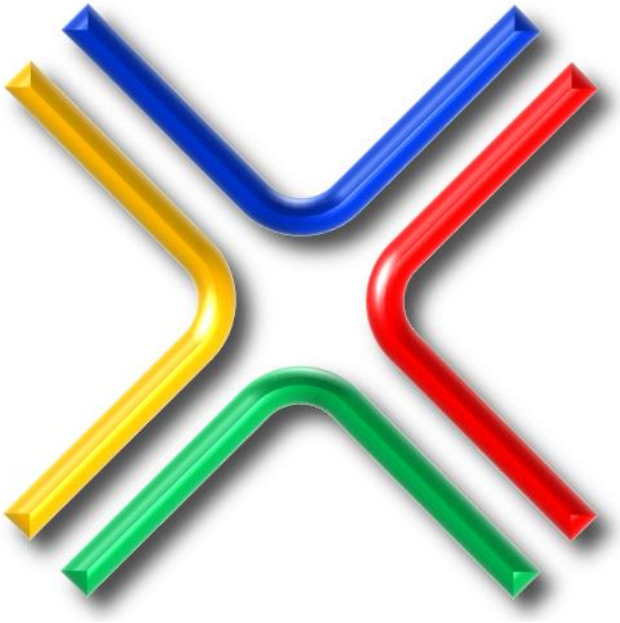
TRANSPATH®

flat files



BIOBASE

BIOLOGICAL DATABASES

Documentation

2012.2

geneXplain GmbH
Am Exer 10b
D-38302 Wolfenbüttel
Germany

E-mail: info@genexplain.com

URL: http://www.genexplain.com

Content

1. Overview

1.1. Preface

TRANSPATH® is an information system on gene-regulatory pathways. It focuses on pathways involved in the regulation of transcription factors. Elements of the relevant signal transduction pathways such as hormones, enzymes, complexes and transcription factors are stored together with information about their interaction.

TRANSPATH started in 1998 as a Ph.D.-project and was developed and curated by Frank G. Schacherer in the GBF Braunschweig. The TRANSPATH project was financed by a grant from the German Human Genome Project. The public version was developed using an object-oriented database management system and offered already three gateways to the data: a search page, a browsable index, and dynamically constructed graphical trees of interacting signaling molecules. There were also graphic maps of selected pathways, which allowed access to the data by clicking on images of the agents involved in the pathway.

As an exchange format for the data, XML (eXtensible Markup Language) was chosen. XML flatfiles can be created from and read into the database. In this format, the data can be read by humans but is also easily processable.

The aim of the early academic version as well as of the TRANSPATH® Professional database is

- to guide through a meanwhile overwhelming amount of data in a field which is connected to nearly all areas of modern molecular biology;
- to map the regulatory pathways for the individual molecules, and, ultimately, in the proteome(s) as a whole;
- to develop a tool for the identification of regulatory elements for newly unravelled genomic sequences;
- to provide a basis for a more comprehensive understanding of how the proteome governs transcriptional control.

The TRANSPATH data has been generally extracted from the original literature. Thus, the main responsibility for the correctness of the data is up to the authors, while we have to assume any responsibility for correctly extracting their publications. In the long term, a direct submission system should be envisaged similar to that established for depositing sequence data in the sequence data libraries.

1.2. TRANSPATH tables and their relations

This ASCII flat file release comprises the following tables:

1.2.1. MOLECULE

This table gives information on individual regulatory molecules, their states and complexes. Molecules interact with each other to build pathways. A Molecule in

TRANSPATH® is anything that is subject to reactions. Most Molecules have a mass, be it a small molecule like ATP, or a protein. No difference is made between receptors, enzymes, second messengers, transcription factors or other special kinds of proteins. A Molecule can also be a group of such entities, like a protein family, a state of such an entity, like the phosphorylated form or a complex of several other Molecules. And finally a Molecule can be part of another Molecule, either non-covalently bound as in a complex, or covalently bound as result of ubiquitination or sumoylation reactions. The reason for such a wide scope for this class is to catch anything that does show specific signaling behavior.

1.2.2. REACTION

This table gives information about interaction, reactions and relations between molecules. A reaction is a term for all kinds of interactions between signaling entities in signaling or regulatory events. The character of the interaction is more closely defined in the effect field by a set of terms. Reactions as processes are not physical entities like molecules, yet they are the central point in a signal transduction database. By representing these reactions between molecules as separate nodes in the graph, it becomes possible to store their properties and annotate them. Since many reactions in signal transduction are catalyzed, and most catalyzed reactions are quasi-unidirectional, all reactions stored in the database are by default unidirectional. Equilibrium reactions are identified in the effect field.

1.2.3. GENE

Gene information in TRANSPATH® was formerly provided in the molecule table. To have a clear separation between genes and gene products, they are now stored in a separate table. All genes are linked to the TRANSFAC® Gene table, where users can find information about the structure of gene regulatory regions, including individual binding sites for transcription factors, and other information as well. Genes have been included to provide information about the last step in signal transduction pathways: the regulation of target genes by activated transcription factors. Thus, TRANSPATH® presents information about complete signaling pathways: starting with the activation of a receptor at the membrane through a cascade of kinases into the nucleus, where a particular transcription factor is activated and regulates a set of target genes.

1.2.4. PATHWAY

The pathway flat file contains two types of entities: pathways and chains. Pathways reflect canonical pathways for specific signaling molecules (mostly ligands or receptors) and are made up of one or more chains. Chains are sets of consequent reactions, i.e. the reactions contain common molecules. Chains that are linked to pathways consist of reactions of the type 'pathway step' and in some cases 'semantic'. Evidence chains are derived from the information of one reference with a consistent experimental configuration. Chains can have bifurcations and even loops (if they carry a regulatory meaning).

1.3. How to cite TRANSPATH®

The TRANSPATH® database has been described in the following article. Please cite us if you apply TRANSPATH® for your research:

Krull, M.; Pistor, S.; Voss, N.; Kel, A.; Reuter, I.; Kronenberg, D.; Michael, H.; Schwarzer, K.; Potapov, A.; Choi, C.; Kel-Margoulis, O. and Wingender, E. (2006) "TRANSPATH®: An Information Resource for Storing and Visualizing Signaling Pathways and their Pathological Aberrations" Nucleic Acids Res. 34, D546-D551

