BIOCREATIVE VI: Annotation manual of CHEMPROT interactions between Chemical Entities Mentions (CEMs) and Gene and Protein Related Objects (GPROs).

Version 6.0 (1st August 2017)

This document describes the guidelines used for the construction of the annotations of chemical-protein interactions (CHEMPROT) between Chemical Entities Mentions (CEMs) and Gene and Protein Related Objects (named as GPROs throughout this manuscript) of the BioCreative VI CHEMPROT corpus. It provides the basic details of the chemical-protein interaction annotation task and the conventions that should be followed during the corpus construction process. The CHEMPROT annotation guidelines have been refined after iterative cycles of annotations of sample documents. It also incorporated suggestions made by curators as well as observations of annotation inconsistencies encountered when comparing results from different human curators. In brief, the annotated CHEMPROT interactions include direct interactions (when a physical contact exits between a CEM and a GPRO, in most cases this GPRO being a protein or protein family and alters its function/activity) as well as indirect regulatory interactions between CEMs and GPROs (including genes, gene products (proteins, RNA), DNA/protein sequence elements and protein families, domains and complexes) that alter either the function or the quantity of the GPRO. The aim of the iterative manual annotation cycles was to improve the quality and consistency of the guidelines, in order to make them more intuitive and easier to follow. During the preparation process of the guidelines some rules had to be reformulated to make them more explicit and additional rules were added when necessary to better cover the practical annotation scenario and for being more complete.

The manual annotation task basically consisted of labeling or marking up manually through a customized web-interface the mentions of CHEMPROT interactions in text. The text that was labeled consisted of PubMed abstracts (titles and abstracts in English) from scientific papers published between 2005 and 2014. A subset of these abstracts comes from the previous CHEMDNER-BioCreative IV task for the annotation of CEMs. The collection was enriched with abstracts cited in the DrugBank database.[1,2] CEMs and GPROs were separately annotated according to the guidelines described for the CHEMDNER-BioCreative IV and CHEMDNER-Patents-BioCreative V tasks, respectively. **These guidelines for CHEMPROT provide curation rules to evaluate if a sentence within an abstract contains a description of a chemical-protein interaction**; in particular, if sufficient detail/evidence is provided for co-mentioned CEMs and GPROs. Additionally, it describes curation rules and definitions to assign each identified CHEMPROT to any of the **5 classes and 16 subclasses** in Table 1, below.

CLASS	SUBCLASS	DEFINITION
PART_OF (tag: part_of)		CEM that are structurally related to a GPRO: e.g. specific amino acid residues of a protein.
REGULATOR (tag: undef_regul)		CEM that clearly regulates a GPRO, but for which there is no further information on whether the regulation is direct or indirect.
	UPREGULATOR (aka INCREMENTS) **(tag: undef_regul_up)**	CEM that increments a GPRO signal, without any insight on the mechanism. Warning! Despite this subclass is named UPREGULATOR, it comprises positive regulation (increments): direct protein activation by binding or indirect up-regulation of GPRO expression/protein levels or post translational modification).
	DOWNREGULATOR (aka DECREASES) **(tag: undef_regul_down)**	CEM that decreases a GPRO signal, without any insight on the mechanism. Warning! Despite this subclass is named DOWNREGULATOR, it comprises negative regulation (decrements): direct protein inhibition by binding or indirect down-regulation of GPRO expression/protein levels or post translational modification).
DIRECT REGULATOR (tag: direct_regul)		Binder/Ligand: CEM that directly binds to a GPRO (typically a protein) through a direct physical interaction and changes its activity/function (typically a protein activity/function).
	INHIBITOR **(tag: direct_inhi)**	CEM that binds to a GPRO (typically a protein) and decreases its activity.
	ACTIVATOR **(tag: direct_acti)**	CEM that binds to a GPRO (typically a protein) and increases its activity. Conceptual synonyms are Stimulator, Inducer, Potentiator and Enhancer.
	ANTAGONIST **(tag: direct_anta)**	CEM that reduces the action of another CEM, generally an agonist. Many antagonists act at the same **receptor** macromolecule as the agonist.
	AGONIST **(tag: direct_agon)**	CEM that binds to a **receptor** and alters the receptor state resulting in a biological response. Conventional agonists increase receptor activity, whereas *inverse agonists* reduce it. If no information is provided on whether the CEM activates or reduces GPRO activity, this general subclass should be assigned (instead of AGONIST-ACTIVATOR and AGONIST-INHIBITOR, below).
	AGONIST-ACTIVATOR **(tag: direct_agon_acti)**	Agonists that bind to a **receptor** and increase its biological response. Typically, for full agonists and most partial agonists (depending on concentration).
	AGONIST-INHIBITOR **(tag: direct_agon_inhi)**	Agonists that bind to a **receptor** and decrease its biological response. Typically, for inverse agonists.
	MODULATOR **(tag: direct_modu)**	CEM that acts as **allosteric modulator**, compound that increases or decreases the action of an (primary

CLASS	SUBCLASS	DEFINITION
		or orthosteric) agonist or antagonist by combining with a distinct (allosteric or allotropic) site on the receptor macromolecule. If no information is available on whether the CEM activates or reduces GPRO activity, this general subclass should be assigned (instead of MODULATOR-ACTIVATOR or MODULATOR-INHIBITOR, below).
	MODULATOR-ACTIVATOR **(tag: direct_modu_acti)**	CEM that acts as allosteric modulator and increases GPRO activity. It should be clearly stated or inferred that the response is increased, otherwise assign the CEM to the MODULATOR subclass.
	MODULATOR-INHIBITOR **(tag: direct_modu_inhi)**	CEM that acts as allosteric modulator and decreases GPRO activity. It should be clearly stated or inferred that the response is decreased, otherwise assign the CEM to the MODULATOR subclass.
	COFACTOR **(tag: direct_cofac)**	CEM that is required for a protein's biological activity to happen
	SUBSTRATE **(tag: direct_subs)**	CEM upon which a GPRO (typically protein) acts. It should be understood as the substrate of a reaction carried out by a protein ("reactant") or as transporter substrate.
	PRODUCT OF **(tag: direct_prod)**	CEM that is a product of enzymatic reaction or a transporter.
	SUBSTRATE / PRODUCT OF **(tag: direct_subsprod)**	CEM that is both, substrate and product of enzymatic reaction.
INDIRECT REGULATOR (tag: indirect_regul)		CEM that indirectly regulates a GPRO activity or quantity by interacting with cellular components other than the GPRO in a way that alters GPRO protein activity or quantity. Protein activity is typically altered via post translational modifications (PTMs) like e.g. phosphorylations or by altered protein-protein interactions. Protein quantity is typically determined by gene expression (e.g. transcription or gene transcript stability), protein expression and/or stability, or by protein release/uptake). This class is reserved for indirect interactions between a CEM and a GPRO. If no information on the effect is provided (upregulator or downregulator), this class should be assigned.
	INDIRECT UPREGULATOR **(tag: indirect_regul_up)**	CEM upregulates GPRO via other target. GPRO is upregulated by CEM as a result from CEMs effect on other target which results in upregulation of GPRO.
	INDIRECT DOWNREGULATOR **(tag: indirect_regul_down)**	CEM downregulates GPRO via other target. GPRO is downregulated by CEM as a result from CEMs effect on other target which results in downregulation of GPRO.
NOT		This class should be used to define the NEGATIVE

CLASS	SUBCLASS	DEFINITION
(tag: not)		occurrence of a chemical-protein interaction, without providing any further information on the specific negative CHEMPROT class or class

Table 1. CHEMPROT interaction classes and subclasses defined for the CHEMPROT task of BioCreative VI. For each CHEMPROT class and subclass a short description is provided. Illustrative example cases are listed below.

On this guideline, CEMs appear background colored in green (CEM) and GPROs in blue (GPRO). Explicit CHEMPROT interaction terms (e.g. agonist) or exemplary words from which the interaction is inferred are in gray (e.g. affinity). These terms in gray should be interpreted as indications for the annotation, but not as terms to be explicitly annotated with the interface. *Feature terms* associated to interactions in the sentence that provide further details on the annotation class are underlined (e.g. covalent). Exemplary terms that should not be labeled as CEMs, GPROs or CHEMPROT interactions are bordered: CEM.

