BEL 2.0 Specification

Natalie Catlett

Version 2.0, 2014-06

Table of Contents

This is the home of the BEL Language Documentation v2.0.

Please cite the BEL v2.0 language documentation as "BEL v2.0 Language Documentation, https://github.com/OpenBEL/language ", along with the date accessed.

Overview

The Biological Expression Language (BEL) is a language for representing biological observations in a computable form, along with contextual information. BEL is intended as a knowledge capture and interchange medium. BEL is used to qualitatively represent causal and correlative relationships involving biological measurements (e.g., RNA, protein, phosphorylated proteins). Each BEL statement stands alone as an individual observation or fact, and can be integrated with related observations into a cohesive network.

BEL is a human-readable and -writable language designed to be easy for life scientists to learn and use. BEL is comprised of a relatively small set of function and relationships types that can be used in conjunction with widely used vocabularies like HGNC human gene symbols, Gene Ontology, ChEBI, and MeSH. As a language of discourse for biological findings, BEL is designed to be "white-boardable" as well as written.

BEL History

BEL was initially designed in 2003 at Selventa (operating as Genstruct®) by Dexter Pratt. BEL was designed with a focus on capturing qualitative causal relationships that could be used for inference. From 2003 to 2010, BEL evolved in response to daily use by scientists representing findings derived from tens of thousands of abstracts and full-text articles.

In 2011, it was proposed to make BEL an open standard. BEL has been refined, formalized, and extended to meet the needs of a broader community to represent, manage, and share scientific findings in the life sciences. BEL and associated software was released as open-source technology in 2012. OpenBEL became a Linux Foundation Collaborative project in 2013.

BEL Version History

- BEL v1.0 – initial open source release, 2012
- BEL v2.0 – major revisions and refinements, 2014

Summary of Changes for BEL v2.0

These additions and modifications enhance the BEL language by providing new representation capability (e.g., DNA and RNA variants, protein cleavage fragments, cellular location of abundances) and enabling the use of external vocabularies (post-translational modifications, activities).

Variants

- Now represent sequence variants at DNA, RNA, and protein levels.
- Now represent multiple substitutions within the same gene/RNA/protein
- New BEL abundance modifier function `variant("")` / `var("")` is used for most variant types, replacing `substitution()` / `sub()` and `truncation()` / `trunc()`. Human Genome Variation Society (HGVS) nomenclature adopted to describe variants (Dunnen and Antonarakis, 2000) within the `var("")` modifier function, expanding supported types of variation to include insertions, deletions, duplications as well as non-specific variants.
- Usage of `fus()` changed. Instead of a modifier function for a gene/RNA/protein abundance, `fus()` is used to compose new entities that can be used in place of a namespace value for abundance functions.

Protein Cleavage Fragments

- New abundance modifier function `fragment()` / `frag()` to be used within protein abundances to specify protein fragments based on amino acid sequence range.

Post-Translational Protein Modifications

- The `proteinModification()` / `pmod()` abundance modifier function can now use external vocabularies (e.g., PSI-MOD) for modification types, enabling users to add types without requiring a language change.
- Now multiple `pmod()` expressions can be used within a protein abundance.

Translocations and Cellular Location

- New abundance modifier function to specify location - `location()` / `loc()`.
- Change in `translocation()` / `tloc()` function format, to explicitly add BEL location functions to location arguments.

Activity Functions

- The ten distinct BEL activity functions, e.g., `kinaseActivity()` / `kin()`, `catalyticActivity()` / `cat()`, `transcriptionalActivity()` / `tscript()`, are consolidated to a single activity function `activity()` / `act()`.
- New modifier function `molecularActivity()` / `ma()` can be used to specify specific activity types, using external vocabularies, e.g., GO Molecular Function, or a default BEL vocabulary.

Regulates Relationship

- New causal relationship `regulates` to represent cases where A is reported to affect B, but it cannot be determined if A increases or decreases B.

BEL Script Format Changes

- Citation annotation requirement removed for *Name* field
- Citation annotation *DOI* and *URL* added as accepted types
- *BEL Script Evidence* Annotation renamed to *Support*
- BEL version set in document header

1. Language Structure

Knowledge in BEL is expressed as BEL Statements. Generally, BEL Statements have the form of a *subject - predicate - object* triple, where the subject is a BEL Term, the predicate is one of the BEL relationship types (e.g., `increases`), and the object can be either a BEL Term or a BEL Statement. A BEL Statement may also be comprised of a subject term only.

BEL Terms are composed of BEL Functions applied to concepts referenced using Namespace identifiers. Each BEL Term represents either an abundance of a biological entity, e.g., human AKT1 protein, or a process such as apoptosis.

BEL Annotations are applied to BEL Statements to optionally express additional information about the statement itself such as the citation for the publication reporting the observation, or the context in which the observation was made (e.g., species, tissue, cell line).

1.1. Namespaces

BEL is specifically designed to adopt external vocabularies and ontologies, and represent life-science knowledge in the language and schema of the organization collecting or using the knowledge. Thus, BEL Terms are defined by reference to concepts in external vocabularies, which provide a set of well-known domain values, such as the official human gene symbols provided by HGNC (http://www.genenames.org/) . While we consider it good practice to define biological entities with respect to well-defined domains such as public ontologies, no specific vocabulary is essential to the use of BEL, and users are free to define and reference their own vocabularies as needed.

BEL uses Namespaces to unambiguously reference concepts. The user associates a Namespace prefix with an external vocabulary and uses the prefix to refer to elements of the vocabulary. For example, if we associate the Namespace prefix HGNC with the vocabulary of symbols managed by the HGNC committee, we can then compose BEL Terms by referencing the HGNC Namespace prefix and any concept from the HGNC namespace together with a relevant BEL Function, e.g., `proteinAbundance(HGNC:AKT1)` or `rnaAbundance(HGNC:TNF)`.

1.1.1. Equivalencing between Namespaces

Values from different Namespaces may correspond to the same biological concept. For example, the name AKT1 in the HGNC Namespace refers to the same gene referenced with ID 207 in the EGID (Entrez Gene Identifier) Namespace. The BEL Framework assembles knowledge into a cohesive network, mapping equivalent BEL Terms, e.g., `proteinAbundance(HGNC:AKT1)` and `proteinAbundance(EGID:207)`, to a single node in the network. This correspondence of Namespace

values is handled in the BEL Framework separately from BEL knowledge representation.

1.2. Terms

Two general categories of biological entities are represented as BEL Terms: **abundances** and **processes**.

1.2.1. Abundances

Life science experiments often measure the abundance of a type of thing in a given sample or set of samples. BEL Abundance Terms represent classes of abundance, the abundances of specific types of things. Examples include the *protein abundance of TP53*, the *RNA abundance of CCND1*, the *abundance of the protein AKT1 phosphorylated at serine 21*, or the *abundance of the complex of the proteins CCND1 and CDK4*.

1.2.2. Processes

BEL Process Terms represent classes of complex phenomena taking place at the level of the cell or the organism, such as the biological process of *cell cycle* or a disease process such as *Cardiomyopathy*. In other cases, BEL Terms may represent classes of specific molecular activities, such as the kinase activity of the AKT1 protein, or a specific chemical reaction like conversion of superoxides to hydrogen peroxide and oxygen.

Measurable biological parameters such as *Blood Pressure* or *Body Temperature* are represented as process BEL Terms. These BEL Terms denote biological activities that, when measured, are reduced to an output parameter.

1.2.3. BEL Terms as Functional Expressions

BEL Terms are denoted by expressions composed of a BEL Function and a list of arguments. BEL v2.0 specifies a set of approximately 20 functions allowed in term expressions.

The combination of a BEL function and its arguments fully specifies a BEL Term. The BEL Term expression `f(a)` denotes a BEL Term defined by function `f()` applied to an argument `a`. Wherever the same function is applied to the same arguments, the resulting BEL Term references the same biological entity.

The semantics of a BEL Term are determined by the function used in the term expression. For example, the function `proteinAbundance()` is defined such that any term expression using `proteinAbundance()` represents a class of abundance of protein. Many BEL functions take only single values as arguments, providing a structured method of using ontologies and vocabularies in BEL. For example, values in the HUGO Gene Nomenclature Committee (HGNC) vocabulary of official human gene symbols can be used to designate gene, RNA, and protein abundances. The function `proteinAbundance()` could then be applied to an HGNC gene symbol, *AKT1* for example, to indicate the class of protein abundances produced by the AKT1 gene, producing the BEL Term `proteinAbundance(HGNC:AKT1)`.

1.3. Statements

A BEL Statement represents an experimental observation, generally reported in a scientific publication or unpublished experimental data. Generally, BEL Statements express a causal or correlative relationship between two biological entities. Because BEL Terms are functionally composed, a BEL Statement can consist of a single BEL Term; this simple statement indicates that the biological entity represented by the term has been observed.