1.4. PUBLICATIONS

1. Kel, A.; Voss, N.; Jauregui, R.; Kel-Margoulis, O. and Wingender, E. (2006) "Beyond microarrays: Find key transcription factors controlling signal transduction pathways" BMC Bioinformatics, 7 (Suppl. 2), S13.
2. Krull, M.; Pistor, S.; Voss, N.; Kel, A.; Reuter, I.; Kronenberg, D.; Michael, H.; Schwarzer, K.; Potapov, A.; Choi, C.; Kel-Margoulis, O. and Wingender, E. (2006) "TRANSPATH®: An Information Resource for Storing and Visualizing Signaling Pathways and their Pathological Aberrations" Nucleic Acids Res. 34, D546-D551
3. Kel-Margoulis, O.; Matys, V.; Choi, C.; Reuter, I.; Krull, M.; Potapov, A. P.; Voss, N.; Liebich, I.; Kel, A.; Wingender, E. (2005) "Databases on Gene Regulation" In: "Information Processing And Living Systems", eds.: Bajic, V. B. and Wee, T. T., World Scientific Publishing Co.
4. Choi, C., Crass, T., Kel, A., Kel-Margoulis, O., Krull, M., Pistor, S., Potapov, A., Voss, N., Wingender, E. (2004) "Consistent re-modeling of signaling pathways and its implementation in the TRANSPATH database" Genome Inf. Ser. 15, 244-254.
5. Choi, C.; Krull, M.; Kel, A.; Kel-Margoulis, O.; Pistor, S.; Potapov, A.; Voss, N.; Wingender, E. (2004) "TRANSPATH®: a high quality database focused on signal transduction" Comp. Funct. Genom. 5, 163-168.
6. Krull, M.; Voss, N.; Choi, C.; Pistor, S.; Potapov, A.; Wingender, E. (2003) "TRANSPATH®: an integrated database on signal transduction and a tool for array analysis" Nucleic Acids Res. 31, 97-100.
7. Kloos, D. U.; Choi, C.; Wingender, E. (2002) "The TGF-beta-Smad network: introducing bioinformatic tools" Trends Genet. 18, 96-103.
8. Schacherer, F.; Choi, C.; Götze, U.; Krull, M.; Pistor, S.; Wingender, E. (2001) "The TRANSPATH signal transduction database: a knowledge base on signal transduction networks" Bioinformatics 17, 1053-1057.
9. Potapov, A. and Wingender, E. (2001) "Modeling the architecture of regulatory networks." In Wingender, E.; Hofestaedt, R.; Liebich, I. (eds.), Proceedings of the German Conference on Bioinformatics (GCB '01), GBF Braunschweig, 6-10.
10. Wingender, E.; Chen, X.; Fricke, E.; Geffers, R.; Hehl, R.; Liebich, I.; Krull, M.; Matys, V.; Michael, H.; Ohnhäuser, R.; Prüß, M.; Schacherer, F.; Thiele, S.; Urbach, S. (2001) "The TRANSFAC system on gene expression regulation" Nucleic Acids Res. 29, 281-283.

11. Wingender, E.; Chen, X.; Hehl, R.; Karas, H.; Liebich, I.; Matys, V.; Meinhardt, T.; Prüß, M.; Reuter, I.; Schacherer, F. (2000) "TRANSFAC: an integrated system for gene expression regulation" *Nucleic Acids Res.* 28, 316-319.

12. Heinemeyer, T.; Chen, X.; Karas, H.; Kel, A. E.; Kel, O. V.; Liebich, I.; Meinhardt, T.; Reuter, I.; Schacherer, F.; Wingender, E. (1999) "Expanding the TRANSFAC database towards an expert system of regulatory molecular mechanisms" Nucleic Acids Res. 27, 318-22.

2. Table MOLECULE

Molecules interact with each other to build pathways. A molecule is anything that is subject to reactions. Most molecules have a mass, be it a small molecule like ATP, or a protein. No difference is made between receptors, enzymes, second messengers, transcription factors or other special kinds of proteins. A molecule can also be a group of such entities, like a protein family, a state of such an entity, like the phosphorylated form or a complex of several other molecules. And finally a molecule can be part of another molecule, either non-covalently bound as in a complex, or covalently bound such as ubiquitin-like proteins. The reason for such a wide scope for this class is to catch anything that shows specific signaling behavior.

A molecule can serve in four fundamental roles in the context of a reaction: it can be an educt, a product, an enzyme or a modulator. Enzyme, educt and modulator are the inputs of the reactions, while products are the outputs. For semantic reactions molecules are only grouped into signal donors, which are the inputs, and signal acceptors, which are the outputs.

We use the hierarchy forming categories orthofamily/family, orthogroup/isogroup and orthobasic/basic to refer to relations between different entries for proteinaceous molecules. To keep the states/modified forms and complexes separated from the family grouping hierarchy, we use the categories modified forms and complexes.

Further categories comprise non-proteinaceous molecules. Ions, metabolites, co-enzymes, secondary messengers, etc. are collectively addressed as small molecules. On the other hand e.g. miRNA's create different categories, miRNA basic and pre-miRNA basic, depending on their processing status.

2.1. Orthofamily/Family

Important attributes of all components are the collections that allow expression of family relationships. Each component can act as a group of other components or as a member in such a group. In the case of molecule, this grouping represents family relationships between the molecules. In the case of reaction, this grouping can be used to assign reactions to a general class of reactions, as it is done successfully in the EC nomenclature. The third possibility is the use of mixed groups, which contain both reactions and molecules, and correspond to nets or explicitly stated pathways. This is interesting since it would support the appealing idea of building networks from subnets in a modular fashion.

Grouping molecules into families is essential for a usable signal transduction database. When the family relationships are stated in a formal way, it is easy to write algorithms that exploit them in user queries, while marking inherited properties as inferred information. The prefix "ortho" is used when the family entry is not specific for a certain species or higher taxon (Fig. 1).

The family relation is implemented by the groups/members collections.

2.2. Orthogroup/Isogroup

For a single gene, different isoforms such as splice variants may exist. Sometimes in the literature a signaling activity is first attributed to a single molecule, and later it is discovered that there is a whole group of similar molecules. Therefore, a special type of molecule entry is needed, which we term "isogroup" for taxon-specific entries and "orthogroup" for orthologous, non-species-specific ones. To these abstracted group entries all the information can be assigned, where it is not known which specific isoform is involved.

2.3. Orthobasic/Basic

Molecules of the type "basic" contain data for a specific isoform, e.g. splice variant, to which an amino acid sequence can be assigned. Again, the prefix "ortho" is used to generalize information for orthologous isoforms from different species.

A modification tag '_mod' is added to the type terms listed above if the entry carries a modification such as phosphorylation. Complexes and modified forms change availability of the molecule for reactions.

2.4. Classification methods

Since it is easier to examine the sequence of a gene or protein than to investigate its function, a certain behavior is usually shown in detail for only few members of a protein family, and homologs are added to the functional group by sequence or structural similarity.

Based on this assumption, various good databases exist that try to classify proteins and map sequence motifs to functional annotation. They cluster proteins by multiple sequence alignments and use common structural motifs, "profile" patterns, or Hidden Markov Models derived from these alignments to classify new proteins. Sometimes these methods can correctly predict the function, but sometimes not. Thus it is common practice to group molecules into families on the basis of sequence similarities, even if they do not share common behavior.

For a signaling database, it is advantageous to group molecules that show common signaling behavior, that is, to group them by function. Since it is the function we are interested in, we would like to group only by function. On the other hand, we would like to draw as much as possible upon expert knowledge, and stay coherent with the groupings that already exist. To solve this dilemma, the best option seems to be that we group the molecules as it is done traditionally, but link signaling only to those molecules for which it has been shown.

Given the hierarchical tree, if the reaction has been demonstrated for A(h), we link that reaction to A(h) only. We link statements made on a generalized level, like the ones from review papers, to nodes on a higher level in the molecule hierarchy. Given the tree we can link the general activation of A-like proteins to the orthofamily entry "A-like". The context of the original literature statement is preserved.

2.5. States (modified forms)

The unmodified form of a protein and all its modified forms are its states, where the modification can be by covalent binding, by complexation, or by change of environment. The protein per se is a concept that is based on the observation that there is only one gene coding for each protein sequence. All the states share the same gene and consequently part of their structure, the amino acid chain. They are functionally related, often even reversibly transformable into each other.

A basic molecule entry captures this concept and is the class of all states of a protein. These states are different molecules, and we store them as different molecule entries. As molecules, they can be used in a pathway assembly. We store general information like the amino acid sequence in the basic molecule entry and link its states to it.