In the examples below, not all CHEMPROT interactions found within a sentence are described, but only those that properly exemplify the case under discussion.

2. CHEMPROT Interactions

The definition of chemical-protein interactions that were annotated for the CHEMPROT task of BioCreative VI was primarily concerned with capturing those types of relationships that are of practical relevance, both for 1) end users of the extracted data as well as 2) for the named entity recognition systems. Therefore the covered chemical protein interactions had to be annotated at a sufficient level of granularity to be able to determine whether the labeled interaction can or cannot be linked to a specific well-defined relationship between a chemical and a protein/gene. The annotation carried out for the CHEMPROT task was exhaustive for the types of interactions previously specified. This implies that mentions of other relationships between chemicals and GPROs (e.g. phenotypic and biological responses) should not be labeled as CHEMPROTs (see rule N1). The CHEMPROT interactions were defined following the concept "what a CEM does to a GPRO" (CEM → GPRO direction) and not the opposite direction (GPRO → CEM direction) ("what a GPRO does to a CEM", see rule N5).

In order to establish a homogenous nomenclature and avoid redundant class definitions, we reviewed several chemical repositories that integrate chemical – biology information (e.g. **DrugBank**[1,2], **Therapeutic Targets Database** (**TTD**)[3,4] and **ChEMBL**[5,6]), assay normalization ontologies (**BAO**)[7,8] and previously existing formalizations for the annotation of relationships: Biological Expression Language (**BEL**) developed for Track 4 of the BioCreative challenge,[9,10] Curation guidelines for transcription regulation interactions (DNA-binding transcription factor-target gene interaction[11], and **SIGNOR**, a database of causal relationships between biological entities.[12,13] Each of these resources inspired the subclasses definitions of the DIRECT REGULATOR class (e.g. DrugBank, ChEMBL, BAO and SIGNOR) or the INDIRECT REGULATOR class (e.g. BEL, Curation guidelines for transcription regulation interactions and SIGNOR). For example, DrugBank relationships for drugs included a total of 22 definitions, some of them overlapping with CHEMPROT subclasses (e.g. "Inhibitor", "Antagonist", "Agonist",…), some of them being regarded as highly specific for the purpose of this task (e.g. "intercalation", "cross-linking/alkylation") or referring to biological roles (e.g. "Antibody", "Incorporation into and Destabilization") and others, partially overlapping between them (e.g. "Binder" and "Ligand"), that were merged into a single class. Concerning indirect regulatory aspects, the five classes of casual relationships between a subject and an object term defined by BEL ("decreases", "directlyDecreases", "increases", "directlyIncreases" and "causesNoChange")[10] were highly inspiring. Subclasses definitions of pharmacological modes of action were defined according to the UPHAR/BPS Guide to Pharmacology in 2016.[14,15]

The annotation process itself also relied heavily on common sense and domain background knowledge. A prerequisite in order to be able to carry out the manual annotation task is that the annotators must have an academic training in chemistry, biology (molecular biology, genetics) or biochemistry and with expertise in Medicinal Chemistry and Pharmacology projects to make sure the annotations are correct and of high quality.

This also allowed us to have shorter and more compact annotation rules rather than requiring very detailed guidelines for non-experts. A few general rules on how to proceed with the annotation of CHEMPROT interactions are listed below:

G0. Not all pairs of CEM and GPRO relationships must be annotated.

In a sentence, there can be a list of CEMs and a list of GPROs and not all pairs of relationships should be annotated. The CHEMPROT relationship should be annotated only for those pairs for which there is a relationship. Do not label cases in which the relationship between a CEM and GPRO is not clear at all because very general, vague terms describe the association. For example, they might appear co-mentioned in a sentence, but nothing else can be inferred.

Example from PubMed ID 10372729:

There was no change with regard to dobutamine (beta1-adrenoceptor sensitivity) or clonidine, (alpha2-adrenoceptor sensitivity) or to beta1-adrenoceptor density.

Dobutamine : alpha2-adrenoceptor → NOT LABEL

Nothing is stated about this CEM-GPRO pair.

Furthermore, sinapic acid reduced hepatic hydroxyproline content, which correlated with a reduction in the expression of type I collagen mRNA and histological analysis of collagen in liver tissue.

sinapic acid : type I collagen → INDIRECT DOWNREGULATOR

sinapic acid : collagen → NOT LABEL

general effects on "histological analysis" are not annotated.

Effects of nucleotides on N-acetyl-d-glucosamine 2-epimerases (renin-binding proteins): comparative biochemical studies.

Nucleotides : N-acetyl-d-glucosamine 2-epimerases → NOT LABEL

Nucleotides : renin-binding proteins → NOT LABEL

Only general effects where studied, without any idea on the relationship. Note that this type of introductory sentence is very common in many abstracts.

5-HT1A receptor function in normal subjects on clinical doses of fluoxetine: blunted temperature and hormone responses to ipsapirone challenge.

fluoxetine : 5-HT1A → REGULATOR

ipsapirone : 5-HT1A → NOT LABEL

Success of pyridostigmine, physostigmine, eptastigmine and phosphotriesterase treatments in acute sarin intoxication.

pyridostigmine : phosphotriesterase → NOT LABEL

There is no relationship between this CEM and this GPRO.

sarin : phosphotriesterase → NOT LABEL

Effects of GPROs on CEMs are not annotated (see rule N5, below).

G1. Very restricted use of external knowledge sources

CHEMPROT interactions should be inferred from the sentence as a whole, and with the support of the whole abstract in doubtful cases. Consultation of external knowledge resources (UniProt, Wikipedia, DrugBank,…) should be restricted at maximum, and only recommended for very specific doubts on CEMs and GPROs. **In case of doubt about specific subclass assignations, the annotation should be restricted at the highest class level.**

G2. Unclear mentions

If the annotator is not sure about a given CHEMPROT interaction, the corresponding mention should not be labeled.

G3. Guidelines under-specification

In case the annotator encounters cases of mention types that could be related to chemical-protein interactions but the guidelines do not specify their labeling, these should be reported together with examples to request for refinement of the annotation rules.

G4. Each CHEMPROT interaction mention can only be marked with a unique label within a given up-level class

This means that a specific CHEMPROT interaction in a sentence cannot be labeled for instance as DIRECT REGULATOR and AGONIST at the same time (only one label within each up-level class). If possible, CHEMPROT interactions should be annotated at the highest level of granularity (**i.e. subclass preferred over class**) based on the information extracted from the sentence.

For example, different sentences within the same abstract may define a CEM as a DIRECT REGULATOR (e.g. binder, without any further information) or as an AGONIST of a GPRO. Then, each CHEMPROT interaction should be tagged accordingly, even if the CEM-GPRO interaction is differently labeled (up-level class and more granular class) across different sentences of the same abstract.

Example of PubMed ID: 10448105:

We investigated the effect of changing the length and degree of unsaturation of the fatty acyl chain of N-(3-methoxy-4-hydroxy)-benzyl-cis-9-octadecenoamide (olvanil), a ligand of vanilloid receptors, on its capability to: (i) inhibit anandamide-facilitated transport into cells and enzymatic hydrolysis, (ii) bind to CB1 and CB2 cannabinoid receptors, and (iii) activate the VR1 vanilloid receptor

Olvanil:vanilloid receptors → DIRECT REGULATOR

Olvanil:CB1 and CB2 cannabinoid receptors → DIRECT REGULATOR

Olvanil:VR1 vanilloid receptor → ACTIVATOR (subclass of DIRECT REGULATOR)

Example for different sentences in Abstract with PubMed ID 19057128:

Captopril attenuates matrix metalloproteinase-2 and -9 in monocrotaline-induced right ventricular hypertrophy in rats.