1.3.1. Example BEL Statements

Subject Term Only

```
complex(p(HGNC:CCND1), p(HGNC:CDK4))
```

The abundance of a complex formed from protein abundances designated by *CCND1* and *CDK4* in the HGNC namespace. This is a subject term only statement, and indicates that the entity specified by the term has been observed.

Causal

```
p(HGNC:CCND1) => act(p(HGNC:CDK4))
```

The abundance of the protein designated by *CCND1* in the HGNC namespace directly increases the activity of the abundance of the protein designated by *CDK4* in the HGNC namespace.

Causal

```
p(HGNC:BCL2)-| bp(MESHPP:Apoptosis)
```

The abundance of the protein designated by *BCL2* in the HGNC namespace decreases the biological process designated by *apoptosis* in the MESHPP (phenomena and processes) namespace.

Nested Statement - Object Term is Statement

```
p(HGNC:GATA1) => ( act(p(HGNC:ZBTB16)) => r(HGNC:MPL) )
```

The abundance of the protein designated by *GATA1* in the HGNC namespace directly increases the process in which the activity of the protein abundance designated by *ZBTB16* in the HGNC namespace directly increases the abundance of RNA designated by *MPL* in the HGNC namespace.

1.4. Annotations

Each BEL Statement can optionally be annotated to express knowledge about the statement itself. Some important uses of annotations are to specify information about the:

- biological system in which the observation represented by the statement was made
- experimental methods used to demonstrate the observation
- knowledge source on which the statement is based, such as the citation and specific text supporting the statement

Examples of annotations that could be associated with a BEL Statement are the:

- PubMed ID specifying the publication in which the observation was reported,
- Species, tissue, and cellular location in which the observations were made, and
- Dosage, exposure and recovery time associated with the observation.

2. BEL Functions

This section provides a listing and explanation of all BEL functions that are included in the BEL v2.0 Language Specification.

2.1. Abundance Functions

The following BEL Functions represent classes of abundances of specific types of biological entities like RNAs, proteins, post-translationally modified proteins, and small molecules. Biological experiments frequently involve the manipulation and measurement of entities in samples. These BEL functions specify the type of entity referred to by a namespace value. For example,`geneAbundance(HGNC:AKT1)`, `rnaAbundance(HGNC:AKT1)`, and `proteinAbundance(HGNC:AKT1)`, represent the abundances of the AKT1 gene, RNA, and protein, respectively.

2.1.1. abundance(), a()

`abundance(ns:v)` or `a(ns:v)` denotes the abundance of the entity designated by the value `v` in the namespace `ns`. abundance is a general abundance term that can be used for chemicals or other molecules not defined by a more specific abundance function. Gene, RNA, protein, and microRNA abundances should be represented using the appropriate specific abundance function.

Examples - small molecule and chemical

```
a(CHEBI:"oxygen atom")
a(CHEBI:thapsigargin)
```

2.1.2. complexAbundance(), complex()

The `complexAbundance()` or `complex()` function can be used with either a namespace value or with a list of abundance terms.

`complexAbundance(ns:v)` or `complex(ns:v)` denotes the abundance of the molecular complex designated by the value `v` in the namespace `ns`. This form is generally used to identify abundances of named complexes.

Example - named complex

```
complex(SCOMP:"AP-1 Complex")
```

`complexAbundance(<abundance term list>)` denotes the abundance of the molecular complex of members of the abundances denoted by `<abundance term list>`, a list of abundance terms supplied as arguments. The list is unordered, thus different orderings of the arguments should be interpreted as the same term. Members of a molecular complex retain their individual identities. The `complexAbundance()` function does not specify the duration or stability of the interaction of the members of the complex.

Example - composed complex

```
complex(p(HGNC:FOS), p(HGNC:JUN))
```

2.1.3. compositeAbundance(), composite()

The `compositeAbundance(<abundance term list>)` function takes a list of abundance terms. The `compositeAbundance()` or `composite()` function is used to represent cases where multiple abundances synergize to produce an effect. The list is unordered, thus different orderings of the arguments should be interpreted as the same term. This function should not be used if any of the abundances alone are reported to cause the effect. `compositeAbundance()` terms should be used only as subjects of statements, not as objects.

Example - BEL Statement with compositeAbundance term

```
composite(p(HGNC:IL6), complex(GOCC:"interleukin-23 complex")) increases bp(GOBP:"T-
helper 17 cell differentiation")
```

In the above example, IL-6 and IL-23 synergistically induce Th17 differentiation.

2.1.4. geneAbundance(), g()

`geneAbundance(ns:v)` or `g(ns:v)` denotes the abundance of the gene designated by the value v in the namespace ns. `geneAbundance()` terms are used to represent the DNA encoding the specified gene. `geneAbundance()` is considered decreased in the case of a homozygous or heterozygous gene deletion, and increased in the case of a DNA amplification mutation. Events in which a protein binds to the promoter of a gene can be represented using the `geneAbundance()` function.

Example - promoter binding event represented using geneAbundance

```
complex(p(HGNC:TP53), g(HGNC:CDKN1A))
```

In the above example, the p53 protein binds the CDKN1A gene.

2.1.5. microRNAAbundance(), m()

`microRNAAbundance(ns:v)` or `m(ns:v)` denotes the abundance of the processed, functional microRNA designated by the value v in the namespace ns.

Example - microRNA abundance

```
m(HGNC:MIR21)
```

2.1.6. proteinAbundance(), p()

`proteinAbundance(ns:v)` or `p(ns:v)` denotes the abundance of the protein designated by the value v in the namespace ns, where v references a gene or a named protein family.

Examples - protein abundances

```
p(HGNC:AKT1)
p(SFAM:"AKT Family")
```

2.1.7. rnaAbundance(), r()

`rnaAbundance(ns:v)` or `r(ns:v)` denotes the abundance of the RNA designated by the value v in the namespace ns, where v references a gene. This function refers to all RNA designated by ns:v, regardless of splicing, editing, or polyadenylation stage.

Example - RNA abundance

```
r(HGNC:AKT1)
```

2.2. Abundance Modifier Functions

The following BEL functions are special functions that can be used only as an argument within an abundance function. These functions modify the abundance to specify sequence variations (gene, RNA, microRNA, protein), post-translational modifications (protein), fragment resulting from proteolytic processing (protein), or cellular location (most abundance types).

2.2.1. Protein Modifications

proteinModification(), pmod()

The `proteinModification()` or `pmod()` function can be used only as an argument within a `proteinAbundance()` function to indicate modification of the specified protein. Multiple modifications can be applied to the same protein abundance. Modified protein abundance term expressions have the general form:

```
p(ns:protein_value, pmod(ns:type_value, <code>, <pos>))
```

type_value (required) is a namespace value for the type of modification , **<code>** (optional) is a single-letter or three-letter code for one of the twenty standard amino acids, and <pos> (optional) is the position at which the modification occurs based on the reference sequence for the protein. If **<pos>** is omitted, then the position of the modification is unspecified. If both **<code>** and **<pos>** are omitted, then the residue and position of the modification are unspecified. NOTE - A default BEL namespace includes commonly used protein modification types.

Examples

AKT1 phosphorylated at Serine 473

default BEL namespace and 1-letter amino acid code:

```
p(HGNC:AKT1, pmod(Ph, S, 473))
```

default BEL namespace and 3-letter amino acid code:

```
p(HGNC:AKT1, pmod(Ph, Ser, 473))
```

PSI-MOD namespace and 3-letter amino acid code:

```
p(HGNC:AKT1, pmod(MOD:PhosRes, Ser, 473))
```

MAPK1 phosphorylated at both Threonine 185 and Tyrosine 187

default BEL namespace and 3-letter amino acid code:

```
p(HGNC:MAPK1, pmod(Ph, Thr, 185), pmod(Ph, Tyr, 187))
```

Palmitoylated HRAS

HRAS palmitoylated at an unspecified residue. Default BEL namespace:

```
p(HGNC:HRAS, pmod(Palm))
```

Modification Types Provided in Default BEL Namespace

Additional modification types can be requested as needed, or an external vocabulary can be used. Like other BEL namespace values, these modification types can be equivalenced to values in other vocabularies.