In the simplest case there are only two states- an inactive one, and an active one. In other cases, there are more. For example, a transcription factor can be

1. de-phosphorylated in the cytosol

2. phosphorylated in the cytosol or

3. phosphorylated and bound to DNA in the nucleus,

to name a few possibilities. The same protein will exhibit distinctly different signaling functions in these three states. For example in state 1 it will be susceptible to phosphorylation, in state 2 to translocation into the nucleus or dephosphorylation and in state 3 it will activate transcription.

The number of states for a molecule is the product of the number of its modified forms and the number of locations where it is found. Only compounds which share the same location interact in nature.

It is impractical to enter a separate state for each location. Most molecules can be found in several tissues, at several development stages, in several cellular compartments, several organs and several celltypes. To enter a state for each possible combination would lead to an explosion in the number of states, and redundancy in the reactions. This problem is circumvented by using a list of positive and negative locations that is linked to the basic molecule.

In each state, the molecule is available only for a subset of all reactions for that molecule. Receiving a signal changes the molecule's state, usually leading to a new state from which reactions are triggered.

Motifs are structural and functional features of a protein and are often responsible for its signaling behavior. A single protein can have several signaling motifs. Motifs are listed in the feature field together with position information and references.

More detailed information about the data model of TRANSPATH can be found at:

Choi, C., Crass, T., Kel, A., Kel-Margoulis, O., Krull, M., Pistor, S., Potapov, A., Voss, N., Wingender, E. (2004) "Consistent re-modeling of signaling pathways and its implementation in the TRANSPATH database" Genome Inf. Ser. 15, 244-254.

2.6. Molecule naming conventions

2.6.1. Complex binding

* :

e.g. A:B:C

2.6.2. Modification notation

* {mod}

modification; covalent binding

* {mod}n

multiple modification (n stands for any number), e.g. hyperphosphorylation; {mod}2 double modification, etc.

* {mod(n)}

quantifier for modifications at one position, e.g. polyubiquitination at a certain Lys residue:

{ub(4)K145}; but also to indicate monoubiquitination: {ub(1)}

* {modAAposition}{modAAposition}{...

indicates modifications and positions, e.g. {pS175} or {pY53}{pT99}

* there are no separate molecule entries for different states of modifications, unless they behave differently

2.6.3. Types of modification {mod}

* {ace} acetylated

* {ami} amidated

* {car} carboxylated

* {cit} citrulline (derived from methylated arginine)

* {chol} cholesterol modified

* {drib} ADP ribosylated

* {elon} conjugation with elongin B

* {far} farnesylated

* {ger} geranylgeranylated

* {glut} S-glutathionylation

* {gly} glycosylated

* {hyd} hydroxylated

* {met} methylated

* {myr} myristoylated

* {nedd} neddylated; conjugation with Nedd8

* {no} nitrosylated

* {no2} nitrated

* {ox} oxidized, e.g. cysteine residues that are linked via an S-bridge (disulfide-bridge) in the oxidized protein intramolecularly, which is disconnected in the reduced form

* {p} phosphorylated

* {pal} palmitoylated

* {pren} prenylated

* {re} reduced, e.g. {reCys123Cys125}

* {sol} soluble

* {sul} sulfated

* {sumo} sumoylated; conjugation with SUMO-1

* {sumo2} sumoylated; conjugation with SUMO-2

* {sumo3} sumoylated; conjugation with SUMO-3

* {ub} ubiquitinated

2.7. Location

Location describes spatial and temporal environments. It distinguishes molecules and reactions found in different kinds of cells, tissues and organs. Without it we would get the model of a super-cell, which can do everything. Imagine the network for each cell being drawn on an overhead-transparency. Then the view without location information would be all these transparencies layered on top of each other.

Signaling needs the location context. Subcellular location plays an important role in the activity of many compounds. For example, transcription factors like NF-kappaB are only active in the nucleus and Ras is only active after recruitment to the inner plasma membrane. Since all signaling takes place in the complex matrix that is the multicellular organism, we first need a good model of the spatio-temporal segmentation of the organisms.

Each component in the network has two lists of location objects. One list contains all locations where experiments have confirmed the presence of the component, the other list contains all locations where experiments have confirmed its absence - which is necessary to distinguish between cases where the component is absent and cases where its status has not been investigated. For molecules, the locations describe where the molecules have been shown to be expressed, or have been shown not to be expressed. For reactions, the locations describe in what kind of system the reactions have been confirmed, or have been found not to happen.

Relevant aspects of location include:

* subcellular location

Examples: in the nucleus, plasma-membrane associated or spanning, mitochondrial, in the cytosol

* tissue/cellular location

  Examples: liver, kidney, body, glomeruli, Schwann cells, adipose tissue

* developmental stage

  Example: Carnegie stages in human.

* species

  Examples: vertebrata, mammalia, human, rat, mouse

The location items often relate to each other in tree-like schemes. A human, for example, is a vertebrate in the taxonomic dimension. Glomeruli are a substructure of the kidney. This classification is necessary to avoid state explosion, otherwise we would have to insert a separate entry for each subtype, when the supertype is mentioned in the literature.

Sequences and molecular weights differ between orthologs. While the species is categorized as part of the location of a molecule, these orthologs are not different states of the same molecule, they are different members of a group of molecules. The molecules from different species are stored as members of their orthogroup entry.

The experimental proof for a reaction is usually obtained in artificial systems. We cannot claim that a reaction takes place in a cell in vivo simply because all involved molecules are found in this cell. We have to link the reaction to the location for which it was proven.

When we have further information that depends on the location, we store this information in the relation between Location and Molecule/Reaction.

Locations in TRANSPATH® can contain up to five different parameters that are listed below. These attributes are used separately or in combination.

species

The same nomenclature as in TRANSFAC® is used. Examples:

* human, Homo sapiens

* mouse, Mus musculus

* rat, Rattus norvegicus

tissue

Names of organs or tissues.

compartment

The intracellular location of the protein.

In many cases, this is very important for signaling. Since TRANSPATH® Professional 2.3, Gene Ontology terms (Cellular Component ontology) are used to describe subcellular locations of TRANSPATH® Molecules.

Associations of Molecules with Gene Ontology terms (Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium (2000) Nature Genet. 25: 25-29) were mainly obtained by electronic annotation using a gene association data list from Compugen Inc. (http://www.cgen.com/, http://www.labonweb.com).

stage

The developmental stage.

celltype

Names cell lines used in experiments or natural cell types. For cell lines hyperlinks to HyperCLDB (Cell Line Data Base) are provided.

2.8. Fields

It should be noted that in individual entries, some fields may be empty. In this case, these fields are not displayed.

Field Content and format

AC Accession number

The accession number is the unique identifier for each entry. Its format is "MO" in capital letters followed by nine digits (e.g. MO000012345). If two entries have to be merged , the AC of the primary name is retained. The other AC will be stored in the secondary accession numbers (AS) field.

AS Accession numbers, secondary

The secondary accession numbers are optional alternate identifiers for each entry. They are of the form defined separated by semicolons, and are created when two entries are merged.

DT Created by

The name of the curator who created the entry. E-mail your feedback directly to the curator.

Updated by

The name of the curator who last updated the entry. E-mail your feedback directly to the curator.