Captopril : matrix metalloproteinase-2 and -9 → DOWNREGULATOR

monocrotaline : matrix metalloproteinase-2 and -9 → NOT LABEL

MMP-2 and MMP-9 expressions and activities in right ventricles increased significantly in monocrotaline-injected rats and captopril inhibited them.

Captopril : MMP-2 INHIBITOR and INDIRECT DOWNREGULATOR

Captopril : MMP-9 INHIBITOR and INDIRECT DOWNREGULATOR

Monocrotaline : MMP-2 ACTIVATOR and INDIRECT UPREGULATOR

Monocrotaline : MMP-9 ACTIVATOR and INDIRECT UPREGULATOR

These findings indicate that captopril attenuates the development of monocrotaline-induced right ventricular hypertrophy in association with inhibition of MMP-2 and MMP-9 in rats.

Captopril : MMP-2 → INHIBITOR

Captopril : MMP-9 → INHIBITOR

Monocrotaline : MMP-2 → NOT LABEL

Monocrotaline : MMP-9 → NOT LABEL

As exemplified above, there might be cases in which a CEM interacts with a GPRO in two different ways, and they should be annotated (BRAT Annotation allows it).

Example from PubMed ID 15573147:

We previously reported that chronic administration of doxazosin, an Alpha1-adrenoceptor (AR) antagonist, causes an up-regulation in the mRNA expression of all three alpha1-AR subtypes in the rat prostate.

Doxazosin: Alpha1-adrenoceptor → ANTAGONIST (within DIRECT REGULATOR).

Doxazosin: AR → ANTAGONIST (within DIRECT REGULATOR).

Doxazosin: Alpha1-AR → INDIRECT UPREGULATOR

Example from PubMed ID 15573147:

Whereas the addition of apomorphine in the low micromolar range produced an irreversible activation of the channel, application of higher concentrations caused a reversible voltage-dependent inhibition of heterologously expressed TRPA1 channels, resulting from a reduction of single-channel open times.

apomorphine : TRPA1 → INHIBITOR

apomorphine : TRPA1 → ACTIVATOR

In PubMed ID 16336943:

In Chinese hamster ovary (CHO) cells stably expressing 5-HT2 receptors, aripiprazole displayed a dual agonist/antagonist profile for 5-HT2C receptor (VNI isoform) mediated calcium signaling (EC50 1070 nM, IC50 281 nM).

aripiprazole : 5-HT2C → AGONIST

aripiprazole : 5-HT2C → ANTAGONIST

G5. Explicit and implicit CHEMPROT interactions

CHEMPROT annotations between a valid (annotated) CEM and a valid GPRO should be inferred from the total content of the sentence, either stated explicitly or implicitly. In doubtful cases, implicit relationships may be deducible from the sentence only by domain experts, not just from the expressed semantics alone.

Example for PubMed ID 23358194:

There was a significant overlap (42 genes) between the 3 and 30 ppb differentially expressed gene lists, with two of these genes (CYP17A1 and SAMHD1) present in all three atrazine treatments.

Atrazine : CYP17A1 → INDIRECT REGULATOR

Atrazine : SAMHD1 → INDIRECT REGULATOR

Example for PubMed ID 10027850:

The involvement of 5-HT1A receptors in the antiaggressive actions of these drugs was confirmed by showing that the selective 5-HT1A receptor antagonist WAY-100635 (N-[2-[4-(2-methoxyphenyl)-1-piperazinyl]ethyl]-N-(2-pyridinyl)cyclohexanecarboxamide trihydrochloride), which was inactive alone, fully prevented the antiaggressive effects of alnespirone, 8-OH-DPAT, and buspirone and partly reversed those of ipsapirone and eltoprazine.

WAY-100635 : 5-HT1A receptor → ANTAGONIST

The antagonist relationship is explicitly mentioned.

alnespirone/8-OH-DPAT/buspirone/ipsapirone/… : 5-HT1A receptor → AGONIST

For the case of the last mentioned CEMs (such as alnespirone), the relationship AGONIST is not explicitly mentioned. However, the expert user can implicitly deduce that these compounds are agonists, as it is explicitly mentioned that antagonists revert the effects of these compounds. Therefore, implicitly these compounds are to be annotated as AGONISTS (general class, because nothing is mentioned on their effects).

G6. Speculation and hypothesis

Speculations and hypotheses represent valid interaction description as we understand that the authors found this worthwhile of investigation and, hence, that there are reasons to believe that there may be an effect. Linguistic markers of speculations, hedging or uncertainties (e.g. expressions like 'might be ', 'could be' , 'may indicate') should not be considered to judge a particular relation, but have to rely on the biological/biochemical evidence provided in the sentence.

These results suggest that MK-801 induced upregulation of NMDA (NR2B) receptor subunit might be mediated by tyrosine kinases.

MK-801 : NMDA (NR2B) receptor subunit → INDIRECT UPREGULATOR

This relationship is clearly established, and also its indirect nature (as a potential mediator is mentioned).

MK-801 : tyrosine kinases → REGULATOR

The relationship is “might be mediated”, so there is a relationship, although it is no clear whether there is physical interaction or not (DIRECT or INDIRECT) and no clear

response on whether MK-801 upregulates (increases response) /downregulates (decreases response) tyrosine kinases.

Example of PubMed ID 23479193:

Synthesis and in-silico studies of some diaryltriazole derivatives as potential cyclooxygenase inhibitors

Diaryltriazole : cyclooxygenase → INHIBITOR

Example for PubMed ID 8100523:

They either start with Tyr-Ala, His-Ala or His-Ser which might be in part potential targets for dipeptidyl-peptidase IV, a highly specialized aminopeptidase removing dipeptides only from peptides with N-terminal penultimate proline or alanine.

Tyr-Ala : dipeptidyl-peptidase IV → SUBSTRATE

(same for His-Ala and His-Ser)

G7. Annotation of Feature Terms

Feature terms that detail further information for the specific interaction, such as the (ir)reversible character of binders, mechanism (competitive, non-competitive), selectivity, time-dependent character…are to be annotated as NOTES (see below, 5. BRAT Annotation).

In the following, definitions and exemplary cases of CHEMPROT interactions are detailed.

Class 1. PART_OF Class

For CEMs that are structurally related to a GPRO: e.g. amino acid residues of a protein.

Identification of functionally important cysteine residues of the human renin-binding protein as the enzyme N-acetyl-D-glucosamine 2-epimerase

cysteine : human renin-binding protein → PART_OF

cysteine : N-acetyl-D-glucosamine 2-epimerase → PART_OF

Mutations of aromatic residues in the first transmembrane helix impair signalling by the secretin receptor.

aromatic : secretin receptor → PART_OF

L-carnitine binds to a preformed pocket in the active site tunnel of carnitine acetyltransferase aligned with His(322)

His : carnitine acetyltransferase → PART_OF

Class 2. REGULATOR Class

CEMs that clearly regulate a GPRO, but for which there is no further information on whether the regulation is direct or indirect.

Testosterone and dihydrotestosterone act via the androgen receptor but a defective receptor function results in different degrees of genital malformations.

Testosterone : androgen receptor in that case the words “act via” adds an association of type REGULATOR between them, despite it is not clear at all (could be a binder/ligand or could be downstream the pathway).

Glycylsarcosine coadministration could inhibit the uptake of cefadroxil in PEPT2(+/+) mice (p < 0.01) but not PEPT2(-/-) mice.

glycylsarcosine : PEPT2 → REGULATOR

Example from PubMed ID: 23535185

The inhibitory effect of galangin on theses pro-inflammatory cytokines was related with c-Jun N-terminal kinases, and p38 mitogen-activated protein kinase, nuclear factor-κB, and caspase-1.

galangin : *c-Jun* → REGULATOR

galangin : p38 → REGULATOR

galangin : nuclear factor-κB → REGULATOR

galangin : caspase-1 → REGULATOR

There was no change with regard to dobutamine (beta1-adrenoceptor sensitivity) or clonidine, (alpha2-adrenoceptor sensitivity) or to beta1-adrenoceptor density.

