Last updated: August 31, 2014

Patterns of RLIMS-P

Textual patterns, implemented with regular expressions and supporting rules, are used in RLIMS-P to extract trigger-argument relations. These relations are categorized into two cases:

* simple relations … relations between a predicate (e.g., “phosphorylate”) and its immediate argument (agent or theme)
* extended relations ... relations between a predicate and its argument reported nearby, which is not an immediate argument syntactically

**NOTE**

In the patterns introduced below, common variants of grammars (negation, coordination, and any optional phrases, such as appositive, parenthesized phrases, and prepositional phrases) and those of triggers (various prefixes, such as {hyper, hypo, under, mono, di, …}-phosphorylation) are not repeatedly presented. They are handled separately in RLIMS-P, e.g., simplification of input text before applying patterns. Anaphoric expressions and other linking expressions are also handled separately, as discussed in the paper, i.e., “<protein>” in the pattern could be a distinctive protein name or an anaphoric expression, such as “*this protein*”. The case of auto-phosphorylation requires an additional treatment, where the extracted substrate (or kinase) is treated also as the kinase (or substrate).

# Simple relations

## Trigger = ADJ

This section introduces patterns, in which the trigger is in the adjective form, e.g., “phosphorylated” as in “*phosphorylated IGFBP-1”*.

### Target = {agent, agent|theme}

This subsection explains patterns where the role of the target entity is an agent (or agent-or-theme, which is denoted as “agent|theme”). As shown below, the target entity may be found “inside” of the trigger phrase or “outside”.

#### Target-position: inside

* <protein>-catalyzed phosphorylated <head>

where <head> may be a specific protein name (see the below section, target=theme), region, amino acid, or general noun such as “protein”, “peptide”, and “site”, and the trigger, “phosphorylated”, may be abbreviated as “phospho-”.

Extract <protein> as “agent” of the trigger “phosphorylated”.
* <protein> phosphorylated <head>

Extract <protein> (or the phrase “its”) as the “agent” if <head> contains a protein name (e.g., “MAPK-phosphorylated p53” (MAPK must be the kinase, since the substrate is known to be p53) or, otherwise, it is extracted initially as “agent|theme” and the disambiguation step is deferred until all the other extraction methods are applied to this trigger, e.g., “MAPK phosphorylated regions”, which might mean regions phosphorylated by MAPK or regions of MAPK that are phosphorylated.

#### Target-position: outside

* phosphorylation <site> for <protein>

where <site> may be a phrase, such as “site”, or specific amino acids and/or positions.

Extract <protein> as agent

### Target={theme, agent|theme}

This subsection explains patterns where the role of the target entity is a theme (or sometimes agent-or-theme). As before, these targets may be found “inside” of the trigger phrase or “outside”.

#### Target-position: inside

* phosphorylated <head>
* phosphorylatable <head>

Extract <head> (i.e., a word(s) embedded in NP) as theme if it is protein or protein-part, or, otherwise, extract the entire NP as theme.
* <protein>’s phosphorylated <head>

Extract <protein> (or “its”) as theme. If <head> contain protein-part (e.g., amino acid and/or position), it should be extracted as well.

#### Target-position: outside

* phosphorylation <site> of <protein>

Extract <protein> as “agent|theme”.
* phosphorylated ... (or phosphorylation {site, protein, form, mutant, rate}) {of, in, on, at} <protein, protein-part, or un-typed phrase>

Extract <protein, protein-part, or un-typed phrase (any phrase at this position)> as “theme”.
* phosphorylated … (or phosphorylation ...) in {its, their, the} phosphorylated {form, state, condition}

Extract the last word (“form”) and/or the anaphor (“its”, “their”) as “theme” as appropriate.
* phosphorylated {substrate, protein, mutant}

Extract the entire phrase as theme.

## Trigger=NP (“phosphorylation”)

This section explains the case where the trigger is NP: “phosphorylation”.