CO Copyright-information

NA Name

A human-readable identifier for the component. The most common spelling in the literature is used, with an emphasis on forms with a dash (e.g. Grb-2 instead of Grb2), if both forms exist. Non-abbreviated molecule names are written according to their most common appearance, in most cases beginning with a lower case letter. Note that Greek letters are expanded to alpha, beta, gamma etc.

Molecules (transcriptions factors) with a lot of detailed information available at TRANSFAC® have a hyperlink with the respective accession number in TRANSFAC in front of their name.

If a molecule is modified chemically and in a different state the name is tagged with an abbreviation in {} brackets.

For a list of abbreviations and their meanings, see Section 2.6 (nametag_loc.txt)

There are also tags to differentiate the species the molecule comes from. This short identifier is useful in reaction names, because the experimental evidences for the reactions are frequently based on molecules from different species. The tag (v.s.) for vertebrate species is used when the exact (vertebrate) species has not been described in the reference and could also not be investigated from cited references or websites.

Isoforms such as splice variants that have no standard or common name or abbreviation yet, are named using the isogroup name and a BIOBASE-specific extension (-xbb1, -xbb2, ...).

For a list of species tags, see Chapter 7.

Stoichiometric factors are used, if quantitative details are known. The notation for homodimers is A:A (or (A)2); in larger complexes the number of equal components is written behind the respective molecule, for example A:(B)2:C:(D)4.

SY Synonyms

Other names for the same component. This is needed, since the names in biosciences change often. The field lists other names or abbreviations. Specific names from orthologous species are not stored here. They are accessible via the group-member hierarchy. Different synonyms are separated by a semicolon.

GE Encoding gene

The corresponding gene encoding for this protein.

OS Species

The species that the molecule entry belongs to. Given is the common name (if it exists), followed by the Latin denomination, as it is done in TRANSFAC.

CL Classification

This sorts the entry into all the groups it belongs to. The classification stretches over all hierarchical levels, that are "above" the entry. As a molecule can belong to more than one functional group several classification paths can exist. Main paths are marked.

TY Type

The type of this molecule entry. Possible values are:

* orthofamily, group entry for homologous superfamilies, families
* family, for species-specific superfamilies, families
* orthogroup, group entry for the products of orthologous genes (equivalent for orthogene entries in gene table)
* isogroup, group entry for the products of one gene in one taxon (usually species), which have emerged from gene duplication, alternative splicing of mRNA etc.
* orthobasic, entry for othologous isoforms (e.g. splice variants)
* basic, for taxon-specific isoforms (usually species)
* orthocomplex, group entry for othologous complexes, consisting of non-covalently bound molecules
* complex, for taxon-specific complexes, consisting of non-covalently bound molecules A modification tag '_mod' is added to the type terms listed above if the entry carries a modification such as phosphorylation.
* group (XOR), is a grouping entry for molecules with the same function (e.g. isoforms) in a pathway; should be interpreted in a way that exactly one member (exclusive OR) of this group is involved in a concrete instance of the pathway
* miRNA basic, for mature miRNA (micro RNA) entries
* pre-miRNA basic, for pre-miRNA (precursor micro RNA) entries
* mRNA basic, for mRNA (messenger RNA) entries
* smallmolecule, for small non-proteinaceous compounds
* other, for entries that don't fit into one of the other type categories

HP Superfamilies

Lists all groups or families this component belongs directly to (one hierarchical level above). This is a very important field, since abstracting common signaling behaviour is needed to avoid the explosion of entries. For molecules, a group is a set of molecules which share a common signaling behaviour. For example, all isoforms of the gamma subunit of G-proteins could be grouped into one group. Also, orthologous forms from different species can be grouped together on hierarchical levels. On the bottom of the whole hierarchy, there should be proteins (and other molecules) which physically exist.

HC Subfamilies

Lists all molecule entries that are a step downward in the hierarchy, e.g. splice variants of a molecule or members of a group/family.

SZ Sequence length, molecular weight

The total number of amino acids in the given sequence. The calculated molecular weight of the protein in Dalton or Kilodalton (derived from cDNA /genomic clones). Experimental molecular mass (or range) in kDa (experimental method, e. g. SDS PAGE, GF/gel filtration).

IP Isoelectric point

The isoelectric point (pI) is the pH value at which a protein has not net electrical charge. Here, the calculated pI value for the protein is given. Experimental isoelectric point (experimental method), [reference].

SC Sequence source

Names the data source (e.g. database) the amino acid sequence has been derived from.

SQ Sequence

Shows the amino acid sequence in one-letter-code.

DR External database hyperlink

Database name (e. g. Swiss-Prot): database accession number; identifier.

The focus lies on linking TRANSPATH with Swiss-Prot, EMBL, InterPro, Entrez Gene, UniGene, GO, DIP, BIND, and HyperCLDB. For some of the molecules links to PDB, PROSITE, Flybase, MGD, and others are also provided. Also, corresponding Affymetrix micro-array probe set identifiers are listed. For the following chips data is available:

* AFFY_HG_FOCUS
* AFFY_HG_U133A
* AFFY_HG_U133A_2
* AFFY_HG_U133B
* AFFY_HG_U133_PLUS_2
* AFFY_HG_U95AV2
* AFFY_HG_U95B
* AFFY_HG_U95C
* AFFY_HG_U95D
* AFFY_HG_U95E
* AFFY_MG_U74AV2
* AFFY_MG_U74BV2
* AFFY_MG_U74CV2
* AFFY_MOUSE430_2
* AFFY_MOUSE430A_2
* AFFY_MU11KSUBA
* AFFY_MU11KSUBB
* AFFY_RAT230_2
* AFFY_RG_U34A
* AFFY_RG_U34B
* AFFY_RG_U34C
* AFFY_U133_X3P
* HuGeneFL

The format is AFFYMETRIX:chip:probeset. Except for those from chip HuGeneFL, the Affymetrix links are based on those in Ensembl, version 27.35a for human, 27.33c for mouse, and 27.3e for rat. External database hyperlink (of encoding gene) -- Hyperlinks of the encoding gene. They are included to retrieve matches in an array data analysis that is focused on gene products.

FT Features (motifs)

Lists all features/motifs/domains of the molecule (e.g. SH2-domain, signal peptide...), that are important in the diverse signaling cohesions the molecule is involved in. Gives the first and last position of the feature/motif in N -> C direction on the AA sequence. Structural and functional features/motifs, that are annotated from the literature, are named by the common motif name and have a reference link.

Automatically annotated features/domains that were calculated using Pfam hidden Markov models (HMM) are also shown in this field. These features are characterized by their Pfam model name, a link

to the corresponding Pfam entry, a raw score and an E-value (expectation value). For further explanations please take a look at the Report 1, 0001 (2003).

CC Comments

A list of comments. Further information about the different categories of comments is available in Chapter 6.

GO GO: biological process, molecular function

A list of Gene Ontology (GO) terms from the ontologies 'molecular function' and 'biological process'. Associated terms from the third ontology from GO -cellular component- can be found in the field 'Location positive and experiment(s)'. All links to GO terms have been retrieved from Ensembl.

CP Location positive and experiment(s)

Lists all locations (tissues, physiological cell types, cell lines, cellular compartments, developmental stages) where the compound was found. Gives experiment abbreviation, the signal strength in the experiment as expression level and cites reference (if available). For further information, please see Section 2.7 (location.txt).

Abbreviations for the types of experiments (methods) used to verify the abundance of the molecule are given in brackets. For an acronym explanation table please see Chapter 8 (quality.txt).