Label	Synonym
Ac	acetylation
ADPRib	ADP-ribosylation
ADP-rybosylation	adenosine diphosphoribosyl
Farn	farnesylation
Gerger	geranylgeranylation
Glyco	glycosylation
Hy	hydroxylation
ISG	ISGylation
ISG15-protein conjugation	Me
methylation	Me1
monomethylation	mono-methylation
Me2	dimethylation
di-methylation	Me3
trimethylation	tri-methylation
Myr	myristoylation
Nedd	neddylation
NGlyco	N-linked glycosylation
NO	Nitrosylation
OGlyco	O-linked glycosylation
Palm	palmitoylation
Ph	phosphorylation
Sulf	sulfation
sulphation	sulfur addition
sulphur addition	sulfonation
sulphonation	Sumo
SUMOylation	Ub
ubiquitination	ubiquitinylation
ubiquitylation	UbK48
Lysine 48-linked polyubiquitination	UbK63
Lysine 63-linked polyubiquitination	UbMono
monoubiquitination	UbPoly

Supported One- and Three-letter Amino Acid Codes

Amino Acid	1-Letter Code	3-Letter Code
Alanine	A	Ala
Arginine	R	Arg
Asparagine	N	Asn
Aspartic Acid	D	Asp
Cysteine	C	Cys
Glutamic Acid	E	Glu
Glutamine	Q	Gln
Glycine	G	Gly
Histidine	H	His
Isoleucine	I	Ile
Leucine	L	Leu
Lysine	K	Lys
Methionine	M	Met
Phenylalanine	F	Phe
Proline	P	Pro
Serine	S	Ser
Threonine	T	Thr
Tryptophan	W	Trp
Tyrosine	Y	Tyr
Valine	V	Val

2.2.2. Variants

variant(""), var("")

The `variant("<expression>")` or `var("<expression>")` function can be used as an argument within a `geneAbundance()`, `rnaAbundance()`, `microRNAAbundance()`, or `proteinAbundance()` to indicate a sequence variant of the specified abundance. The `var("")` function takes HGVS variant description expression, e.g., for a substitution, insertion, or deletion variant. Multiple `var("")` arguments may be applied to an abundance term.

Protein examples

reference allele

```
p(HGNC:CFTR, var("="))
```

This is different than `p(HGNC:CFTR)`, the root protein abundance, which includes all variants.

unspecified variant

```
p(HGNC:CFTR, var("?"))
```

substitution

```
p(HGNC:CFTR, var("p.Gly576Ala"))
p(REF:NP_000483.3, var("p.Gly576Ala"))
```

CFTR substitution variant Glycine 576 Alanine (HGVS *NP_000483.3:p.Gly576Ala*). Because a specific position is referenced, a namespace value for a non-ambiguous sequence like the RefSeq ID in the lower example is preferred over the HGNC gene symbol. The *p.* within the `var("")` expression indicates that the numbering is based on a protein sequence.

deletion

```
p(HGNC:CFTR, var("p.Phe508del"))
p(REF:NP_000483.3, var("p.Phe508del"))
```

CFTR ΔF508 variant (HGVS *NP_000483.3:p.Phe508del*). Because a specific position is referenced, a namespace value for a non-ambiguous sequence like the RefSeq ID in the lower example is preferred over the HGNC gene symbol. The *p.* within the `var("")` expression indicates that the numbering is based on a protein reference sequence.

frameshift

```
p(HGNC:CFTR, var("p.Thr1220Lysfs"))
p(REF:NP_000483.3, var("p.Thr1220Lysfs"))
```

CFTR frameshift variant *(*HGVS *NP_000483.3:p.Thr1220Lysfs*7).* Because a specific position is referenced, a namespace value for a non-ambiguous sequence like the RefSeq ID in the lower example is preferred over the HGNC gene symbol. The *p.* within the `var("")` expression indicates that the numbering is based on a protein reference sequence.

DNA (gene) examples

These are all representations of CFTR *ΔF508*.

SNP

```
g(SNP:rs113993960, var("delCTT"))
```

chromosome

```
g(REF:NC_000007.13, var("g.117199646_117199648delCTT"))
```

gene - coding DNA reference sequence

```
g(HGNC:CFTR, var("c.1521_1523delCTT"))
g(REF:NM_000492.3, var("c.1521_1523delCTT"))
```

Because a specific position is referenced, a namespace value for a non-ambiguous sequence like the RefSeq ID in the lower example is preferred over the HGNC gene symbol. The *c.* within the `var("")` expression indicates that the numbering is based on a coding DNA reference sequence.The coding DNA reference sequence covers the part of the transcript that is translated into protein; numbering starts at the A of the initiating ATG codon, and ends at the last nucleotide of the translation stop codon.

RNA examples

These are all representations of CFTR *ΔF508*.

coding reference sequence

```
r(HGNC:CFTR, var("c.1521_1523delCTT"))
r(REF:NM_000492.3, var("c.1521_1523delCTT"))
```

Because a specific position is referenced, a namespace value for a non-ambiguous sequence like the RefSeq ID in the lower example is preferred over the HGNC gene symbol. The *c.* within the `var("")` expression indicates that the numbering is based on a coding DNA reference sequence. The coding DNA reference sequence covers the part of the transcript that is translated into protein; numbering starts at the A of the initiating ATG codon, and ends at the last nucleotide of the translation stop codon.

RNA reference sequence

```
r(HGNC:CFTR, var("r.1653_1655delcuu"))
r(REF:NM_000492.3, var("r.1653_1655delcuu"))
```

Because a specific position is referenced, a namespace value for a non-ambiguous sequence like the RefSeq ID in the lower example is preferred over the HGNC gene symbol. The *r.* within the `var("")` expression indicates that the numbering is based on an RNA reference sequence. The RNA reference sequence covers the entire transcript except for the poly A-tail; numbering starts at the transcription initiation site and ends at the transcription termination site.

2.2.3. Proteolytic fragments

fragment(), frag()

The `fragment()` or `frag()` function can be used within a `proteinAbundance()` term to specify a protein fragment, e.g., a product of proteolytic cleavage. Protein fragment expressions take the general form:

```
p(ns:v, frag(<range>, <descriptor>))
```

where `<range>` (required) is an amino acid range, and `<descriptor>` (optional) is any additional distinguishing information like fragment size or name.

Examples

For these examples, *HGNC:YFG* is 'your favorite gene'. For the first four examples, only the `<range>` argument is used. The last examples include use of the optional `<descriptor>`.

fragment with known start/stop

```
p(HGNC:YFG, frag(5_20))
```

amino-terminal fragment of unknown length

```
p(HGNC:YFG, frag(1_?))
```

carboxyl-terminal fragment of unknown length

```
p(HGNC:YFG, frag(?_*))
```

fragment with unknown start/stop

```
p(HGNC:YFG, frag(?))
```

fragment with unknown start/stop and a descriptor

```
p(HGNC:YFG, frag(?, 55kD))
```

2.2.4. Cellular location

location(), loc()

`location()` or `loc()` can be used as an argument within any abundance function except

`compositeAbundance()` to represent a distinct subset of the abundance at that location. Location subsets of abundances have the general form:

```
f(ns:v, loc(ns:v))
```

Examples

Cytoplasmic pool of AKT1 protein

```
p(HGNC:AKT1, loc(MESHCS:Cytoplasm))
```

Endoplasmic Reticulum pool of Ca^{2+}

```
a(CHEBI:"calcium(2+)", loc(GOCC:"endoplasmic reticulum"))
```

2.3. Process Functions

The following BEL Functions represent classes of events or phenomena taking place at the level of the cell or the organism which do not correspond to molecular abundances, but instead to a biological process like angiogenesis or a pathology like cancer.

2.3.1. biologicalProcess(), bp()

`biologicalProcess(ns:v)` or `bp(ns:v)` denotes the process or population of events designated by the value v in the namespace ns.

Examples

```
bp(GOBP:"cell cycle arrest")
bp(GOBP:angiogenesis)
```

2.3.2. pathology(), path()

`pathology(ns:v)` or `path(ns:v)` denotes the disease or pathology process designated by the value v in the namespace ns. The +pathology()` function is included to facilitate the distinction of pathologies from other biological processes because of their importance in many potential applications in the life sciences.

Examples

```
pathology(MESHD:"Pulmonary Disease, Chronic Obstructive")
pathology(MESHD:adenocarcinoma)
```

which one class of abundance is transformed or changed into a second class of abundance by translocation, degradation, or participation in a reaction. All types of abundance terms except compositeAbundance() may be used within these transformation functions.

2.5.1. Translocations

BEL translocation functions include `translocation()` as well as `cellSurfaceExpression()` and `cellSecretion()`, two functions intended to provide a simple, standard means of expressing commonly represented translocations.

translocation(), tloc()

For the abundance term A, `translocation(<abundance>, fromLocation(ns1:v1), toLocation(ns2:v2))` or `tloc(<abundance>, fromLoc(ns1:v1), toLoc(ns2:v2))` denotes the frequency or number of events in which members of `<abundance>` move from the location designated by the value v1 in the namespace ns1 to the location designated by the value `v2` in the namespace `ns2`. Translocation is applied to represent events on the cellular scale, like endocytosis and movement of transcription factors from the cytoplasm to the nucleus. Special case translocations are handled by the BEL functions: `cellSecretion()`, `cellSurfaceExpression()`.