Dobutamine : beta1-adrenoceptor → REGULATOR

Clonidine : alpha2-adrenoceptor → REGULATOR

Furthermore, the data provide evidence for a major involvement of these 5-HT1A receptors in the modulation of aggressive behavior by 8-OH-DPAT, ipsapirone, buspirone, and eltoprazine.

8-OH-DPAT : 5-HT1A receptors → REGULATOR

2.1. UPREGULATOR (AKA INCREMENTS)

CEM up-regulates GPRO, without any insight on the mechanism (direct or indirect). Warning! Despite this subclass is named UPREGULATOR, it comprises positive regulation (increments): direct protein activation by binding or indirect up-regulation of GPRO expression/protein levels or post translational modification).

Example from PubMed ID: 23282998

Results of the present study also showed that arsenic caused cytotoxicity by elevating morphological alterations, TUNEL-positive nuclei, caspase-3 activity and DNA damage and reducing cell adhesion and cell proliferation in a time-dependent manner.

arsenic : caspase-3 → UPREGULATOR

AMP-activated protein kinase (AMPK) is activated by metformin, phenformin, and the AMP mimetic, 5-aminoimidazole-4-carboxamide-1-beta-D-ribofuranoside (AICAR)

Metformin : AMP-activated protein kinase → UPREGULATOR

2.2. DOWNREGULATOR (AKA DECREASES)

CEM down-regulates GPRO, without any insight on the mechanism (direct or indirect). Warning! Despite this subclass is named DOWNREGULATOR, it comprises negative regulation (decrements): direct protein inhibition by binding or indirect down-regulation of GPRO expression/protein levels or post translational modification).

Example from PubMed ID: 10503933

Plasma alanine amino transferase activity was reduced by GVG treatment and this was not further modified by cocaine administration.

CVG : alanine amino tranferase → DOWNREGULATOR

Class 3. DIRECT REGULATOR

CEM that directly binds to a GPRO (typically a protein), i.e. there is a direct physical interaction between CEM and GPRO. **GPRO responses (especially of the type increase activity / decrease activity mediated) should only be annotated if there is no doubt on the existence of a direct interaction between the CEM and GPRO. Otherwise, assign the relationship as of class REGULATOR (class 2).** By default, without any further information on the specific type of binding or mode of action (e.g. inhibitor, agonist…), relations of this type should be annotated at this highest level. Features providing further information (irreversible binding, reversible binding, competitive binding, non-competitive binding, saturation binding, covalent binding) should also be annotated.

Preincubation of PPARgamma with L-764406 prevented binding of the [3H]TZD, suggesting a covalent interaction with the receptor; in addition, structurally related analogues of L-764406, which would be predicted not to interact with PPARgamma in a covalent fashion, did not displace [3H]TZD binding to PPARgamma.

L-764406 : PPARgamma → DIRECT REGULATOR

(covalent should be annotated as a feature term)

3H]TZD : PPARgamma → DIRECT REGULATOR

Note: for the concept "structurally related analogues of L-764406", according to the CEM guidelines, only "L-764406" would be annotated. Therefore, the whole concept of analogues and their potential interactions are not annotated.

L-carnitine binds to a preformed pocket in the active site tunnel of carnitine acetyltransferase aligned with His(322)

L-carnitine : carnitine acetyltransferase → DIRECT REGULATOR

We have identified two novel compounds (RTI 3021-012 and RTI 3021-022) that demonstrate similar affinities for human progesterone receptor (PR) and display equivalent antiprogestenic activity

RTI 3021-012 : human progesterone receptor → DIRECT REGULATOR

Using this approach, a novel non-TZD compound (L-764406) was shown to be a potent (apparent binding IC50 of 70 nM) PPARgamma ligand

L-764406 : PPARgamma → DIRECT REGULATOR

Different sub-classes within the DIRECT REGULATOR class:

3.1. INHIBITOR

CEM that binds to a GPRO (typically a protein) and decreases its activity. As some subclasses below (e.g. inverse agonists, negative allosteric modulators) also refer to "decreased GPRO activity", the INHIBITOR class should be used for cases in which the specific mechanism is unknown or clearly these terms do not apply because they do not refer to a receptor or allosterism. If mechanistic information is available, the more specific subclass assignation (e.g. modulator or antagonist or agonist-inhibitor) should be assigned. Features providing further information (competitive inhibition, uncompetitive inhibition, non-competitive inhibition, partial inhibition, reversible inhibition, irreversible inhibition, tight binding inhibition, time dependent inhibition,…) should also be annotated.

The results show that chronic treatment of cortical neurons with tyrosine kinase inhibitor (genistein) resulted in downregulation of the NR2B subunit polypeptide levels, while daidzein, an inactive analog of genistein, did not alter the levels of NR2B subunit, implying that tyrosine kinases may be involved in the regulation of the NMDA NR2B subunit content.

genistein : tyrosine kinase → INHIBITOR

Exemestane is a safe drug due to its steroidal structure, which blocks aromatase at a different site to nonsteroidal AIs (eg., anastrozole and letrozole)

Exemestane : aromatase → INHIBITOR

anastrozole : aromatase → INHIBITOR

letrozole : aromatase → INHIBITOR

3.2. ACTIVATOR

CEM that binds to a GPRO (typically a protein) and increases its activity. This class, named as ACTIVATOR should also include close synonyms and concepts such as STIMULATOR,

INDUCER, POTENTIATOR, ENHANCER… in those cases in which an increase in CEM causes an increase in GPRO activity (through direct physical interaction). As for the INHIBITOR class, if mechanistic information is provided on the activation (e.g. Agonist or positive allosteric modulation), then, the more specific subclass having mechanistic information should be assigned. Features providing further information (partial activation, full-activation,…) should also be annotated.

These selective antiaggressive actions of alnespirone are mediated by stimulating 5-HT1A receptors, presumably the somatodendritic autoreceptors at the raphe nuclei

alnespirone : 5-HT1A → ACTIVATOR

In that case, general knowledge establishes that alnespirone is an agonist. However, without any mention on the mechanism in the full abstract (despite it is in fact mentioned, PubMed Id 10027850), the ACTIVATOR class should be assigned. External resources must not be used to establish the specific mechanism if not mentioned in the abstract.

A-769662 directly stimulated partially purified rat liver AMPK (EC50 = 0.8 μM) and inhibited fatty acid synthesis in primary rat hepatocytes (IC50 = 3.2 μM)

A-769662 : AMPK → ACTIVATOR

In that case, the word *directly* defines that there is physical interaction.

3.3. ANTAGONIST

CEM that reduces the action of another CEM, generally an agonist. Many antagonists act at the same **receptor** macromolecule as the agonist. In pharmacology, antagonists have affinity but no efficacy for their cognate receptors. Also, they are commonly denominated BLOCKERS.

Chronic treatment of cortical neurons with the NMDA receptor antagonist, (+)-5-methyl-10,11-dihydro-5H-dibenzo[a, d]cycloheptane-5,10-iminemaleate (MK-801) enhanced the membrane associated tyrosine kinase activity and upregulated the NR2B receptor subunit

MK-801 : NMDA receptor → ANTAGONIST

5-methyl-10,11-dihydro-5H-dibenzo[a, d]cycloheptane-5,10-iminemaleate : NMDA receptor → ANTAGONIST

To help determine whether NMDA-Rs mediating feeding might contain these subunits, we conducted behavioral studies using LHA-administered ifenprodil, an antagonist selective for NR2A- and/or NR2B-containing NMDA-Rs at the doses we used (0.001-100 nmol)

ifenprodil : NMDA-Rs → ANTAGONIST

The selective CGRP-receptor antagonist 8-37 CGRP, but not the cannabinoid CB1 receptor blocker SR141716A, inhibited the vasodilator effect of anandamide

SR141716A : CB1 receptor → ANTAGONIST

Note: 8-37 CGRP is a peptide of > 15 aminoacids. According to the CEM guidelines, it should not be annotated as CEM. Again, same as above, the explicit word "Blocker" does not help to distinguish between "inhibitors" and "antagonists". However, using domain knowledge, it can be inferred that it is an ANTAGONIST.

