctionalStatus

geneIdentifier

This attribute is for storing information about what is the gene that this transcript/protein originates from. This is really just a shortcut. If the current instance is a gene identifier then this should be left empty. Accepts multiple values in order to be able to point to the SAME gene in different databases, e.g. EMBL, HUGO, Ensembl.

geneName

GN lines from sequence record.

hasComponent

Holds instances of **PhysicalEntity** that are components of a **Complex**.

hasEvent

Holds instances of **Event** that are grouped into a **Pathway**. They should be entered in the order as they appear in the **Pathway**. On **BlackBoxEvent** this attribute holds **Event** instances to describe known steps within the **BlackBoxEvent**, e.g. the reactions of the fatty acid beta oxidation cycle.

hasMember

Takes individual reactions for an instance of **ReactionlikeEvent** that represents a group of reactions via the use of an **EntitySet**. Such individual reactions do not need to be spelled out, unless they have a distinct feature like a literature reference or specific regulation.

identifier

This attribute holds the actual database identifier number for a given **DatabasIdentifier** instance.

inferredFrom

Points to the event or entity in another species that this event/entity has been inferred from. If the inference is based on computation only, this is indicated under evidenceType (= IEA).

initial

This attribute holds the initials of a person represented in a **Person** instance.

input

The input **PhysicalEntities** of a given **ReactionLikeEvent** are each entered individually.

instanceOf

Only used for GO classes and is populated/updated automatically.

InteractionEvent

This class is designed to model simple interactions between any physical entities. It is introduced to be used for data imported from other data sources including protein-protein interactions or other types of interactions.

journal (on LiteratureReference)

The name of the scientific journal in which the reference was published (or the title of a book). For references retrieved from PubMed with the curator and author tools, this slot is filled automatically.

keyword

Contains keywords associated with the sequence. This is pulled in (if available) automatically from the external database.

literatureReference (on summation)

Paper/s that support the text description, but not the primary evidence for the event. This shows up as hyperlinked lines on a web browser.

modification

For a given `modifiedResidue` instance, enter the `simpleEntity` that represents the specific modification of the residue within the modified protein. For example, for the "ModifiedResidue" instance: "Orthophosphate on Serine [nucleus] 428 of SPTREMBL:O46469", the modification would be "Orthophosphate".

modified

Filled in automatically when a database instance is modified.

name

A short textual description of the event/entity/etc. For some classes it is a defining attribute, so it should be chosen carefully to be unique.

Ontology

Internal definition and not instantiable for holding data. It is the content of the protegé file defining the database.

orthologousEvent

Points to equivalent events in other species. In contrast to 'inferredFrom' this attribute is attached to the events in both species - it only indicates equivalence, not inference.

output

The output **PhysicalEntities** of a given event are each entered individually.

pages

The inclusive page numbers of a **LiteratureReference**. For references retrieved from PubMed with the curator and author tools, this attribute is filled automatically.

precedingEvent

Points to the preceding event(s), which is usually events whose output is used as input for the present event. The preceding event can also point to a pathway, but make sure the connection is always given on the reaction level as well (this is important when it comes to visualization).

proteinIdentifier

This attribute is for storing information about what the protein this transcript/gene produces. This is really just a shortcut. If the current instance is a protein identifier then this should be left empty. Accepts multiple values in order to be able to point to the same proteins in different databases, e.g. EMBL, HUGO, Ensembl, RefSeq.

pubMedIdentifier (on literatureReference)

The pubmed identifier (number only). This information must be manually supplied by the curator.

reactionType

A *controlledVocabulary* that provides attributes to classify *reactionLikeEvents*. Currently in use only to distinguish reactions involving drugs by modes of action.

drugActionType

Reactions that have *drugs* and *sets or complexes containing drugs* as inputs and output have a drugAction attribute taken from the IUPHAR / Guide to Pharmacology [glossary](#) to classify the effect of the drug on the reaction (agonist, channel blocker, etc.).*

requiredInputComponent

Takes an input component that is essential for this event to happen, e.g. a defined domain of a protein. Use Domain, with the appropriate coordinates filled in.