CN Location negative and experiment(s)

Lists all locations where the compound was NOT found. By doing this, having no entry is used to tell that nothing is known about the components location.

Abbreviations for the types of experiments (methods) used to verify the absence of the molecule are given in brackets.

ST Complex or modified form of

A list of molecules this molecule entry is a modified form of. Or, if it is a complex, its subunits.

CX Complexes

A list of complexes this molecule is engaged in.

MF Modified forms

A list of modified forms for this molecule. Modified forms have no own AA sequence, as it is given in the linked "Modified form of"-entry.

XA Reaction downstream

A list of reactions which go out from this molecule (in the semantic view). Here the molecule serves as a signal donor.

XB Reaction upstream

A list of reactions which lead to this molecule (in the semantic view). So the molecule serves as a signal acceptor.

XC Reaction catalyzed

A list of reactions which are catalyzed by this molecule.

XI Reaction inhibited

A list of reactions which are inhibited by this molecule.

PW Pathways

Indicates all the pathways and chains in which the respective molecule is involved.

RN Reference number

[consecutive entry reference number]. A list of the papers from which the information in this entry was extracted.

RX Pubmed database hyperlink

The PMID number in PubMed

RA Reference author(s)

List of authors

RT Reference title

RL Reference publication

3. Table REACTION

A reaction in TRANSPATH® is a term for all kinds of interactions between signaling entities (molecule or gene) in signaling or regulatory events. The character of the interaction is more closely defined in the effect field by a set of terms.

Reactions as processes are not physical entities like molecules, yet they are the central point in a signal transduction database. By representing these reactions between molecules as separate nodes in the graph, it becomes possible to store their properties and annotate them.

Since many reactions in signal transduction are catalyzed, and most catalyzed reactions are quasi-unidirectional, all reactions stored in the database are by default unidirectional. Equilibrium reactions are identified by the term "reversible" in the effect field.

Basically there four kinds of reactions in the database: molecular evidence, pathway step, decomposition, and semantic. Of these molecular evidence, pathway step, and semantic can have two specifications, direct or indirect. Metabolic reactions are curated on pathway step level only. They are assigned pathway step; metabolic. These terms are stored in the type field.

3.1. Reaction hierarchy

The reaction hierarchy has been introduced in release 5.1 to connect the patchwork of molecular evidence reactions (different species and modified forms in the various experiments) to a mechanistic level that models consecutive pathways. This pathway level consists of 'direct' pathway step reactions, 'indirect' pathway step reactions (in cases where the direct mechanism for a part of the pathway is not known) and auxiliary decomposition reactions. In addition there are 'metabolic' pathway step reactions. In combination with reaction chains, this level is used to model canonical pathways and networks. The semantic projection level gives a broader shortcut overview of the signal transduction pathways without biochemical detail.

3.2. Mechanistic reactions: molecular evidence, pathway step, decomposition

The classical chemical reaction notation can model aggregation into a complex, dissociation of acomplex, chemical modification and catalysis in the detail necessary to capture data from the primary literature.

We call this representation mechanistic, because it reflects how the molecules interact, and not how the signal is transported. It is well-established and familiar and is used for the pathway step, decomposition and molecular evidence reaction types.

Molecules act as enzymes, substrates, modulators, or products, and any reaction has exactly one or no enzyme, one or more substrates, any number of modulators, either inhibitors or activators, and one or more products. There is semantic information in assigning some of the inputs to the enzyme or modulator collections, which tells us something about the local meaning of the molecule for the reaction.

This view demands a detailed knowledge of the mechanisms of transduction. It makes no assumptions about the molecule states. All the different states of a molecule have to be listed as separate entities.

3.3. Semantic reactions

We call this representation semantic, because it assigns a meaning to the states of the molecules for the overall network, "active" or "inactive". It is easy to understand and it is familiar to scientists from the papers. It shows how the signal flows through the network. The reactions need additional information, that is if they activate or inhibit the target molecule.

The reactions in this scheme are binary. Molecules act as signal donors or signal acceptors to the reactions.

This view implicitly assumes that each molecule exists in an active and an inactive state. It is not necessary to differentiate between them, as it is understood that incoming activating reactions always refer to the inactive state, incoming inactivating reactions and outgoing reactions refer to the active state. Because the states/modified forms are implied for semantic reactions, we link them to basic molecule entries in the database.

Saying that a molecule has an "active" or "inactive" state is a semantic statement. Both states undergo reactions, and the decision as to which is which can consequently only be determined in the larger context of the whole network.

3.4. Translocations

In a translocation, the same molecule enters and leaves the reaction with a changed spatial context. This process takes some time, which is important for the dynamic behavior of the network.

Translocations cannot be represented in the basic mechanistic model which assumes all reaction partners are present in the same reaction space. Since a molecule is just associated with a list of locations, we cannot differentiate between, for example, the cytosolic and the nuclear form of the molecule. A reaction that moves a molecule from one form to the other is then a loopback from the molecule to itself. Therefore we associated the two locations with the connection entries between the molecule and the reaction. The term translocation is assigned in the effect field.

More detailed information about the data model of TRANSPATH can be found at:

Choi, C., Crass, T., Kel, A., Kel-Margoulis, O., Krull, M., Pistor, S., Potapov, A., Voss, N., Wingender, E. (2004) "Consistent re-modeling of signaling pathways and its implementation in the TRANSPATH database" Genome Inf. Ser. 15, 244-254.

3.5. Name syntax for reactions

1. Binding

Examples:
cyclin B1(h):Cdk1(h) + p73alpha(h) <==> cyclin B1(h):Cdk1(h):p73alpha(h)

(TGFbeta1(h))2 + 2 TGFbetaR-II(h) <==> (TGFbeta1(h))2:(TGFbetaR-II(h))2
ErbB2(r){pY1227} + CrkII(h) <==> ErbB2(r){pY1227}:CrkII(h)

2. Dissociation

Example:
ROS(m):SHP-1(h) <==> ROS(m) + SHP-1(h)

3. Phosphorylation

General view:

ProtA + ATP --EnzymeC--> ProtA{p} + ADP

Examples:

GATA-4(m) + ATP --ERK2(m)--> GATA-4(m){pS105} + ADP

stathmin(x) + 3ATP --Plk1(x)--> stathmin(x){pS16}{pS25}{pS39} + 3ADP
Here, three residues have been phosphorylated and thus there are 3ATP and 3ADP.

KIF23(h) + ATP --cyclin B1(h):Cdk1(h)--> KIF23(h){p} + ADP
The enzyme is in a complex.

cyclin A(h):Cdk2(h):Cdc25A(h) + ATP --> cyclin A(h):Cdk2(h):Cdc25A(h){p} + ADP
The enzyme is in a stable complex with the substrate and is therefore not separately linked as enzyme.

4. Dephosphorylation

General view:

ProtA{p} --EnzymeC--> ProtA + p

Examples:

Jak2(v.s.){pY1007} --SHP-2(v.s.)--> Jak2(v.s.) + p

C-Nap1(h){p} --PP1-gamma1(h):Nek2A(h)--> C-Nap1(h) + p
Here, an enzyme is in a complex.

Pyk2(r){pY402}{pY579}{pY580} --PTP-PEST(h)--> Pyk2(r) + 3p

Cdk2(h){pT160}:KAP(h) --> Cdk2(h):KAP(h) + p
Enzyme is in a stable complex with the substrate and is therefore not separately linked as enzyme.