Example

endocytosis (translocation from the cell surface to the endosome) of the epidermal growth factor receptor (EGFR) protein can be represented as:

```
tloc(p(HGNC:EGFR), fromLoc(GOCC:"cell surface"), toLoc(GOCC:endosome))
```

cellSecretion(), sec()

For the abundance term A, `cellSecretion(<abundance>)` or `sec(<abundance>)` denotes the frequency or number of events in which members of `<abundance>` move from cells to regions outside of the cells. `cellSecretion(<abundance>` can be equivalently expressed as:

```
tloc(<abundance>, fromLoc(GOCC:intracellular), toLoc(GOCC:"extracellular space"))
```

The intent of the `cellSecretion()` function is to provide a simple, standard means of expressing a commonly represented translocation.

cellSurfaceExpression(), surf()

`cellSurfaceExpression(<abundance>)` or `surf(<abundance>)` denotes the frequency or abundance of events in which members of `<abundance>` move to the surface of cells. `cellSurfaceExpression(<abundance>)` can be equivalently expressed as:

```
tloc(<abundance>, fromLoc(GOCC:intracellular), toLoc(GOCC:"cell surface"))
```

RNA abundance of fusion with known breakpoints

```
r(fus(HGNC:TMPRSS2, "r.1_79", HGNC:ERG, "r.312_5034"))
```

The *r.* designation in the range fields indicates that the numbering uses the RNA sequence as the reference. RNA sequence numbering starts at the transcription initiation site. You use *c._* for g() fusions and *p._* for p() fusions. These *r.*, *c.*, and *p.* designations come from HGVS variation description convention.

RNA abundance of fusion with unspecified breakpoints

```
r(fus(HGNC:TMPRSS2, "?", HGNC:ERG, "?"))
```

3. BEL Relationships

The following BEL Relationship types are included in the BEL v2.0 language specification:

- Causal Relationships
- Correlative Relationships
- Genomic Relationships
- Other Relationships
- Deprecated Relationships

The most used BEL relationships should be the causal and correlative relationship categories. Relationships not used in the written BEL language, but introduced by the BEL Framework during compilation of a BEL network are not covered in this document.

3.1. Causal Relationships

These relationship types denote a causal relationship, or the absence of a causal relationship between a subject and an object term.

3.1.1. increases, →

For terms A and B, `A increases B` or `A → B` indicate that increases in A have been observed to cause increases in B.

`A increases B` also represents cases where decreases in A have been observed to cause decreases in B, for example, in recording the results of gene deletion or other inhibition experiments.

A is a BEL Term and B is either a BEL Term or a BEL Statement.

The `increases` relationship does not indicate that the changes in A are either necessary for changes in B, nor does it indicate that changes in A are sufficient to cause changes in B.

TIP	In most cases, these relationships will be introduced by the BEL Namespace resources, and are not needed for creation of BEL Statements and BEL Documents.

3.3.1. orthologous

For terms A and B, `A orthologous B` indicates that A and B represent entities in different species which are sequence similar and which are therefore presumed to share a common ancestor. For example,

```
g(HGNC:AKT1) orthologous g(MGI:AKT1)
```

indicates that the mouse and human AKT1 genes are orthologs.

3.3.2. transcribedTo, :>

For RNA abundance term R and gene abundance term G, `G transcribedTo R` or `G :> R` indicates that members of R are produced by the transcription of members of G. For example:

```
g(HGNC:AKT1) :> r(HGNC:AKT1)
```

indicates that the human AKT1 RNA is transcribed from the human AKT1 gene.

3.3.3. translatedTo, >>

For RNA abundance term R and protein abundance term P, `R translatedTo P` or `R >> P` indicates that members of P are produced by the translation of members of R. For example:

```
r(HGNC:AKT1) >> p(HGNC:AKT1)
```

indicates that AKT1 protein is produced by translation of AKT1 RNA.

3.4. Other Relationships

Additional miscellaneous relationship types. Icon In most cases, these relationships will be introduced by the BEL Namespace resources, and are not needed for creation of BEL Statements and BEL Documents.

3.4.1. hasMember

For term abundances A and B, `A hasMember B` designates B as a member class of A. A member class is a distinguished sub-class. A is defined as a group by all of the members assigned to it. The member classes may or may not be overlapping and may or may not entirely cover all instances of A. A term may not appear in both the subject and object of the same hasMember statement.

3.4.2. hasMembers

The `hasMembers` relationship is a special form which enables the assignment of multiple member classes in a single statement where the object of the statement is a set of abundance terms. A statement using `hasMembers` is exactly equivalent to multiple `hasMember` statements. A term may not appear in both the subject and object of the same `hasMembers` statement.

For the abundance terms A, B, C and D, `A hasMembers list(B, C, D)` indicates that A is defined by its member abundance classes B, C and D.

3.4.3. hasComponent

For complex abundance term A and abundance term B, `A hasComponent B` designates B as a component of A, that complexes that are instances of A have instances of B as possible components. Note that, the stoichiometry of A is not described, nor is it stated that B is a required component. The use of `hasComponent` relationships is complementary to the use of functionally composed complexes and is intended to enable the assignment of components to complexes designated by names in external vocabularies. The assignment of components can potentially enable the reconciliation of equivalent complexes at knowledge assembly time.

3.4.4. hasComponents

The `hasComponents` relationship is a special form which enables the assignment of multiple complex components in a single statement where the object of the statement is a set of abundance terms. A statement using `hasComponents` is exactly equivalent to multiple `hasComponent` statements. A term may not appear in both the subject and object of the same hasComponents statement.

For the abundance terms A, B, C and D, `A hasComponents list(B, C, D)` indicates that A has components B, C and D.

3.4.5. isA

For terms A and B, `A isA B` indicates that A is a subset of B.

All terms in BEL 1.0 represent classes, but given that classes implicitly have instances, `A isA B` is interpreted to mean that any instance of A must also be an instance of B. This relationship can be used to represent GO and MeSH hierarchies:

```
pathology(MESH:Psoriasis) isA pathology(MESH:"Skin Diseases")
```

3.4.6. subProcessOf

For process, activity, or transformation term A and process term P, `A subProcessOf P` indicates that instances of process P, by default, include one or more instances of A in their composition. For example, the reduction of HMG-CoA to mevalonate is a subprocess of cholesterol biosynthesis:

```
rxn(reactants(a(CHEBI:"(S)-3-hydroxy-3-methylglutaryl-CoA"),a(CHEBI:NADPH),
a(CHEBI:hydron)),\
 products(a(CHEBI:mevalonate), a(CHEBI:"CoA-SH"), a(CHEBI:"NADP(+)"))) subProcessOf\
 bp(GOBP:"cholesterol biosynthetic process")
```

3.5. Deprecated Relationships

WARNING	These BEL v1.0 relationships are supported in BEL v2.0, but are slated to be removed in the next major version.

3.5.1. analogous

For terms A and B, `A analogousTo B` indicates that A and B represent abundances or molecular activities which function in a similar manner, but do not share sequence similarity or a common ancestor.

3.5.2. biomarkerFor

For term A and process term P, `A biomarkerFor P` indicates that changes in or detection of A is used in some way to be a biomarker for pathology or biological process P.

3.5.3. prognosticBiomarkerFor

For term A and process term P, `A prognosticBiomarkerFor P` indicates that changes in or detection of A is used in some way to be a prognostic biomarker for the subsequent development of pathology or biological process P.

4. Appendices

Additional information supporting the BEL Language specification.

- Namespaces Used in Examples
- BEL Examples
- BEL Best Practices - Updated for BEL v2

4.1. Namespaces Used in Examples

Namespaces are a reference to the specific vocabulary that a value used in a BEL Term comes from. The examples in this documentation use the following set of BEL Namespaces (v20131211) to reference external ontologies and vocabularies:

Namespace Abbreviation	Namespace Description
EGID	Entrez Gene IDs

HGNC	HGNC human gene symbols
MGI	MGI mouse gene symbols
RGD	RGD rat gene symbols
SP	SwissProt accession numbers
MESHD	Medical Subject Heading Disease names
MESHCS	Medical Subject Heading Cellular Structure names
MESHPP	Medical Subject Heading Process names
CHEBI	Chemicals of Biological Interest names
GOBP	Gene Ontology Biological Process names
GOCC	Gene Ontology Cellular Component names
SCOMP	Selventa Named Complexes
SFAM	Selventa Protein Families

4.2. BEL Examples

The following pages contain examples of BEL Terms and BEL Statements. BEL Terms are used to represent biological entities including abundances and processes. These terms are used as the basis of BEL Statements that link one or more BEL Terms together with a relationship and/or additional context information to represent biological knowledge.