3.4. AGONIST

A CEM that binds to a **receptor** and alters the receptor state resulting in a biological response. Conventional agonists increase receptor activity, whereas *inverse agonists* reduce it. If no information is provided on whether the CEM activates or reduces GPRO activity, this general subclass should be assigned (instead of AGONIST-ACTIVATOR and AGONIST-INHIBITOR, below). Features providing further information (allosteric agonism, partial agonism, full agonism, inverse agonism, allosteric antagonism, selective agonist, non-selective agonist) should also be annotated.

Example from PubMed ID: 10025919

The P2Y2 agonists ATP and UTP stimulated a small release of PGE2 that was potentiated after pretreatment with rHuIL-1alpha.

ATP : P2Y2 → AGONIST

Example from PubMed ID: 10036860

A selective 5-HT1B/1D agonist sumatriptan (0.01-10 mumol/L) caused nitrite production.

sumatriptan : 5-HT1B/1D → AGONIST

3.5. AGONIST-ACTIVATOR

Agonists that bind to a **receptor** and increase its biological response. Typically, for full agonists and partial agonists.

Example from PubMed ID: 16336943

In human embryonic kidney (HEK) cells transiently expressing 5-HT2C receptor isoforms, aripiprazole exhibited full agonism at the unedited INI, but partial agonism at the partially edited VNI and fully edited VSV isoforms (EC50s of 571, 1086 and 2099 nM, respectively).

aripiprazole : 5-HT2C → AGONIST

(general class for the general isoform class)

aripiprazole : INI → AGONIST-ACTIVATOR

full agonism information is provided, therefore, activation.

aripiprazole : VNI/VSV → AGONIST

(no further information on whether it increases activity or not)

3.6. AGONIST-INHIBITOR

Agonists that bind to a **receptor** and decrease its biological response. Typically, for inverse agonists.

Example from PubMed ID: 16336943

In contrast, while lacking agonist activity at the VNI and VSV, olanzapine showed inverse agonism at the INI isoform (IC50 594 nM), reaching a maximal attenuation of 20%.

olanzapine : INI → AGONIST-INHIBITOR

3.7. MODULATOR

CEM that acts as **allosteric modulator**, compound that increases or decreases the action of an (primary or orthosteric) agonist or antagonist by combining with a distinct (allosteric or allotropic) site on the receptor macromolecule. If no information is available on whether the CEM activates or reduces GPRO activity, this general subclass should be assigned (instead of MODULATOR-ACTIVATOR or MODULATOR-INHIBITOR, below).

Example from PubMed ID: 16489449

The present study shows that cannabidiol is an allosteric modulator at mu and delta opioid receptors.

cannabidiol : mu and delta opioid receptors → MODULATOR

3.8. MODULATOR-ACTIVATOR

CEM that acts as **allosteric modulator** and increases GPRO activity. It should be clearly stated or inferred that the response is **increased**, otherwise assign the CEM to the MODULATOR subclass.

Example from PubMed ID: 15608073

We found that 3-cyano-N-(1,3-diphenyl-1H-pyrazol-5-yl)benzamide (CDPPB), is a potent and selective positive allosteric modulator of the metabotropic glutamate receptor subtype 5 (mGluR5).

3-cyano-N-(1,3-diphenyl-1H-pyrazol-5-yl)benzamide : mGluR5

→ MODULATOR –ACTIVATOR

CDPPB : mGluR5 → MODULATOR –ACTIVATOR

3.9. MODULATOR-INHIBITOR

CEM that acts as **allosteric modulator** and decreases GPRO activity. It should be clearly stated or inferred that the response is **decreased**, otherwise assign the CEM to the MODULATOR subclass.

Example from PubMed ID: 28560482

The crystal structure of mGlu5 in the complex with the negative allosteric modulator mavoglurant was recently reported, providing a fundamental model for designing new allosteric modulators.

mavoglurant : mGlu5 → MODULATOR - INHIBITOR

3.10. COFACTOR

CEMs that are required for a protein's biological activity to happen. In case of doubt, label the corresponding CEM as of class SUBSTRATE.

Adenosine kinase (AK) is an enzyme responsible for converting endogenous adenosine (ADO) to adenosine monophosphate (AMP) in an adenosine triphosphate- (ATP-) dependent manner

adenosine triphosphate : Adenosine kinase → COFACTOR

ATP : Adenosine kinase → COFACTOR

3.11. SUBSTRATE

CEM upon which a GPRO (typically protein) acts. It should be understood as the substrate of a reaction carried out by a protein ("reactant") or as transporter substrate. Features providing further information (transporter substrate, endogenous,…) should also be annotated.

Adenosine kinase (AK) is an enzyme responsible for converting endogenous adenosine (ADO) to adenosine monophosphate (AMP) in an adenosine triphosphate- (ATP-) dependent manner

Adenosine : Adenosine kinase → SUBSTRATE

ADO : Adenosine kinase → SUBSTRATE

The recently cloned organic cation transporter, OCTN2, isolated as a homologue of OCTN1, has been shown to be of physiological importance in the renal tubular reabsorption of filtered L-carnitine as a high-affinity Na+ carnitine transporter in man

L-carnitine : OCTN2 → SUBSTRATE (Transporter Substrate)

Note: the relationship between OCTN1 and L-carnitine should not be labeled as the term "isolated as a homologue" is not enough as to provide a relationship between them.

The uptake of ergothioneine mediated by OCTN1 and of L-carnitine mediated by OCTN2 was decreased during oxaliplatin exposure

Ergothioneine : OCTN1 → SUBSTRATE (Transporter Substrate)

L-carnitine : OCTN2 → SUBSTRATE (Transporter Substrate)

oxaliplatin : OCTN1 → DOWNREGULATOR

oxaliplatin : OCTN2 → DOWNREGULATOR

These findings demonstrate that PEPT2 is the primary transporter responsible for cefadroxil uptake in the choroid plexus.

cefadroxil : PEPT2 → SUBSTRATE

3.12. PRODUCT OF

CEM as the product of an enzymatic reaction or a transporter.

Adenosine kinase (AK) is an enzyme responsible for converting endogenous adenosine (ADO) to adenosine monophosphate (AMP) in an adenosine triphosphate- (ATP-) dependent manner

adenosine monophosphate : Adenosine kinase → PRODUCT OF

AMP : Adenosine kinase → PRODUCT OF

Hydrogen sulfide (H(2)S), a regulatory gaseous molecule that is endogenously synthesized by cystathionine gamma-lyase (CSE) and/or cystathionine beta-synthase (CBS) from L-cysteine (L-Cys) metabolism

Hydrogen sulfide : cystathionine gamma-lyase → PRODUCT OF

L-cysteine : cystathionine gamma-lyase → SUBSTRATE

3.13. SUBSTRATE / PRODUCT OF

CEMs that are both, substrate and products of enzymatic reactions.

Carnitine acetyltransferase catalyzes the interchange between L-carnitine and acetyl-L-carnitine

L-carnitine : Carnitine acetyltransferase → SUBSTRATE / PRODUCT OF

acetyl-L-carnitine : Carnitine acetyltransferase → SUBSTRATE / PRODUCT OF

Class 4. INDIRECT REGULATOR

CEM that indirectly regulates a GPRO activity or quantity by interacting with cellular components other than the GPRO in a way that alters GPRO protein activity or quantity. Protein activity is typically altered via post translational modifications (PTMs) like e.g. phosphorylations or by altered protein-protein interactions. Protein quantity is typically determined by gene expression (e.g. transcription or gene transcript stability), protein expression and/or stability, or by protein release/uptake. This class is reserved for non-direct interactions between a CEM and a GPRO. In order to ease the annotation response, all these responses are grouped within this class.

By default, without additional information on the response (increase or decrease), relationships of this type should be annotated at this highest level. Features providing further information (time-dependent regulation,…) should also be annotated.