5. Acetylation

General view:

ProtA + AcCoA --EnzymeA--> ProtA{ace} + CoA

Example:
RIP140(h) + AcCoA --p300(m)--> RIP140(h){aceK446} + CoA

6. Deacetylation

General view:

ProtA{ace} --EnzymeB--> ProtA + acetyl

7. Ubiquitination

Ubiquitin chains are formed by adding several 76aa ubiquitins to an already bound ubiquitin one after another. Degradation of ubiquitinated proteins by the proteasome do not degrade the ubiquitin. The 76aa ubiquitin fragments have been shown to be recycled.

Examples:

alpha-synuclein(m.s.) + ubiquitin(m.s.) --parkin(m.s.)--> alpha-synuclein(m.s.){ub}
Notation {ub} is applied when nothing is known about the ubiquitin chain.

alpha-synuclein(h){ub(2)} + n ubiquitin(v.s.) --UCH-L1(h)--> alpha-synuclein(h){ub(n)}
Notation {ub(2)} reflects the fact that a chain of two ubiquitins was formed.

p53(h) + n ubiquitin(v.s.) --PIRH2(v.s.)--> p53(h){ub(n)}

Mdm2(h) + n ubiquitin(v.s.) --> Mdm2(h){ub(n)}

Rad23A(mo) + n ubiquitin(v.s.) --E6-AP(h)--> Rad23A(mo){ub(n)}
Notation {ub(n)} is applied when authors detect polyubiquitination.

8. Deubiquitination

Removal of the ubiquitin modification. Besides ubiquitin withdrawal the protein is unchanged.

Examples:
AF-6(h){ub} --Fam(m)--> AF-6(h) + ubiquitin(h)
TRAF2(m.s.){ub} --CYLD(m.s.)--> TRAF2(m.s.) + ubiquitin(m.s.)

9. Degradation

Often, a degradation reaction for ubiquitinated proteins is shown.

Example:
TGFbetaR-I(v.s.){ub} --> protein remnants + ubiquitin(v.s.)

A degradation reaction will yield numerous break-down products of undefinable size which become smaller as degradation continues. Hence polyacrylamid gel electrophoresis (PAGE) or on Western blots etc. will show a smear if a degradation reaction has been followed. In case the protein in question was flagged for degradation by ubiquitin the ubiquitin-flag is recycled. Degradation of ubiquitinated proteins by proteasomes do not degrade the ubiquitin. The 76aa ubiquitin fragments have been shown to be recycled.

10. Sumoylation

Examples:

AP-2alpha(h) + SUMO-1(h) --> AP-2alpha(h){sumo}
Position is not known, enzyme is not known

AP-2gamma(h) + SUMO-1(h) --> AP-2gamma(h){sumoK10}
Position is known, enzyme is not known.

Note that binding reactions of SUMO-1 with other proteins are possible, without covalent modification.

11. Palmitoylation

Example:
Shh-N(h){chol} + palmityl-CoA --> Shh-N(h){chol}{palC1} + CoA
Here, enzyme is not known.

12. Depalmitoylation

Example:
H-Ras(m.s.){pal} --PPT(b)--> H-Ras(m.s.) + palmityl

13. Cholesterol modification

Example:
cholesterol + SHH(h) --> Shh-C(h) + Shh-N(h){chol}
Here, there is a combined effect, cleavage and cholesterol modification.

14. Myristoylation

Example:
G-alpha-i-1(m.s.) + myristoyl-CoA --> Galpha-i-1(m.s.){myr} + CoA
Here, enzyme is not known.

15. Demyristoylation

Example:
Galpha-i-1((m.s.){myr} --enzyme--> G-alpha-i-1(m.s.) + myristoyl

16. ADP-ribosylation

General view:

ProtA + NAD --EnzymeC--> ProtA{drib} + nicotinamide

Example:
G-alpha-i-2(r):G-beta(r):EGFR(r) + NAD --> G-alpha-i-2(r){drib}:G-beta(r):EGFR(r) + nicotinamide

17. Prenylation

(general mechanism including farnesylation and (geranyl)geranylation)

Example:
Rab3A(m.s.) --RabGGTase-alpha(m.s.)--> Rab3A(m.s.){pren}

18. (Geranyl)geranylation

Examples:
Rab7 + geranyl-PPi + NADPH --> Rab7{ger} + PPi + NADP
Here, enzyme is not known.

Rab7 + 2 geranyl-PPi + 2NADPH --> Rab7{ger(2)} + 2PPi + 2NADP
Case of geranylgeranylation.

19. Farnesylation

Example:
ProtA + farnesyl-PPi + NADPH --> ProtA{far} + PPi + NADP

20. Cleavage

A cleavage reaction gives rise to at least 2 protein fragments. With appropriate detection methods it is possible to monitor at least one of the specific cleavage products. Frequently, the amino acid sequence of the target site of the cleaving enzyme (or the cleavage site) are discussed in the paper. Products of a cleavage reaction may still be functional.

General view:

ProtA --Enzyme--> ProtA' (+ ProtA'' + ...) (+ protein remnants)

Example:
cTnT(r) --Caspase-3(r)--> cTNTp25(r) + protein remnants

21. Hydrolysis

Hydrolysis is a chemical process in which a molecule is split into two parts by the addition of a molecule of water.

Examples:

Hydrolysis of nucleotides, e.g. GTP to GDP (by GTPases):
RhoA(h):GTP --Graf2(m)--> RhoA(h):GDP + p

Hydrolysis of phospholipids:
PIP2 --PLCbeta3(h)--> DAG + IP3

22. Exchange

Dissociation of one molecule (e.g. GDP) and (concurrent) association with another molecule (e.g. GTP).

Example:

Ras:GDP + GTP --NO--> Ras:GTP + GDP
GTP-GDP-exchange on G proteins/GTPases (catalyzed by Guanine nucleotide exchange factors = GEFs)

23. Glycosylation

N-glycosylation occurs at the ER mainly through Asn (= 'N') residues.

O-glycosylation - most O-glycosylation occurs in the Golgi or in the cytosol, only a few in the ER to Ser(=S) or Thr(=T) residues.

General view:

ProtA + 2 NDP-Gly --enzyme(transferase)--> ProtA{gly(n)} + NMP + NDP

Glycosylation involves several steps of addition of different sugars e.g. GlcNAc, Man, Glu, etc. The very first step is the linkage of the first sugar to the protein at a specific amino-acid position. Afterwards several additions and "trimming" of sugars can occur, regulated by different metabolic enzymes. We are interested in the specific modified forms because they have different regulatory functions.

TRANSPATH intends to capture as much detail as possible depending on the knowledge provided by the reference.

Examples for more specific variations of the reaction above:

ProtA + 2 UDP-GlcNAc --enzyme--> ProtA{GlcNAc(2)} + UMP + UDP

ProtA{GlcNAc(2)} + 5 GDP-Man --enzyme--> ProtA{GlcNAc(2)}{man(5)} + 5 GDP

ProtA{GlcNAc(2)}{man(5)} + 4dolichol-p-Man + 3dolichol-p-Glc --enzyme-->
ProtA{GlcNAc(2)}{man(9)}{glc(3)} + 3dolichol-p + 4dolichol-p

24. Methylation

Protein methylation is one of the posttranslational modification reactions which modulates the function of proteins.