These examples are written in BEL Script format; see documentation for more information.

- BEL Term Examples
- BEL Statement Examples
- Other Examples

4.2.1. BEL Term Examples

- Abundance Term Examples
- Activity Term Examples
- Binding Interaction Term Examples
- Biological Processes and Pathologies Term Examples
- Post-Translationally Modified Protein Term Examples
- Transformation Term Examples (Reactions, Translocations, Degradation)
- Variant (Mutant) Protein Examples

Abundance Term Examples

Measurable entities like genes, RNAs, proteins, and small molecules are represented as abundances in BEL. BEL Terms for abundances have the general form `a(ns:v)`, where `a` is an abundance function, `ns` is `a` namespace reference and `v` is a value from the namespace vocabulary. See Namespaces Used in Examples.

- Chemicals and Small Molecules
- Genes, RNAs, and proteins
- Protein families
- microRNAs
- Complexes
- Composite abundances

Chemicals and Small Molecules

The general abundance function `abundance()` is used to represent abundances of chemicals, small molecules, and any other entities that cannot be represented by a more specific abundance function.

Examples

Long Form

```
abundance(CHEBI:"nitrogen atom")
abundance(CHEBI:"prostaglandin J2")
```

Short Form

```
a(CHEBI:"nitrogen atom")
a(CHEBI:"prostaglandin J2")
```

These BEL Terms represent the abundance of the entities specified by *nitrogen atom* and by *prostaglandin J2* in the CHEBI namespace.

Genes, RNAs, and proteins

The abundance functions `geneAbundance()`, `rnaAbundance()`, and `proteinAbundance()` are used with namespace values like HGNC human gene symbols, EntrezGene IDs, SwissProt accession numbers to designate the type of molecule represented.

Examples

Abundances of the gene, RNA, and protein encoded by the human AKT1 gene are represented as:

Long Form

```
geneAbundance(HGNC:AKT1)
rnaAbundance(HGNC:AKT1)
proteinAbundance(HGNC:AKT1)
```

Short Form

```
g(HGNC:AKT1)
r(HGNC:AKT1)
p(HGNC:AKT1)
```

These BEL Terms represent the gene, RNA, and protein abundances of the entity specified by *AKT1* in the HGNC namespace. Equivalent terms can be constructed using a corresponding value from a different namespace. For example, the abundance of the human AKT1 RNA can also be represented by referencing the EntrezGene ID or SwissProt accession namespaces:

```
r(EGID:207)
r(SP:P31749)
```

The BEL Framework identifies and merges corresponding terms created using different namespaces into a single term through namespace equivalencing.

Protein families

Protein families are used to represent a group of functionally similar proteins. For example, AKT1, AKT2, and AKT3 together form the AKT family. Like other proteins, abundances of protein families are represented using the `proteinAbundance()` function, with namespace values from the Selventa named protein families namespace.

Example

This term represents the protein abundance of the AKT protein family.

```
p(SFAM:"AKT Family")
```

microRNAs

The abundance function `microRNAAbundance()` is used to represent the fully processed, active form of a microRNA. The specific abundance functions allow distinct representations of the gene, RNA, and microRNA abundances for a given namespace value.

Example

These BEL Terms represent the abundances of the gene, RNA, and processed microRNA, respectively, for the entity specified by *Mir21* in the MGI mouse gene symbol namespace.

Long Form

```
geneAbundance(MGI:Mir21)
rnaAbundance(MGI:Mir21)
microRNAAbundance(MGI:Mir21)
```

Short Form

```
g(MGI:Mir21)
r(MGI:Mir21)
m(MGI:Mir21)
```

Complexes

The abundances of molecular complexes are represented using the `complexAbundance()` function. This function can take either a list of abundance terms or a value from a namespace of molecular complexes as its argument.

Example

Both BEL Terms represent the IkappaB kinase complex. The first by referencing a named protein complex within the GO Cellular Component namespace, and the second by enumerating the individual protein abundances that compose the IkappaB kinase complex, CHUK, IKBKB, and IKBKG.

Long Form

```
complexAbundance(GOCC:"IkappaB kinase complex")
complexAbundance(proteinAbundance(HGNC:CHUK), proteinAbundance(HGNC:IKBKB),
proteinAbundance(HGNC:IKBKG))
```

Short Form

```
complex(GOCC:"IkappaB kinase complex")
complex(p(HGNC:CHUK), p(HGNC:IKBKB), p(HGNC:IKBKG))
```

Composite abundances

Multiple abundance terms can be represented together as the subject of a BEL Statement by using the `compositeAbundance()` function. This function takes a list of abundances as its argument and is used when the individual abundances do not act alone, but rather synergize to produce an effect.

Example

This term represents the combined abundances of TGFB1 and IL6 proteins.

Long Form

```
compositeAbundance(proteinAbundance(HGNC:TGFB1), proteinAbundance(HGNC:IL6))
```

Short Form

```
composite(p(HGNC:TGFB1), p(HGNC:IL6))
```

Activity Term Examples

Term activity functions are applied to protein, complex, and RNA abundances to specify the frequency of events resulting from the molecular activity of the abundance. This distinction is particularly useful for proteins whose activities are regulated by post-translational modification. Specific activity types can be indicated using the `molecularActivity()` process modifier function. The default BEL namespace includes molecular activity values corresponding to the BEL v1.0 activity functions, and GO Molecular Function namespace values can be used to indicate more specific molecular activities.

- Non-Specified Activities
- Catalytic Activity
- Peptidase Activity
- G-proteins in the active (GTP-bound) state
- Transporter Activity
- Chaperone Activity
- Transcription Activity

Non-Specified Activities

If the type of molecular activity is not reported, it does not need to be specified. The `activity()` function is sufficient for distinguishing the frequency of events mediated by an abundance from the amount of the abundance. This term represents the ligand-bound activity of the human non-catalytic receptor protein TLR7.

Long Form

```
activity(proteinAbundance(HGNC:TLR7))
```

Short Form

```
act(p(HGNC:TLR7))
```

Long Form - GO Molecular Function Namespace

```
act(p(MGI:Trp53), ma(GOMF:"nucleic acid binding transcription factor activity"))
```

Binding Interaction Term Examples

The `complexAbundance()` function can be used to specify molecular interactions between abundances. This function can take either a list of abundances that define a molecular complex or a namespace value that represents a molecular complex (e.g., many GO Cellular Component values) as an argument. These examples demonstrate the use of the `complexAbundance()` function to represent protein-protein, protein-chemical, and protein-DNA interactions.

- Protein – protein interactions
- Protein – DNA interactions
- Protein – small molecule interactions

Protein – protein interactions

Example - protein-protein interaction as BEL statement

This statement represents that MTOR and AKT1S1 proteins physically interact. Note that this statement has only an object term and no subject term and relationship.

Long Form

```
SET Citation = {"PubMed", "Nat Cell Biol 2007 Mar 9(3) 316-23", "17277771"}
```

```
SET SupportingText = "Here, we identify PRAS40 (proline-rich Akt/PKB substrate
 40 kDa) as a novel mTOR binding partner"
```

```
# disambiguation PRAS40 = HGNC AKT1S1
```

```
complexAbundance(proteinAbundance(HGNC:AKT1S1), proteinAbundance(HGNC:MTOR))
```

Short Form

```
complex(p(HGNC:AKT1S1), p(HGNC:MTOR))
```

Example - protein-protein interaction as Statement object

Here, a protein-protein interaction is the object of a BEL Statement.This statement expresses that the MTOR and STAT3 proteins associate and that increases in the protein abundance of BMP4 can increase the abundance of the complex comprised of MTOR and STAT3.

Long Form

```
SET Citation = {"PubMed", "J Cell Biol. 2003 Jun 9;161(5):911-21.", "12796477"}
```

```
SET SupportingText = "Upon BMP4 treatment, the serine-threonine kinase
FKBP12/rapamycin-associated protein (FRAP), mammalian target of
rapamycin (mTOR), associates with Stat3 and facilitates STAT activation."
```

```
proteinAbundance(HGNC:BMP4) increases complexAbundance(proteinAbundance(HGNC:MTOR),
proteinAbundance(HGNC:STAT3))
```

Short Form

```
p(HGNC:BMP4) -> complex(p(HGNC:MTOR), p(HGNC:STAT3))
```

Protein – DNA interactions

Example - transcription factor protein binding to DNA

This statement expresses that STAT3 protein binds to the CCL11 gene DNA, and that this association is increased by IL17A.