Chronic treatment of cortical neurons with the NMDA receptor antagonist, (+)-5-methyl-10,11-dihydro-5H-dibenzo[a, d]cycloheptane-5,10-iminemaleate (MK-801) enhanced the membrane associated tyrosine kinase activity and upregulated the NR2B receptor subunit

MK-801 : tyrosine kinase → UPREGULATOR

(class REGULATOR as strictly it is not possible to infer whether the relationship is DIRECT or INDIRECT)

MK-801 : NR2B receptor subunit → INDIRECT UPREGULATOR

(class INDIRECT REGULATOR of protein levels)

Regulation of glucocorticoid receptor-mRNA in human blood cells by amitriptyline

Amitriptyline : glucocorticoid receptor → INDIRECT REGULATOR

No further information on the effect.

Different sub-classes within the INDIRECT REGULATOR class:

4.1. INDIRECT UPREGULATOR

Upregulator / inducer / stimulator / enhancer. CEMs that induce/stimulate/enhance the frequency, rate or extent of gene expression or transcription, protein expression, protein release/uptake or protein functions. Increases in CEM cause an increase in the corresponding GPRO (or its response) and also cases in which a decrease of CEM causes a decrease in the corresponding GPRO.

L-764406 exhibited partial agonist activity in cells expressing a chimeric receptor containing the PPARgamma LBD and a cognate reporter gene and also induced the expression of the adipocyte-specific gene aP2 in 3T3-L1 cells.

L-764406 : aP2 → INDIRECT UPREGULATOR

Their activation by nucleotide agonists (ADP and ATP for P2Y1; ATP and UTP for P2Y2) elevates [Ca2+]i and moderately induces expression of the c-fos proto-oncogene.

ADP : c-fos → INDIRECT UPREGULATOR

ATP : c-fos → INDIRECT UPREGULATOR

Annotate it twice, once for each mention

UTP : c-fos → INDIRECT UPREGULATOR

4.2. INDIRECT DOWNREGULATOR

CEMs that decrease gene expression or transcription, protein expression, protein release / uptake or indirectly, protein functions. Increases in CEM cause a decrease in the corresponding

GPRO (or indirectly its response) and also cases in which decreases of CEM causes an increment in the corresponding GPRO.

Oxytocin receptor stimulation of phospholipase C is inhibited by cAMP

cAMP : phospholipase C → INDIRECT DOWNREGULATOR

cAMP : oxytocin receptor → REGULATOR

We have recently demonstrated that the chemotactic peptide N-formyl-methionyl-leucyl-phenylalanine (FMLP) is capable of inducing a time-dependent downregulation of both FcgammaRIIIB and FcgammaRII in human neutrophils, altering FcgammaR-dependent functions.

N-formyl-methionyl-leucyl-phenylalanine : FcgammaRIIIB

→ INDIRECT DOWNREGULATOR

FMLP : : FcgammaRIIIB → INDIRECT DOWNREGULATOR

The c-fos messenger RNA expression induced by acute 6-hydroxydopamine injection was abolished by intraperitoneal pretreatment with the dopamine D2 receptor agonist, quinelorane (2 mg/kg) and strongly reduced by administration of the selective adenosine A2A receptor antagonist SCH-58261 (5 mg/kg).

6-hydroxydopamine : c-fos → INDIRECT UPREGULATOR

quinelorane : c-fos → INDIRECT DOWNREGULATOR

SCH-58261 : c-fos → INDIRECT DOWNREGULATOR

The thiazolidinediones, troglitazone and pioglitazone, decreased basal and TNF-alpha-stimulated PAI-1 secretion and mRNA expression in HUVEC in a dose-dependent fashion.

Thiazolidinediones : PAI-1 → INDIRECT DOWNREGULATOR

troglitazone : PAI-1 → INDIRECT DOWNREGULATOR

pioglitazone : PAI-1 → INDIRECT DOWNREGULATOR

Dexamethasone-mediated inhibition of Ox-LDL-induced GM-CSF mRNA expression and macrophage growth was significantly abrogated by RU-486, a glucocorticoid receptor antagonist

Dexamethasone : GM-CSF → INDIRECT DOWNREGULATOR

RU-486 : GM-CSF → INDIRECT UPREGULATOR

The results reported here show, by using a novel methodological approach, that an acute decrease of dopamine release causes an induction of c-fos messenger RNA in dopamine D2 receptor-containing striatopallidal neurons

dopamine : c-fos → INDIRECT DOWNREGULATOR

Under the drug screening process of synthetic diphenylpyrazole derivatives, we discovered compound yuwen02f1 possesses anti-inflammatory effects in decreasing the release of pro-inflammatory cytokines including TNFα and IL-6, nitric oxide, reactive oxygen species (ROS) as well as inhibiting migration of LPS-stimulated phagocytes

yuwen02f1 : *cytokines* → INDIRECT DOWNREGULATOR

yuwen02f1 : TNFfα → INDIRECT DOWNREGULATOR

yuwen02f1 : IL-6 → INDIRECT DOWNREGULATOR

Class 5. NOT. Negative CHEMPROT Interactions

Together with positive evidence of a chemical-protein interaction, sentences might state whether a chemical-protein interaction does not occur. This class should be used to define the NEGATIVE occurrence of a chemical-protein interaction, without providing any further information on the specific negative CHEMPROT class or subclass.

The results show that chronic treatment of cortical neurons with tyrosine kinase inhibitor (genistein) resulted in downregulation of the NR2B subunit polypeptide levels, while daidzein, an inactive analog of genistein, did not alter the levels of NR2B subunit, implying that tyrosine kinases may be involved in the regulation of the NMDA NR2B subunit content.

Daidzein : NR2B subunit → NOT

Explicitly stated that has no effect on GPRO levels (REGULATOR).

Daidzein : tyrosine kinase → NOT

Note: implicitly stated that it is not an inhibitor by the terms “inactive analog of genistein” which is an active compound.

Whereas growth hormone-releasing hormone (GHRH), gonadotropin-releasing hormone (GnRH), thyrotropin-releasing hormone (TRH), neuropeptide Y (NPY) and cholecystokinin

(CCK) failed to induce any change in GH release, corticotropin-releasing hormone (CRH) dose-dependently stimulated GH release with a significant effect at 1 nM and a maximal effect (> or =400% of controls at 24 h) at 100 nM.

thyrotropin-releasing hormone : GH → NOT

Note, according to the CEM guidelines, only peptides having less than 15 amino acids should be annotated. Therefore, only GnRH (10 aa) and TRH (3 aa) should be annotated as CEM (and also as GPRO). The rest of these peptides would be annotated exclusively as GPRO. In positive, this would be of class REGULATOR.

Example from PubMed ID: 16336943

In contrast, while lacking agonist activity at the VNI and VSV, olanzapine showed inverse agonism at the INI isoform (IC50 594 nM), reaching a maximal attenuation of 20%.

olanzapine : VNI → NOT

olanzapine : VSV → NOT

Example from PubMed ID: 10047461

The changes in cell cycle regulatory proteins associated with growth inhibition and DNA damage by Tomudex are not p53 dependent.

Tomudex : p53 → NOT

However, general indications of selectivity that do not provide specific negative associations should not be mapped:

Example from PubMed ID: 10027835

In contrast to esuprone and L-deprenyl, the selective MAO-B inhibitor LU 53439 was not effective in the kindling model; this substantiates the previous notion that the anticonvulsant activity of L-deprenyl is not related to MAO-B inhibition, but to other effects of this drug, such as inhibition of MAO-A.

LU 53439 : MAO-A → NOT LABEL, as the mention of LU 53439 being selective is very general and cannot be fully confirmed.