General view:

ProtA + S-Adenosylmethionine --enzyme--> ProtA{met} + S-Adenosylhomocysteine

Several articles report that a specific arginine residue is methylated at a certain position. With corresponding experimental evidence, this can be written as:

ProtA + S-Adenosylmethionine --enzyme--> ProtA{metR234} + S-Adenosylhomocysteine

25. Nitrosylation

In general, S-nitrosylation is referred to as nitrosylation. S-nitrosylation is a modification of a protein (adding NO group) at sulfur bearing amino acid residues, e.g. cysteine. Protein S-nitrosylation is an oxidative modification of protein sulfhydryl groups that may significantly modify cellular metabolism.

General view:

ProtA + NO --enzyme--> ProtA{no}

Example:
Dynamin is nitrosylated at cysteine residue 607 by an enzyme, nitric oxide synthase:
dynamin + NO --NO synthase--> dynamin{noC607}

26. Sulfation

Tyrosine residues in proteins may become sulfated shortly before they exit from the Golgi apparatus. The sulfation depends on the sulfate donor 3'-phosphoadenosine-5'-phosphosulfate (PAPS), that is transported from the cytosol into the lumen of the *trans*-Golgi network.

General view:

ProtA + PAPS --TPST--> ProtA{sulY} + 3',5'-ADP

TPST = tyrosylprotein sulfotransferase (EC 2.8.2.20)

PAPS = adenosine 3'-phosphate 5'-phosphosulfate (universal sulfate donor)

TPST catalyzes the transfer of sulfate from PAPS to the hydroxyl group of a peptidyltyrosine residue to form a tyrosine O4-sulfate ester and 3',5'-ADP.

27. Neddylation

Neddylation is a specific conjugation such as ubiquitination or sumoylation. NEDD8, the mammalian homolog of yeast Rub1, can likewise become covalently linked.

General view:

ProtA + Nedd8 --(E1),(E2),E3--> ProtA{neddK}

Deneddylation, which removes the Nedd8 moiety, requires the isopeptidase activity of the COP9 signalosome (CSN).

28. Hydroxylation

General view:

ProtA --enzyme--> ProtA{hyd}

(direct: direct; reversible: false; effect term: hydroxylation)

> Example:
> HIF-1(ce) --EGL-9(ce)--> HIF-1(ce){hydP621}
> In this example a proline residue at position 621 has been modified by hydroxylation.

Hydroxylation is any chemical process that introduces one or more hydroxyl groups (-OH) into a compound (or radical) thereby oxidising it. Thus hydroxylation can be regarded as a special case of an oxidising reaction. Hydroxylation reactions are often facilitated by enzymes called hydroxylases. Frequently proline and lysine residues are modified by hydroxylation.

29. Demethylimination

Conversion of a methylated arginine residue to citrulline.

General view:

ProtA{metR} --PAD4--> ProtA{cit} + methylamine

3.5.1. Indirect Reactions

30. DNA binding

General view:

ProtA --> GeneB

(direct: indirect; reversible: false; effect term: DNA binding)

Indirect effect.

Binding of protein A to the promoter/enhancer of gene X has been shown by electrophoretic gel mobility-shift assay (EMSA)/gel retardation assays or some kind of footprint analysis or South-Western blotting (or less frequent by protein DNA cross-linking with subsequent detection of the protein by an antibody).

Since "DNA binding" is not a term which implies quantification it may be combined with "transactivation", "transrepression" to show a quantitative effect. For the automatically created reactions, effect terms are combined as necessary.

31. Transactivation

General view:

ProtA --> GeneB

(direct: indirect; reversible: false; effect term: transactivation)

Indirect effect.

Protein A acts on (/regulates the transcription of) gene B. The action of protein A is shown by some kind of reporter gene assay (CAT, luciferase, β-gal…) Sometimes only the gene B promoter (or a respective promoter -enhancer construct) is used in reporter gene assays.

The action of protein A may also be shown by Northern analysis or quantitative RT-PCR of gene B derived mRNA or in array analysis.

As this term implies quantification, higher levels luciferase activity/mRNA should have been measured upon co-transfection/transformation of protein A. In references which focus more on gene regulation reporter gene assays using mutated and wildtype promoters/enhancers may be compared. In these cases the activity seen with the wildtype promoter/ promoter-enhancer construct should be higher.

32. Transrepression

General view:

ProtA --/ GeneB

(direct: indirect; reversible: false; effect term: transrepression)

Indirect effect.

Experimental set-up is similar as for "transactivation" except that lower levels of reporter gene activity or mRNA are mesured. In cases where mutated and wildtype promoters/enhancers are compared the activity seen with the wildtype promoter/ promoter-enhancer construct should be lower.

Note: transregulation

In former times under a different paradigm this term was used to indicate the action of a protein on a gene/transcriptional regulation of a gene by a protein. Since experimental set-up's show up- or down regulation of transcription(/"gene expression"), the terms transactivation or transrepression are applied instead now.

33. Increase of abundance

General view:

ProtA --> ProtB

(direct: indirect; reversible: false; effect term: increase of abundance)

> Example XN000061049:
> SMAR1(m) --> p21Cip1(h)
> The effect of murine SMAR1 was shown with the help of Western blots. The authors of the respective paper mentioned it as "increase of expression".

Indirect effect.

Protein A acts on Protein B.

Upon action of protein A more protein B is detected by Western analysis/immunoblot experiment, ELISA, scintillation counting.

34. Decrease of abundance

General view:

ProtA --/ ProtB

(direct: indirect; reversible: false; effect term: decrease of abundance)

Indirect effect.

Experimental set-up is similar as for "Increase of abundance" except that less protein B is detected upon action of protein A.

Note: regulation of abundance

Frequently used in former times under a different paradigm that is discontinued. Because experiments show up- or down regulation of protein B this quantitative effect is curated rather than just the regulation itself.

Note: Regulation/increase/decrease of abundance of mRNA is captured as: transactivation or transrepression. Please see the paragraphs on transactivation, transrepression and transregulation above.

35. Expression

General view:

GeneB --> ProtB

GeneB --> ProtB-xbb1

(direct: indirect; reversible: false; effect term: expression)

Indirect effect.

Gene B gives rise to protein B. Frequently there is no explicit expression reaction in the reference. However, from the use/appearance of a protein in another type of reaction it can be deduced that the gene encoding it gives rise to it. Because expression is a multi step process involving transcription and translation, expression is regarded to be an indirect effect.

36. Activation

General view:

ProtA --> ProtB

(direct: indirect; reversible: false; effect term: activation) where ProtA is, for example, a transcriptional co-factor.

Is used, for example, when a reporter gene assay had been used to show that a transcription co-factor potentiated/enhanced (acts on) the activation of a transcription factor.

37. Increase of binding

General view:

ProtA --> ProtB:ProtC

Example:
Crk(m.s.) --> p110alpha(h):H-Ras(m.s.)
v-Crk enhances the interaction of H-Ras with PI3K p110 Ras-binding domain.

38. Increase of phosphorylation

General view:

ProtA --> ProtB{p}

Example:
ATR(m.s.) --> WRN(m.s.){p}
Imported from Proteome HumanPSD - mammalian ATR indirectly increases the phosphorylation of mammalian WRN.