Long Form

```
SET Citation = {"PubMed", "J Immunol 2009 Mar 15 182(6) 3357-65", "19265112"}
```

```
SET SupportingText = "IL-17A induced at 1 h a marked enrichment of
 STAT3- associated CCL11 promoter DNA"
```

```
proteinAbundance(HGNC:IL17A) increases \
 complexAbundance(proteinAbundance(HGNC:STAT3), geneAbundance(HGNC:CCL11))
```

Short Form

```
p(HGNC:IL17A) -> complex(p(HGNC:STAT3), g(HGNC:CCL11))
```

Protein – small molecule interactions

Example - protein binding to a small molecule

This statement represents that PIP3 binds AKT1 protein.

Long Form

```
SET Citation = {"PubMed", "Breast Cancer Res 2005 7(4) R394-401", "15987444"}
```

```
SET Evidence = "After PIP3 binding, Akt1 is activated"
```

```
# disambiguation PIP3 = CHEBI 1-phosphatidyl-1D-myo-inositol 3,4,5-trisphosphate
```

```
complexAbundance(abundance(CHEBI:"1-phosphatidyl-1D-myo-inositol 3,4,5-
trisphosphate"), proteinAbundance(HGNC:AKT1))
```

Short Form

```
complex(a(CHEBI:"1-phosphatidyl-1D-myo-inositol 3,4,5-trisphosphate"), p(HGNC:AKT1))
```

Biological Processes and Pathologies Term Examples

Biological phenomena that occur at the level of the cell or the organism are considered processes. These terms are represented by values from namespaces like GO and MeSH.

- Biological Processes
- Diseases and Pathologies

Biological Processes

Cellular senescence can be represented by:

Long Form

```
biologicalProcess(GOBP:"cellular senescence")
```

Short Form

```
bp(GOBP:"cellular senescence")
```

Diseases and Pathologies

Disease pathologies like muscle hypotonia can be represented by:

Long Form

```
pathology(MESHD:"Muscle Hypotonia")
```

Short Form

```
path(MESHD:"Muscle Hypotonia")
```

Post-Translationally Modified Protein Term Examples

The `proteinModification()` or `pmod()` function is used within a protein abundance to specify post-translational modifications. Types of post-translational modification are specified by a namespace value; the default BEL namespace provides many commonly used modification types. Abundances of modified proteins take the form `p(ns:v, pmod(ns:type_value, <code>, <pos>))`, where `<type>` (required) is the kind of modification, `<code>` (optional) is the one- or three- letter amino acid code for the modified residue, and `<pos>` (optional) is the sequence position of the modification.

- Hydroxylation
- Phosphorylation
- Acetylation
- Glycosylation
- Methylation
- Ubiquitination

Hydroxylation

This term represents the abundance of human HIF1A protein hydroxylated at asparagine 803.

Long Form

```
proteinAbundance(HGNC:HIF1A, proteinModification(Hy, Asn, 803))
```

Short Form

```
p(HGNC:HIF1A, pmod(Hy, N, 803))
```

Phosphorylation

This term represents the phosphorylation of the human AKT protein family at an unspecified amino acid residue.

```
p(SFAM:"AKT Family", pmod(Ph))
```

Acetylation

This term represents the abundance of mouse RELA protein acetylated at lysine 315.

```
p(MGI:Rela, pmod(Ac, Lys, 315))
```

Glycosylation

This term represents the abundance of human SP1 protein glycosylated at an unspecified amino acid residue.

```
p(HGNC:SP1, pmod(Glyco))
```

Methylation

This term represents the abundance of rat STAT1 protein methylated at an unspecified arginine residue:

```
p(RGD:STAT1, pmod(Me, Arg))
```

Ubiquitination

This term represents the abundance of human MYC protein ubiquitinated at an unspecified lysine residue:

```
p(HGNC:MYC, pmod(Ub, Lys))
```

Transformation Term Examples (Reactions, Translocations, Degradation)

- Reactions
- Translocations
- Degradation

Reactions

The `reaction()` or `rxn()` function expresses the transformation of products into reactants, each defined by a list of abundances.

Example

This BEL Term represents the reaction in which the reactants phosphoenolpyruvate and ADP are converted into pyruvate and ATP.

Long Form

```
reaction(reactants(abundance(CHEBI:phosphoenolpyruvate), abundance(CHEBI:ADP)),\
 products(abundance(CHEBI:pyruvate), abundance(CHEBI:ATP)))
```

Short Form

```
rxn(reactants(a(CHEBI:phophoenolpyruvate), a(CHEBI:ADP)),\
 products(a(CHEBI:pyruvate), a(CHEBI:ATP)))
```

Translocations

Translocations, or the movement of abundances from one location to another, are represented in BEL Terms by the `translocation()` or `tloc()` function. For convenience, the frequently used translocations of abundances from inside the cell to cell surface or extracellular space are represented by the `cellSurface()` and `cellSecretion()` functions, respectively.

Example

This term represents the event in which human NFE2L2 protein is translocated from the cytoplasm to the nucleus.

Long Form

```
translocation(proteinAbundance(HGNC:NFE2L2), fromLoc(MESHCS:Cytoplasm),
toLoc(MESHCS:"Cell Nucleus"))
```

Short Form

```
tloc(p(HGNC:NFE2L2), fromLoc(MESHCL:Cytoplasm), toLoc(MESHCL:"Cell Nucleus"))
```

Example - cell secretion

This term represents secretion of mouse IL6 protein.

Long Form

```
cellSecretion(proteinAbundance(MGI:Il6))
```

Short Form

```
sec(p(MGI:Il6))
```

Example - cell surface expression

This term represents cell surface expression of rat Fas protein.

Long Form

```
cellSurfaceExpression(proteinAbudance(RGD:Fas))
```

Short Form

```
surf(p(RGD:Fas))
```

Degradation

Events in which an abundance is degraded can be represented by the `degradation()` or `deg()` function.

Example

This term represents the degradation of MYC RNA. Degradation decreases the amount of the abundance - when degradation statements are compiled, a directlyDecreases relationship edge is added between the degradation term and the degraded entity.

Long Form

```
degradation(rnaAbundance(HGNC:MYC))
```

Short Form

```
deg(r(HGNC:MYC))
```

Variant (Mutant) Protein Examples

The abundances of mutated and variant proteins can be represented in BEL using the abundance modifier function `variant("")` and the other function `fusion()`.

- Amino Acid Substitutions
- Truncated Proteins
- Fusion Proteins

Amino Acid Substitutions

The abundances of proteins with amino acid sequence variations, such as those resulting from missense mutations or polymorphisms can be specified by using the `variant("")` or `var("")` function within a protein abundance term.

Example

Long Form

```
proteinAbundance(HGNC:PIK3CA, variant("p.Glu545Lys"))
```

Short Form

```
p(HGNC:PIK3CA, var("p.Glu545Lys"))
```

This term represents the abundance of the human PIK3CA protein in which the glutamic acid residue at position 545 has been substituted with a lysine.

Truncated Proteins

The abundances of proteins that are truncated by the introduction of a stop codon can be specified by using the `variant("")` or `var("")` function within a protein abundance term.

Example

Long Form

```
proteinAbundance(HGNC:ABCA1, variant("p.Arg1851*"))
```

Short Form

```
p(HGNC:ABCA1, var("p.Arg1851*"))
```

This term represents the abundance of human ABCA1 protein that has been truncated by substitution of Arginine 1851 with a stop codon.

Fusion Proteins

The abundances of fusion proteins resulting from chromosomal translocation mutations can be specified by using the `fusion()` or `fus()` function within a protein abundance term.

Example

Long Form

```
proteinAbundance(fusion(HGNC:BCR, "p.1_426", HGNC:JAK2, "p.812_1132"))
```

Short Form

```
p(fus(HGNC:BCR, "p.1_426", HGNC:JAK2, "p.812_1132"))
```

This term represents the abundance of a fusion protein of the 5' partner BCR and 3' partner JAK2,

with the breakpoint for BCR at amino acid 426 and JAK2 at 812. *p.* indicates that the protein sequence is used for the range coordinates provided. If the breakpoint is not specified, the fusion protein abundance can be represented as:

```
p(fus(HGNC:BCR, "?", HGNC:JAK2, "?"))
```

The `fusion()` function can also be used within `geneAbundance` and `rnaAbundance` terms to represent genes and RNAs modified by fusion mutations.

4.2.2. BEL Statement Examples

- Causal Statement Examples
- Correlative Statement Examples
- Direct Causal Statement Examples
- Nested Statement Example

Causal Statement Examples

Causal statements connect subject and object terms with a causal `increases`, `decreases`, or `causesNoChange` relationship. Subject terms can be an abundance or process (including activities and transformations) and object terms can be either an abundance, a process, or a second BEL Statement.