3. Interactions that should not be annotated. Negative Rules (N-rules)

N1. Do not label phenotypic responses, general mode of actions, biological roles, physiological roles, pharmacological roles and chemical roles of the kind:

cytotoxicity, apoptosis, differentiation, [(14)C]agmatine accumulation, intercalant, transepithelial Na+ accumulation, antiprogestenic activity, ischemia-induced hyperactivity, locomotor activity, antinociception, motor dysfunction, [3H]thymidine incorporation, behavioral effect, HPA-axis activation, genotoxic stress, neurite shortening, molecular messenger, hormone, neurotransmitter, antibiotic, antimicrobial, antiviral agent, fungicide, antifungal, antibacterial, antiprotozoal, disinfectant, xenobiotic, growth regulator, immunomodulator, metabolite, vaccine, aptamer, vitamin, nutrient, food, carcinogenic agent, allergenic agent, teratogenic agent, hepatotoxic agent, nephrotoxic agent, neurotoxin, cardiotoxic agent, toxin and related (neurotoxin), analgesic, antihistaminic, catalyst, solvent, base, acid, antioxidant, reducing agent, adduct, antigen, hormonal stress response, cell proliferation, cell migration,…

N2. Do not label relationships between general CEM terms/concepts and GPROs:

According to the CEM guidelines, there are general chemical mentions annotated (general class names, functional groups) that should be in general not annotated. An exception for this rule is for some functionalities (e.g. aromatic) that are structurally parts of a GPRO (class PART_OF).

Examples:

Our study examined recent claims of an association of the TaqI A1 allele and the functional -141C Ins allele of the dopamine D2 receptor (DRD2) gene with alcohol dependence.

Alcohol : DRD2 → NOT LABEL

The MWFE polypeptide of mammalian complex I (the proton-translocating NADH-quinone oxidoreductase) is 70 amino acids long, and it is predicted to be a membrane protein.

amino acids : complex I → NOT LABEL

Other examples of general concepts:

These findings demonstrate that NR2B antagonists may have clinical utility for the treatment of neuropathic and other pain conditions in man with a reduced side-effect profile than existing NMDA antagonists.

Here, a specific CEM is not found and no relationship should be tagged.

The central cannabinoid receptor (CB1) mediates the pharmacological activities of cannabis, the endogenous agonist anandamide and several synthetic agonists.

cannabis:cannabinoid receptor→ DIRECT REGULATOR

anandamide:cannabinoid receptor → AGONIST

synthetic agonists:cannabinoid receptor → nothing to be annotated

For the case of fragments or structural parts of compounds that are modified / examined and whose activity / binding against a GPRO is described / assessed, label the CHEMPROT relationship (if proceeds), unless the CEM corresponds to single chemical elements that are replaced.

Example from PubMed ID: 23562060

An HTS campaign identified several weak M5 PAMs (M5 EC50 >10μM) with a structurally related isatin core that possessed a southern phenethyl ether linkage.

isatin : M5 → MODULATOR-ACTIVATOR

phenethyl ether : M5 → MODULATOR-ACTIVATOR

N3. Do not label relationships/interactions between CEMs

In the present study, we investigated the effects of TCDD on testosterone signal transduction pathways and vice versa in the androgen receptor (AR) positive LNCaP prostate cancer cell line.

TCDD: testosterone → NOT LABEL

A selective 5-HT1B/1D agonist sumatriptan (0.01-10 mumol/L) caused nitrite production

sumatriptan : nitrite → NOT LABEL

The effect of substance P and compound 48/80 on histamine and serotonin release from not isolated and isolated mast cells have been compared in experiments in vitro.

substance P : histamine → NOT LABEL

N4. Do not label interactions/relationships between GPROs

The phosphorylation of the glutamate receptor 1 (GluR1) subunit of AMPA receptors by protein kinase A (PKA), protein kinase C (PKC), and Ca2+/calmodulin-dependent protein kinase II (CaMKII) has been characterized extensively.

AMPA receptors : PKA → NOT LABEL

N5. Do not label interactions that address the question "What a GPRO does to a CEM" or "CEM response affected by GPRO" (direction GPRO → CEM).

HEK293 cells overexpressing rOctn1, rOctn2, human OCTgN1, and human OCTN2 showed increased uptake and cytotoxicity of oxaliplatin compared with mock-transfected HEK293 controls; in addition, both uptake and cytotoxicity were inhibited by ergothioneine and L-carnitine.

rOctn1 : oxaliplatin → NOT LABEL (what a GPRO does to a CEM)

ergothioneine : oxaliplatin → NOT LABEL

A selective 5-HT1B/1D agonist sumatriptan (0.01-10 mumol/L) caused nitrite production

5-HT1B/1D : nitrite → NOT LABEL

Only pergolide had high potency and intrinsic activity at the dopamine D1 receptor for stimulating cyclic AMP accumulation.

Dopamine D1 receptor : cyclic AMP → NOT LABEL

Fyn tyrosine kinase reduces the ethanol inhibition of recombinant NR1/NR2A but not NR1/NR2B NMDA receptors expressed in HEK 293 cells

ethanol : NR1/NR2A → INHIBITOR

ethanol : NR1/NR2B NMDA receptors → NOT

Fyn tyrosine kinase : ethanol → NOT LABEL

Stabilized transfection of Tsup-1 cells with a combination of plasmids encoding Edg-2 plus -4 antisense mRNA suppressed the levels of Edg-2 and -4, but not Edg-3 and -5, in Western blots and reduced in parallel the increments in HB-EGF and susceptibility to DT evoked by LPA but not S1P.

Edg-2 plus -4 : LPA → NOT LABEL

Corticotropin-releasing hormone (CRH) is a central regulator of the hormonal stress response, causing stimulation of corticotropin and glucocorticoid secretion.

Corticotropin-releasing hormone : glucocorticoid → NOT LABEL

N6. Unmatched, multiple relationships.

a) Do not label cases in which several CEMs are mentioned together with a list of GPROs and it is not possible to establish which CEM associates with a given GPRO:

Corticosterone, desipramine, O-methylisoprenaline, cirazoline, moxonidine, l-arginine, l-lysine, verapamil, nifedipine, CdCl(2), ondansetron, and l-carnitine failed to inhibit specific [(14)C]agmatine accumulation, thus excluding that it is mediated by amino acid or monoamine carriers, by the putrescine carrier, by 5-HT(3) receptor channels, by Ca(2+) channels or by the organic cation transporters OCT1, OCT2, OCT3, OCTN1, or OCTN2

It could be inferred that these CEMS are DIRECT REGULATORS of these GPROs. However, there is not information on which CEM specifically relates to which GPRO. Therefore, do not label these relationships.

b) Do not label cases in which GPROs are grouped in a MULTIPLE class in such a way that the CEM effect cannot be fully mapped to all the GPROs:

Using radioligand binding techniques, we determined the equilibrium dissociation constants (K(D)) for 37 neuroleptics and one metabolite of a neuroleptic (haloperidol metabolite) for the human serotonin, norepinephrine, and dopamine transporters with [3H]imipramine, [3H]nisoxetine, and [3H]WIN35428, respectively.

As human serotonin, norepinephrine, and dopamine transporters is a single unique annotation (MULTIPLE class), the effect of each CEM ([3H]imipramine, [3H]nisoxetine, [3H]WIN35428) on each GPRO, despite being indicated in the sentence ("respectively"), cannot be assigned to a single GPRO.

[3H]imipramine : human serotonin, norepinephrine, and dopamine transporters

→ NOT LABEL

N7. CEMs that are part of a GPRO mention.

According to the CEM and GPRO guidelines, protein names that contain either the SUBSTRATE or the PRODUCT of the enzymatic reaction or transporter, have double

annotation for CEMs and GPROs. For example, the term "Glutamate Receptor" will be labeled as GPRO and also the sub-term "Glutamate" as CEM.

This kind of relationships (either for SUBSTRATE or PRODUCT) must not be annotated.

The phosphorylation of the glutamate receptor 1 (GluR1) subunit of AMPA receptors by protein kinase A (PKA), protein kinase C (PKC), and Ca2+/calmodulin-dependent protein kinase II (CaMKII) has been characterized extensively.

glutamate : glutamate receptor → NOT LABEL

AMPA : AMPA receptors → NOT LABEL

Thymidylate synthase (TS) is a key enzyme in the de novo synthesis of 2'-deoxythymidine-5'-monophosphate (dTMP) from 2'-deoxyuridine-5'-monophosphate (dUMP), for which 5,10-methylene-tetrahydrofolate (CH2-THF) is the methyl donor.