### Target={agent, agent|theme}

#### Target-position: inside

* <protein>-catalyzed (or <protein>-site) phosphorylation

Extract <protein> as “agent”
* <protein> phosphorylation

Extract <protein> as “agent|theme”

#### Target-position: outside

* phosphorylation is {carried out, executed} {by, via} NP

Extract NP as “agent” if it is a protein. If a protein is embedded in NP, e.g., NP=“<protein> activation”, the embedded protein is recoded as “agent|theme”
* phosphorylation {by, via} NP

Just as above, extract NP as “agent” or “agent|protein”

In the above, if an agent could be extracted, then any protein extracted as “agent|theme” by other patterns are inferred as a theme, e.g., “<protein1> phosphorylation by <protein2>”, where <protein2> is extracted as an agent with the current pattern and, thus, protein1 initially extracted as “agent|theme” by the other pattern is known to be a theme.

### Target=theme

#### Target-position: inside

* {its, their} (<protein>) phosphorylation

Extract the anaphor as “theme” and the optional protein, if present, as “agent”
* <protein> phosphorylation

Extract <protein> as “agent|theme”

Target-position: left (outside of the trigger; to the left of it)

* <protein> {function, functionality} {requires, needs} phosphorylation

Extract <protein> as “theme”
* <protein-part> (or a possibly un-typed phrase ending with site, domain, motif, form) {of, for} phosphorylation

Extract <protein-part> as “theme”
* <protein, protein-part> <be> \* target of phosphorylation

where <be> represents {is, was, are were, …} and \* represents determiners, adjective, and other modifying phrases.

Extract <protein, protein-part> as “theme”

#### Target-position: right

* phosphorylation {at, in, of, on, within} <protein, protein-part>

Extract <protein, protein-part> as “theme”. If it is a protein, any embedded protein is further extracted as “agent”.

## Trigger=NP “enzyme”

### Target=agent

#### Target-position: left

* <protein> {function, work, act, behave} as <enzyme>

where <enzyme> is a noun phrase referring to kinase type, e.g., a phrase with the head word “kinase”.

Extract <protein> as “agent”
* <protein> {possess, exhibit, show, display, demonstrate} <enzyme> activity

Extract <protein> as “agent”

### Target=theme

#### Target-position: right

* ... kinase for <protein>

Extract <protein> as “theme”

## Trigger=NP <phosphate group>

### Target=theme

#### Target-position: left

* <protein> {contain, bear, possess, have, carry} <group>
* <protein> bind <group>
* <protein> be {modified, attached, conjugated} ({at, in} <protein-part>)? by <group>
* <protein-part> with <group>

where <group> is the phrase “{phosphate, phosphoryl, phosphonyl, phosphoamide, phosphodiester} group”.

Extract <protein, protein-part> as “theme”.

#### Target-position: right

* {carry, contain, posses, ...} <group> at <protein-part>
* <group> (be) {modified, attached, added, transferred, ...} {at, onto, on, to, with} <protein, protein-part>
* {modified, attached, added, transferred} ({at, in, on} ...) {by, to, onto} <group>
* {modification, attachment, addition, transfer, ...} of <group> (from ...)? {at, on, to, onto} <protein, protein-part>
* {modified, attached, added, transferred, ...} <group head> {at, on, to, onto} <protein>
* <group> {at, on} <protein>

Extract <protein, protein-part> as “theme”

## Trigger=NP “substrate”

### Target={agent, agent|theme}

#### Target-position: inside

* <protein> substrate

Extract <protein> as “agent|theme”.

#### Target-position: outside

* substrate {of, for} <protein>

Extract <protein> as “agent” if the “substrate” is preceded by “candidate” or, otherwise, “agent|theme”.

### Target=theme

* <protein> <be> a substrate

Extract <protein> as “theme”

* ... substrate

Extract the entire NP as “theme”. If the patter is preceded by “<protein>, one of”, include that protein as “theme” as well.