- Causal increase
- Causal decrease
- Causes no change

Causal increase

Example

These statements use the causal `increases` relationship. These statements are annotated with a citation and supporting evidence text, as well as with the cell line and species context for the experimental observations represented by the statements. These two statements represent the observation that increases in IL6 protein abundance cause increases in the RNA abundance of ENO1 and XBP1. These statements are annotated with CellLine and Species to indicate that the experimental observation was made in the context of the cell line "U266" and species "9606" (Homo sapiens).

Long Form

```
SET Citation = {"PubMed", "Int J Oncol 1999 Jul 15(1) 173-8", "10375612"}
```

```
SET SupportingText = "Northern blot analysis documented that two
 transcription factor genes chosen for further study, c-myc
 promoter-binding protein (MBP-1) and X-box binding protein 1
 (XBP-1), were up-regulated in U266 cells about 3-fold relative
 to the cell cycle-dependent beta-actin gene 12 h after IL-6
 treatment"
```

```
SET CellLine = "U266"
```

```
SET Species = "9606"
```

```
# disambiguation MBP-1 = HNGC ENO1
```

```
proteinAbundance(HGNC:IL6) increases rnaAbundance(HGNC:ENO1)
```

```
proteinAbundance(HGNC:IL6) increases rnaAbundance(HGNC:XBP1)
```

Short Form

```
p(HGNC:IL6) -> r(HGNC:ENO1)
```

```
p(HGNC:IL6) -> r(HGNC:XBP1)
```

Causal decrease

Example

This statement demonstrates a causal statement using the `decreases` relationship. The statement expresses that increases in the abundance of corticosteroid molecules cause decreases in the frequency or intensity of the biological process inflammation. This statement is annotated with an Anatomy and Disease to indicate that the relationship was observed in the context of the *cardiovascular system* and the disease *Stroke*.

Long Form

```
SET Citation = {"PubMed", "J Mol Med. 2003 Mar;81(3):168-74. Epub 2003 Mar 14.",
"12682725"}
```

This statement represents that abundance of protein designated by the name Nr2f2 in the MGI namespace is associated in an unspecified manner with the biological process angiogenesis.

Long Form

```
SET Citation = {"PubMed", "Mech Ageing Dev. 2004 Oct-Nov;125(10-11):719-32.",
"15541767"}
```

```
SET SupportingText = "COUP-TFII is involved in the angiogenic process in the
developing embryos."
```

```
# disambiguation - COUP-TFII refers to MGI Nr2f2
```

```
SET MeSHAnatomy = "Embryo, Mammalian"
```

```
proteinAbundance(MGI:Nr2f2) association biologicalProcess(GOBP:angiogenesis)
```

Short Form

```
p(MGI:NR2F2) -- bp(GOBP:angiogenesis)
```

Direct Causal Statement Examples

The following examples demonstrate the use of direct casual relationships in causal statements. The direct causal relationships `directlyIncreases` and `directlyDecreases` are special forms of the causal `increases` and `decreases` relationships where the mechanism of the causal relationship involves the physical interaction of entities related to the BEL Statement subject and object terms.

- Example - Ligand and Receptor
- Example - Kinase and Substrate
- Example - Catalyst and Reaction
- Example - Self-Referential Relationships
- Example - Direct Transcriptional Control

Example - Ligand and Receptor

In this example, the `directlyIncreases` relationship is used to represent activation of a receptor by its ligand. This statement expresses that amphiregulin (AREG) activates its receptor, the Epidermal Growth Factor Receptor (EGFR). This relationship is direct because ligands directly interact with their receptors.

Citation

Citations are a special type of annotation that references the knowledge source that reports the observation that the statement is based on. Citations are composed of a document type, a document name, a document reference ID, and an optional publication date, authors list and comment field. For example, the citation for a journal article indexed by PubMed would be encoded as:

```
SET Citation = {"PubMed", "Genes Cancer. 2010 Jun;1(6):560-567.", "21533016"}
```

The document name is a text string containing the reference information, the type is PubMed, and the document reference is the PubMed ID.

The citation for a Reactome pathway would be encoded as:

```
SET Citation = {"Online Resource", "p53-Dependent G1 DNA Damage Response",
"REACT_1625.1"}
```

In this case, the document name is the pathway name, the type is *Online Resource*, and the reference is the Reactome identifier.

Support (previously known as Supporting Text)

Support annotations provide the specific text that the statement is derived from. Text should come directly from the abstract or full text of the source referenced by the citation annotation. For example, a support line from the Reactome pathway cited above is:

```
SET Supporting = "The p53 protein activates the transcription of cyclin-dependent
kinase inhibitor, p21.
p21 inactivates the CyclinE:Cdk2 complexes, and prevent entry of the cell into S
phase, leading to G1 arrest."
```

Species

Species annotations indicate the species context for experimental observation represented by the statement. It is good practice to unambiguously assign species context to BEL Statements, even though many BEL Terms are derived from a species-specific namespace (e.g., HGNC, MGI, RGD). Species annotation uses the NCBI taxonomy ID:

```
SET Species = "9606"
```

Sets the species as Homo sapiens.

```
SET Species = "10090"
```

Sets the species as Mus musculus

```
SET Species = "10116"
```

Sets the species as Rattus norvegicus.

Other Annotation Types

Other types of annotations can be added to statements to indicate the context of the experimental observation supported by the statement, including cell line, cell type, and cellular location. For example:

```
SET Cell = "Adipocytes, White"
SET CellLine = "LoVo"
SET Disease = "Lupus Erythematosus, Systemic"
SET Anatomy = "Pulmonary Artery"
```

TIP	In a BEL Document each Annotation Type that will be used, except for Citation and SupportingText, must be defined in the document header, along with the values allowed for each.

Membership Assignment Examples

These examples demonstrate the assignment of members to groups. Because all BEL terms denote classes, membership in a group is an important special case where subsets of a class that define the class are designated.

TIP	The BEL Framework adds family members to protein families and complex components to named complexes during network compilation.

- Protein Family
- Complex Component

Protein Family

In this example, members of a protein family are assigned using the `hasMember` and `hasMembers` relationships.

The `hasMembers` relationship is used to assign a list of protein abundances as members of a protein family. This relationship is a syntactic convenience that is equivalent to the set of two statements using the `hasMember` relationship. These statements designate the protein abundances of MAPK8 and MAPK9 as members of the JNK MAPK protein family. The term representing the JNK family is a protein abundance based on the name *MAPK JNK Family* in the Selventa Protein Families namespace.

```
p(SFAM:"MAPK JNK Family") hasMembers list(p(HGNC:MAPK8), p(HGNC:MAPK9))
```

The `hasMember` relationship is used to assign individual protein abundances to a protein family.

```
p(SFAM:"MAPK JNK Family") hasMember p(HGNC:MAPK8)
```

```
p(SFAM:"MAPK JNK Family") hasMember p(HGNC:MAPK9)
```

Complex Component

In this example components are assigned to a named protein complex using the hasComponent and hasComponents relationships.

The `hasComponents` relationship is similar to the `hasMembers` relationship and is used to assign a list of abundances as components of a complex.These statements designate the protein abundances of RAD9A, RAD1, and HUS1 as components of the complex abundance of the *checkpoint clamp complex*.

```
complex(GOCC:"checkpoint clamp complex") hasComponents list(p(HGNC:RAD9A),
p(HGNC:RAD1), p(HGNC:HUS1))
```

The `hasComponent` relationship is used to assign individual abundances to a named protein complex.

```
complex(GOCC:"checkpoint clamp complex") hasComponent p(HGNC:RAD9A)
```

```
complex(GOCC:"checkpoint clamp complex") hasComponent p(HGNC:RAD1)
```

```
complex(GOCC:"checkpoint clamp complex") hasComponent p(HGNC:HUS1)
```

The single `hasComponents` statement is equivalent to the set of three `hasComponent` statements.

4.3. BEL Best Practices - Updated for BEL v2

These pages contain suggestions and guidelines for representing scientific findings in BEL.

- Representation of Experimental Data
- Statement Annotations
- Modified Proteins
- Reactions
- Protein-Protein Interactions
- Protein Families

experiments that use these constructs.