Note: To be used only for '\non-trivial statements\''. For example, repeating an entity here that is identical to a PhysicalEntity given for input is not helpful.

relatedSpecies (on event)

This attribute is used in events involving more than one species (e.g event involving host-pathogen or symbiotic interactions) to denote the "bystander" species in the event. For example, in the event such as flagellin of Escherichia coli binds to human TLR5, the process is occurring on the human cell membrane and involves protein from the bystander bacterial species. Thus, human would be entered in the *species* attribute and E.coli would be entered as *relatedSpecies*. For additional information and use cases, see this [document](#).

residue

This is to hold the residue (like serine or tyrosine) which is modified. The value is an instance of class ReferenceGroup or ReferenceMolecule, i.e. an instance representing serine.

resourceIdentifier

This attribute is used to add identifiers registered in the identifiers.org web site for ReferenceDatabase instances. It is added for QA purposes and should not affect curation work. The attribute is string, single-valued, and optional.

reverseReaction

Choose a Reaction that is the reverse of the current Reaction.

reviewer (on Event)

The name of the person who reviewed the event. These attribute values are the source of the reviewers given for high-level events listed on the Reactome table of contents.

secondaryIdentifier

Generated automatically.

sequenceOntology

SequenceOntology encompasses terms representing the variant features and attributes of biological sequence derived from an ExternalOntology Sequence Ontology (<http://www.sequenceontology.org/>). SequenceOntology terms are used to populate the *structuralVariant* attribute of **FunctionalStatus**.

species

This attribute holds the name of the species in which the described physicalEntity or Event is occurring. The *species* attribute on event identifies the environment where the reaction is happening. This should almost always be human for human Reactome curation. The exceptions are when we have annotated pathways in other species for inference (e.g Mus) or when we are specifically annotating another species pathways as part of another project (e.g Gallus and Mycobacterium tuberculosis). In disease or immune system processes, the host would be listed as the “species”.

When annotating a process involving host-pathogen interactions, the *relatedSpecies* attribute should be used to list the pathogen (bystander) species. See the *relatedSpecies* attribute definition and use cases for a more detailed description. **Note:** When creating a chimeric reaction to be used for inference (i.e one that involves entities from multiple species and (usually) observed in vitro) ALL species should be listed in the species attribute.

startCoordinate

The amino acid residue location at which an EntityWithAccessionedSequence starts.

structuralVariant

Field takes terms from [SequenceOntology](#) and describes the nature of the underlying genetic change that gives rise to the variant. With FunctionalStatusType, structuralVariant makes up FunctionalStatus, which is used to fill the *functionalStatus* attribute of EntityFunctionalStatus.

summation

A succinct free-text description of the instance, typically an event, with which it is associated.

superTaxon

This refers to the "parent" for a given taxon within the NCBI taxonomy hierarchy. For example the SuperTaxon for both Rattus and Mus is Murina and the SuperTaxon for Murina is Muridae.

surname

The family name of a **person**.

text

Holds the text of the summation.

templateEvent

Allows making a reference to a general process that underlies a **BlackBoxEvent**, e.g. 'Gene Expression' as template for the synthesis of a specific protein. This attribute has never been used - should it be removed from the data model?

title

The title of a **LiteratureReference**. For references retrieved from PubMed with the curator and author tools, this attribute is filled automatically.

transcriptIdentifier

This attribute is for storing information about what is the transcript that this protein originates from or that this gene produces. This is really just a shortcut. If the current instance is a transcript identifier then this should be left empty. Accepts multiple values in order to be able to point to the same gene and proteins in different databases, e.g. EMBL, Ensembl, RefSeq.

url

Holds URL for site which gives "summary information" about the database including a description of what information the db contains. This attribute should not need to be modified manually.

volume

The volume number of the *journal* in a **LiteratureReference**. For references retrieved from PubMed with the curator and author tools, this attribute is filled automatically.

year

The year a **LiteratureReference** was published. For references retrieved from PubMed with the curator and author tools, this slot is filled automatically.

Graph database specific classes

DBInfo

Generic information about the current version of the graph database used.

GO_Term

Gene Ontology term that is used by Reactome events as cross references. Can be GO_BiologicalProcess, GO_CellularComponent or GO_MolecularFunction.

Interaction

Parent label for all kind of molecular interactions brought into Reactome automatically from other databases, based on referenced evidence in a reviewed publication.

TopLevelPathway

A specific kind of Pathway which covers a broad topic of cellular biology. The top-level pathways represent the highest level of the hierarchical organisation of events in Reactome, as displayed in [Pathway Browser](#).

UndirectedInteraction

A kind of molecular interaction for which no information is available about the directionality of the interaction that forms a bond between two molecules.