## Trigger=VP active voice

### Target=agent

#### Target-position=left

* <protein>’s {ability, potential, power} to phosphorylate
* <protein> ({exhibit, show, display, has} {ability, potential} to)? phosphorylate
* <protein>, phosphorylating
* <protein> <be> capable of phosphorylating

Extract <protein> as “agent”.

### Target=theme

#### Target-position=right

* {allow, enable, facilitate} (<protein1>)? to phosphorylate <protein2, protein-part>

Extract <protein2, protein-part> as “theme” and <protein1>, if present, as “agent”
* phosphorylate ({at, on})? <protein, protein-part>

Extract <protein, protein-part> as “theme”

* phosphorylate <protein, protein-part>

Extract <protein, protein-part> as “theme”.

## Trigger=VP passive voice

### Target=agent

* phosphorylated {by, with} <protein>
* phosphorylated in response to <protein>

Extract <protein> as “agent”

### Target=theme

* <protein, protein-part> <be> phosphorylated

Extract <protein, protein-part> as “theme”

## Trigger=VP others

### Target=theme

* <protein-part> phosphorylat\* (e.g., tyrosine phosphorylate)

Extract <protein-part> embedded in verb as “theme”

# Extended relations

There are common textual patterns, from which entities involved in a phosphorylation event can be known, although it may not be an explicit report of a phosphorylation event.

## Keyword=“undergo phosphorylation”

There are a set of verbs (e.g., “undergo”), whose subject can be considered as the theme of the phosphorylation event, if it is followed by the trigger word “phosphorylation”

### Target=theme

* <protein> {undergo, experience, indicate, show, display, exhibit} phosphorylation
* {bind} (to) <protein> in phosphorylation-{sensitive, dependent} {manner, way, fashion}

Extract <protein> as “theme”

## Keyword= “required”

A protein may be reported as a prerequisite of a phosphorylation event, or phosphorylation may be reported as prerequisite of another event pertaining to a protein.

### Target=agent

* <protein> be {required, identified} for phosphorylation

Extract <protein> as “inducer”

### Target=theme

* phosphorylation (<be>) {required, identified, needed} for <event> of <protein, protein-part>

where <event> represents some biological event pertaining to a protein, such as “localization”.

Extract <protein, protein-part> as “theme”

## Keyword= “activate via phosphorylation”

A protein can be inferred to cause phosphorylation of another protein or inferred to be a theme of a phosphorylation event.

### Target=agent

* <protein> {activate, cause, catalyze, enhance, evoke, induce, ...} <event> {by, via} phosphorylation

Extract <protein> as “agent”
* <protein1> <verb> <protein2> to mediate phosphorylation

where <verb> is “activate”, etc.

Extract <protein2>, if present, as “agent”, or <protein1> as “inducer” (kinase candidate)
* <protein> <verb in passive> to mediate phosphorylation

where <verb in passive> may represent any event pertaining to a protein, such as “was activated”.

Extract <protein> as “inducer”

* <protein> {contribute, lead} to ({enhancement, regulation, control} of) phosphorylation

Extract <protein> as “inducer”
* phosphorylation depends on <protein>

Extract <protein> as “inducer”
* <protein> play a role in phosphorylation

Extract <protein> as “inducer”

### Target=theme

* {inhibit, direct, prevent, trigger, recruit, affect, suppress, anchor} <event> of <protein> {by, via, upon, on, through, after} phosphorylation
* {By, Via} phosphorylation, <subject> {inhibit, ...} <protein>

Extract <protein> as “theme”

* phosphorylate <protein-part> to <verb> <protein>

Extract <protein-part> and <protein> as “theme”

* <protein> <verb in passive> {by, via} phosphorylation

where <verb in passive> is “is activated”, etc.

Extract <protein> as “theme”

## Trigger=ADJ

### Target=agent

* <protein> is {essential, mandatory, indispensable, necessary} for phosphorylation

Extract <protein> as “inducer”