Non-phosphorylatable mutant

In this example, mutation of FOXO1 serine 256 to alanine is used to block phosphorylation at 256 (S256A), a site phosphorylated by AKT. The S256A mutation was found to impair phosphorylation of threonine 24 and serine 319 by AKT (PMID 11237865). We could represent this observation as follows:

```
p(HGNC:FOXO1, var("p.Ser256Ala")) =| p(HGNC:FOXO1, pmod(Ph, Ser, 256))
```

```
p(HGNC:FOXO1, var("p.Ser256Ala")) =| (p(SFAM:"AKT Family") => p(HGNC:FOXO1, pmod(Ph,
Thr, 24)))
```

```
p(HGNC:FOXO1, var("p.Ser256Ala")) =| (p(SFAM:"AKT Family") => p(HGNC:FOXO1, pmod(Ph,
Ser, 319)))
```

The first statement indicates that phosphorylation at S256 is blocked by mutation of S256 to alanine. The next two statements indicate that the S256A mutation decreases phosphorylation of FOXO1 threonine 24 and serine 319 by AKT. However, we are not generally interested in the effects of a lab-created mutant like S256A so much as the role of phosphorylation at serine 256 on phosphorylation of the other two sites. Thus, we recommend the following representation:

```
p(HGNC:FOXO1, pmod(Ph, Ser, 256)) =>(p(SFAM:"AKT Family") => p(HGNC:FOXO1, pmod(Ph,
Thr, 24)))
```

```
p(HGNC:FOXO1, pmod(Ph, Ser, 256)) =>(p(SFAM:"AKT Family") => p(HGNC:FOXO1, pmod(Ph,
Ser, 319)))
```

Here, the statements indicate that phosphorylation of FOXO1 at S256 increases the phosphorylation of T24 and S319 by the kinase activity of AKT. While both representations are accurate, the second version is better suited to integrating other information about the role of FOXO1 phosphorylation at S256 into a cohesive, traversable model.

How do I represent observations resulting from manipulation of two or more entities?

In some cases an experiment has a complex perturbation, where manipulations of multiple biological entities are required for an effect. Multiple BEL abundance terms can be represented together as the subject of a BEL Statement by using the `compositeAbundance()` or `composite()` function.

In this example, TGF-beta cooperates with IL-6 to generate T-helper 17 cells (PMID 17918200):

```
composite(p(MGI:Tgfb1), p(MGI:Il6)) -> bp(GOBP:"T-helper 17 cell differentiation")
```

If the two manipulated components are known to physically interact (such as a receptor and it's ligand), we recommend inferring their effects rather than using a composite term.

In this example, both Met and Hgf (the Met ligand) are required for increased expression of integrin Itgav RNA (PMID 16710476):

IMPORTANT

Not recommended:

```
composite(p(MGI:Hgf), p(MGI:Met)) -> r(MGI:Itgav)
```

Recommended:

```
kin(p(MGI:Met) ) -> r(MGI:Itgav)
```

```
p(MGI:Hgf) -> r(MGI:Itgav)
```

Because Hgf binds to and directly activates Met, the effect of Met and Hgf together on Itgav RNA expression can be inferred to result from Met activity.

How should I represent gene knock out or RNAi experiments?

Our general practice is to represent the subject term for experiments where the perturbation is a gene deletion or RNAi knockdown as the abundance of the corresponding protein.

Gene knockouts

In this example, mice with a gene deletion of Nfe2l2 express reduced mRNA of the glutathione S transferase Gsta1 compared to wild-type mice (PMID 11991805):

```
p(MGI:Nfe2l2) -> r(MGI:Gsta1)
```

RNA interference

In this example, knockdown of PTEN using RNA interference results in increased CDKN1A protein levels (PMID 17300726):

```
p(HGNC:PTEN) -| p(HGNC:CDKN1A)
```

We assume that the effects of PTEN RNAi are due to knock down of PTEN protein. Decreased PTEN protein resulting in increased CDKN1A protein is interpreted as PTEN decreases CDKN1A protein.

Correlative

If the observation comes from the comparison of human tumors grouped by the occurrence of a specific mutation, then the relationship should generally be expressed as correlative. In this case there is no experimental perturbation. In this example, most patient tumor samples with an EGFR L858R mutation were observed to exhibit a reduction in ERBB2 tyrosine 1248 phosphorylation compared to wild-type samples (PMID 18687633):

```
p(HGNC:EGFR, var("p.Leu858Arg")) negativeCorrelation p(HGNC:ERBB2, pmod(Ph,Tyr,1248))
```

In this case, no evidence is presented to suggest that the differences in ERBB2 phosphorylation are causally related to the EGFR mutation, only that the two observations are inversely correlated. Note that the subject and object terms are interchangeable for correlative relationships.

Causal

If the observation comes from the comparison of experimentally controlled states, like gene deletion, overexpression, or introduction of a mutant allele into a cell line or animal, the experimental perturbation can generally be represented as the subject term of a causal statement. In this example, DUSP6 RNA is observed to be upregulated in immortalized human bronchial epithelial cells transfected with EGFR mutant L858R, as compared to WT EGFR (PMID 16489012):

```
p(HGNC:EGFR, var("p.Leu858Arg")) -> r(HGNC:DUSP6)
```

In this case, the EGFR mutation is introduced as an experimentally-controlled perturbation.

Object Terms (Measurements)

How should I represent microarray data?

We record the results of experiments like microarrays and RT-PCR, which measure RNA abundances, by representing the object terms as RNA abundances. Only significant effects (e.g., meeting minimum criteria for fold change and statistical significance) should be recorded in BEL Statements.

In a causal BEL Statement, the subject term generally represents an experimentally manipulated entity while the object term represents a measured entity. Our general practice is to represent the object terms in BEL Statements with the terms most closely related to the experimental measurement.

This direct representation of the measurement in BEL supports the creation of KAMs to which 'omic data can be mapped directly and analyzed using automated reasoning applications like Whistle. Inference of the potential downstream consequences of RNA expression changes is supported by connection of RNA abundances to the corresponding proteins during knowledge network compilation

4.3.2. Statement Annotations

How do I annotate a relationship observed in multiple biological contexts?

Often, the scientific literature reports a relationship as occurring across several biological contexts.

Our general practice is to represent each observation with a separate statement. Several annotations can be used to describe the same context, e.g., 'lung' and 'fibroblast', but distinct BEL statements should be used to describe each experimental context that the relationship is observed in.

Example

PMID 18650932 - siRNA knockdown of the atypical PKC-interacting protein Par-4 (PAWR) increases phosphorlyation of AKT at Serine 473 in both human 293 and A549 cells.

"To test whether this is also true in human cells, we used a Par-4 siRNA to deplete endogenous Par-4 levels in human 293 cells and in the A549 human lung adenocarcinoma cell line. Cells were treated with control or Par-4-specific siRNAs, after which they were kept for 24 h in serum-free medium conditions and then stimulated with serum. Data in Figure 5E and F clearly demonstrate that the knockdown of Par-4 provokes enhanced serum-activated phospho-Akt-Ser473 levels in A549 and 293 human cells, respectively."

IMPORTANT

Not recommended:

```
SET CellLine = {A549, 293}
```

```
p(HGNC:PAWR) -| p(SFAM:"AKT Family", pmod(Ph, Ser, 473))
```

Recommended:

```
SET CellLine = A549
```

```
p(HGNC:PAWR) -| p(SFAM:"AKT Family", pmod(Ph, Ser, 473))
```

```
SET CellLine
```

```
p(HGNC:PAWR) -| p(SFAM:"AKT Family", pmod(Ph, Ser, 473))
```

4.3.3. Modified Proteins

- How do I represent a protein modification when specific information is not available?

- How do I represent a protein modification within a complex?
- How do I represent a situation where multiple phosphorylations are required for a protein's activity?
- How do I represent a situation where one protein modification initiates additional modifications?
- How do I represent removal of a protein modification (e.g., dephosphorylation, deubiquitination)?

How do I represent a protein modification when specific information is not available?

BEL terms for post-translational modifications of proteins specify the type of modification, the modified amino acid, and the position of the modified amino acid. The modified amino acid and position are not required, so protein modifications can be represented with less specific information.

Example

Human AKT1 protein modified by phosphorylation at serine 473

```
p(HGNC:AKT1, pmod(Ph, Ser, 473))
```

Human AKT1 protein modified by phosphorylation at an unspecified serine residue

```
p(HGNC:AKT1, pmod(Ph, Ser))
```

Human AKT1 protein that has been modified by phosphorylation at an unspecified amino acid residue

```
p(HGNC:AKT1, pmod(Ph))
```

As a general rule, if specific information is available, it should be used. In some cases, this involves investigation sections of a paper outside of the evidence text or other referenced papers to determine which specific modifications have been measured.

Non-specific protein modification terms have limited value in the context of a knowledge network. For example, phosphorylation at different sites of the same protein can have opposing effects. For example: "Akt-phosphorylated FOXO interacts with the ubiquitin ligase Skp2 and is targeted for proteasomal degradation" (PMID 15917664)

Example

Recommended:

6.1. Java

6.2. Ruby

BEL Parameter

The corresponding (indented) definition.

BEL Term

The corresponding (indented) definition.

BEL Statement

The corresponding (indented) definition.