Thymidylate : Thymidylate Synthase → NOT LABEL

Thymidylate : TS → NOT LABEL

Clinical Implications of Dihydropyrimidine Dehydrogenase (DPD) Deficiency in Patients with Severe 5-Fluorouracil-associated Toxicity: Identification of New Mutations in the DPD Gene.

Dihydropyrimidine : Dihydropyrimidine Dehydrogenase → NOT LABEL

Dihydropyrimidine : DPD → NOT LABEL

However, if there is an independent, separate mention of the same CEM in the same sentence, where it is explicitly mentioned that this CEM acts as a substrate of the GPRO, then that mention (at the given position, separate from the co-mention of the CEM as part of the GPRO mention) should be explicitly labeled:

It has been postulated that cocaine's modulation of serum progesterone levels may in turn alter progesterone receptor activity, thereby contributing to cocaine-induced alterations of neuronal functions and genomic regulations.

progesterone : progesterone receptor → SUBSTRATE

But not label if there is not an explicit mention relating both substrate (CEM) and GPRO:

Progesterone serum levels, progesterone receptor (PR) protein levels, and PR-DNA binding complexes were measured in the striatum by radioimmunoassay, Western blot, and gel shift analyses, respectively.

progesterone : progesterone receptor → NOT LABEL

4. REFERENCES

(1) Wishart, D. S.; Knox, C.; Guo, A. C.; Shrivastava, S.; Hassanali, M.; Stothard, P.; Chang, Z.; Woolsey, J. DrugBank: A Comprehensive Resource for in Silico Drug Discovery and Exploration. *Nucleic Acids Res.* **2006**, *34* (Database issue), D668–D672.

(2) DrugBank Drug & Drug Target Database. http://www.drugbank.ca (accessed Oct 20, 2016).

(3) Yang, H.; Qin, C.; Li, Y. H.; Tao, L.; Zhou, J.; Yu, C. Y.; Xu, F.; Chen, Z.; Zhu, F.; Chen, Y. Z. Therapeutic Target Database Update 2016: Enriched Resource for Bench to Clinical Drug Target and Targeted Pathway Information. *Nucleic Acids Res.* **2016**, *44* (D1), D1069-74.

(4) Therapeutic Targets Database (TTD). http://bidd.nus.edu.sg/group/cjttd/ (accessed Mar 30, 2017).

(5) Bento, A. P.; Gaulton, A.; Hersey, A.; Bellis, L. J.; Chambers, J.; Davies, M.; Kruger, F. A.; Light, Y.; Mak, L.; McGlinchey, S.; Nowotka, M.; Papadatos, G.; Santos, R.; Overington, J. P. The ChEMBL Bioactivity Database: An Update. *Nucleic Acids Res.* **2014**, *42* (Database issue), D1083-1090.

(6) ChEMBL. https://www.ebi.ac.uk/chembl (accessed Oct 20, 2016).

(7) Visser, U.; Abeyruwan, S.; Vempati, U.; Smith, R. P.; Lemmon, V.; Schürer, S. C. BioAssay Ontology (BAO): A Semantic Description of Bioassays and High-Throughput Screening Results. *BMC Bioinf.* **2011**, *12* (1), 257.

(8) BioAssay Ontology. http://bioassayontology.org/ (accessed Mar 30, 2017).

(9) BEL Relationships. https://wiki.openbel.org/display/BLD/BEL+Relationships (accessed Mar 30, 2017).

(10) BEL Causal Relationships. https://wiki.openbel.org/display/BLD/Causal+Relationships (accessed Mar 30, 2017).

(11) Leitner, F.; Krallinger, M.; Vazquez, M.; Kuiper, M.; Valencia, A.; Lægreid, A. Mining Transcription Regulation Information from Literature. *Manuscr. Prep.*

(12) Perfetto, L.; Briganti, L.; Calderone, A.; Perpetuini, A. C.; Iannuccelli, M.; Langone, F.; Licata, L.; Marinkovic, M.; Mattioni, A.; Pavlidou, T.; Peluso, D.; Petrilli, L. L.; Pirrò, S.; Posca, D.; Santonico, E.; Silvestri, A.; Spada, F.; Castagnoli, L.; Cesareni, G. SIGNOR: A Database of Causal Relationships between Biological Entities. *Nucleic Acids Res.* **2016**, *44* (D1), D548-54.

(13) Signor 2.0. The SIGnaling Network Open Resource. http://signor.uniroma2.it/ (accessed Mar 30, 2017).

(14) Southan, C.; Sharman, J. L.; Benson, H. E.; Faccenda, E.; Pawson, A. J.; Alexander, S. P. H.; Buneman, O. P.; Davenport, A. P.; McGrath, J. C.; Peters, J. A.; Spedding, M.; Catterall, W. A.; Fabbro, D.; Davies, J. A.; NC-IUPHAR. The IUPHAR/BPS Guide to PHARMACOLOGY in 2016: Towards Curated Quantitative Interactions between 1300 Protein Targets and 6000 Ligands. *Nucleic Acids Res.* **2016**, *44* (Database issue), D1054-68.

(15) IUPHAR/BPS Guide to Pharmacology. http://www.guidetopharmacology.org/helpPage.jsp#ligand (accessed Mar 30, 2017).

5. BRAT ANNOTATION

CHEMPROT annotation is carried out with BRAT for all sentences in an abstract. CEMs and GPROs are red- and blue colored, respectively (Figure 1).

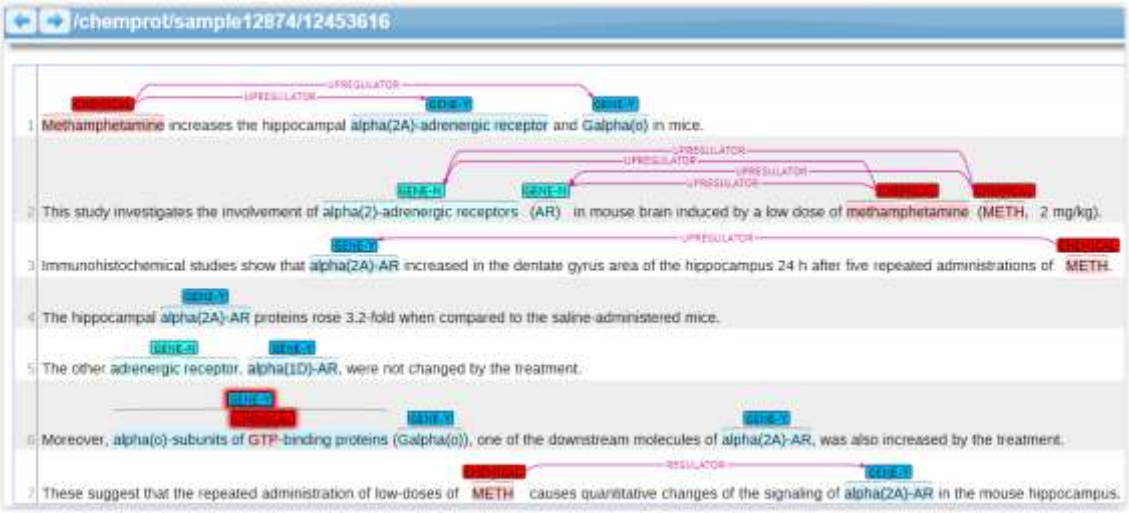

Figure 1. BRAT Annotation interface.

All mentions of a CEM and a given GPRO in an abstract should be labeled. For example, for the sentence "We found that although inactivation facilitated cisapride block of the HERG K+ current, it was not coupled with cisapride block of HERG when the Cs+ current was recorded", the two mentions of "cisapride" should be mapped to the two GPRO mentions of "HERG" (Figure 2).

Figure 2. Example of annotation of CHEMPROT for all occurrences of a given CEM-GPRO in a sentence.

This interface supports multiple CHEMPROT annotation for a given CEM – GPRO pair, by clicking on the selected CHEMPROTs (Figure 3):

Figure 3. Multiple class annotation with BRAF.

BRAT also supports feature annotation by typing selected terms within the Notes text box (Figure 4).

Figure 4. Introduction of feature terms (as indicated in rule G7) within the Notes field.