3.5.2. Reaction Effect Terms

acetylation

one acetyl group at least is attached to a molecule; posttranslational modification of proteins, generally at the N-terminus

activation

increase of protein activity; reactions going out from the target molecule will follow after this reaction

acylation

process of adding an acyl group to a compound

ADP-ribosylation

the transfer of ADP-ribose from NAD to protein amino acids [1]

amidation

addition of an amide group from a glycine to a protein amino acid [1]

assembly

congregation of molecules in a certain subcellular location (e.g. membranous lipid rafts)

automodification

"modification reaction where the enzyme is also the substrate; should be used together with a specific modification term, e.g. phosphorylation"

binding

any kind of chemical/physical bond (covalent bond, hydrogen bond, ion complex, Van-der-Waals) that is established between reaction partners - if not defined more stringently as condensation, acetylation etc.

carboxylation

addition of a carboxy group to a protein amino acid [1]

cholesterol modification

cleavage

peptide bond between amino acids is cleaved by an endoprotease

competition

two or more molecules competing for a reaction partner/ signal

condensation

conjugation

coupling of a molecule with at least one ubiquitin-like molecule (e.g. Nedd8/Rub1)

deacetylation

removal of acetyl group

deacylation

removal of an acetyl group

deamidation

removal of an amide group

decarboxylation

removal of a carboxy group

deconjugation

decrease of abundance

abundance of a protein is decreased by an indirect mechanism

decrease of acetylation

acetylation of a protein is decreased by an indirect mechanism

decrease of acylation

acylation of a protein is decreased by an indirect mechanism

decrease of ADP-ribosylation

ADP-ribosylation of a protein is decreased by an indirect mechanism

decrease of amidation

amidation of a protein is decreased by an indirect mechanism

decrease of binding

binding of a protein to a third component is decreased by an indirect mechanism

decrease of carboxylation

carboxylation of a protein is decreased by an indirect mechanism

decrease of cholesterol modification

decrease of cleavage

cleavage of a protein is decreased by an indirect mechanism

decrease of conjugation

conjugation of a protein is decreased by an indirect mechanism

decrease of DNA binding

DNA binding of a protein is decreased by an indirect mechanism

decrease of folding

folding of a protein is decreased by an indirect mechanism

decrease of glycosylation

glycosylation of a protein is decreased by an indirect mechanism

decrease of hydroxylation

hydroxylation of a protein is decreased by an indirect mechanism

decrease of isomerization

isomerization of a protein is decreased by an indirect mechanism

decrease of methylation

methylation of a protein is decreased by an indirect mechanism

decrease of myristoylation

myristoylation of a protein is decreased by an indirect mechanism

decrease of neddylation

neddylation of a protein is decreased by an indirect mechanism

decrease of N-glycosylation

N-glycosylation of a protein is decreased by an indirect mechanism

decrease of nitration

decrease of nitrosylation

nitrosylation of a protein is decreased by an indirect mechanism

decrease of O-glycosylation

decrease of oligomerization

oligomerization of a protein to binding partners is decreased by an indirect mechanism

decrease of palmitoylation

palmitoylation of a protein is decreased by an indirect mechanism

decrease of phosphorylation

"phosphorylation of a protein is decreased by an indirect mechanism; can occur via dephosphorylation or via inhibition of phosphorylation"

decrease of processing

processing of a protein is decreased by an indirect mechanism

decrease of secretion

secretion of a protein is decreased by an indirect mechanism

decrease of sumoylation

sumoylation of a protein is decreased by an indirect mechanism

decrease of transport

transport of a protein is decreased by an indirect mechanism

decrease of ubiquitination

ubiquitination of a protein is decreased by an indirect mechanism

deglycosylation

removal of a glycosyl group

degradation

"breakdown of a molecule (leaving behind protein remnants; mediated by enzymes, proteasome)"

dehydroxylation

removal of a hydroxy group

demethylation

removal of a methyl group

demyristoylation

removal of a myristoyl moiety

deneddylation

removal of one or more NEDD8 moieties from a protein

denitration

denitrosylation

removal of a nitrosyl group

depalmitoylation

removal of a palmitoyl moiety

dephosphorylation

catalyzed separation of a phosphate group from a molecule (mediated by a phosphatase)

destabilization

additional association of a protein decreases the stability/affinity of an existing complex

desulfation

removal of a sulfate group

desumoylation

removal of one or more SUMO moieties from a protein

deubiquitination

removal of one or more ubiquitin moieties from a protein [1]

dissociation

breakup of a complex

DNA binding

binding of a protein to the promoter/enhancer of a gene details

exchange

catalyzed replacement of a group by another (e.g. exchange of G-protein associated GDP with GTP by a GEF)

expression

transfer of information encoded in the DNA (nucleotide sequence) into the protein (amino acid sequence); reactions with effect expression include several steps: transcription, splicing, capping, and translation

glycosylation

transfer of one or more glycosyl groups to a molecule, resulting in a glycoprotein; posttranslational modification

hydrolysis

catalyzed splitting of a chemical bond with the consumption of water

hydroxylation

any chemical process that introduces one or more hydroxyl groups (-OH) into a compound (or radical) thereby oxidizing it

increase of abundance

abundance of a protein is increased by an indirect mechanism

increase of acetylation

acetylation of a protein is increased by an indirect mechanism

increase of acylation

acylation of a protein is increased by an indirect mechanism

increase of ADP-ribosylation

ADP-ribosylation of a protein is increased by an indirect mechanism

increase of amidation

amidation of a protein is increased by an indirect mechanism

increase of binding

binding of a protein to a third component is increased by an indirect mechanism

increase of carboxylation

carboxylation of a protein is increased by an indirect mechanism

increase of cholesterol modification

increase of cleavage

cleavage of a protein is increased by an indirect mechanism

increase of conjugation

conjugation of a protein is increased by an indirect mechanism

increase of DNA binding

DNA binding of a protein is increased by an indirect mechanism

increase of folding

folding of a protein is increased by an indirect mechanism

increase of glycosylation

glycosylation of a protein is increased by an indirect mechanism

increase of hydroxylation

hydroxylation of a protein is increased by an indirect mechanism

increase of isomerization

isomerization of a protein is increased by an indirect mechanism

increase of methylation

methylation of a protein is increased by an indirect mechanism

increase of mRNA decay

miRNA triggered increase of target mRNA decay thereby also inhibiting translation

increase of myristoylation

myristoylation of a protein is increased by an indirect mechanism

increase of neddylation

neddylation of a protein is increased by an indirect mechanism

increase of N-glycosylation

N-glycosylation of a protein is increased by an indirect mechanism

increase of nitration

increase of nitrosylation

nitrosylation of a protein is increased by an indirect mechanism

increase of O-glycosylation

increase of oligomerization

oligomerization of a protein to binding partners is increased by an indirect mechanism

increase of palmitoylation

palmitoylation of a protein is increased by an indirect mechanism

increase of phosphorylation

phosphorylation of a protein is increased by an indirect mechanism

increase of processing

processing of a protein is increased by an indirect mechanism

increase of secretion

secretion of a protein is increased by an indirect mechanism

increase of sumoylation

sumoylation of a protein is increased by an indirect mechanism

increase of transport

transport of a protein is increased by an indirect mechanism

increase of ubiquitination

ubiquitination of a protein is increased by an indirect mechanism

inhibition

reactions going out from the target molecule will stop after this reaction or will follow if this reaction would not signal; decrease of protein activity

interaction

binary relation which states nothing about directionality between two molecules with yet undefined mechanism; both molecules are entered in the 'molecules upstream' list

isomerization

directed conformational change between cis- and trans-conformation of a peptide chain

methylation

posttranslational modification, covalent transfer of a methyl-group to a substrate

myristoylation

covalent or non-covalent attachment of a myristoyl moiety to a protein amino acid [1]

neddylation

specific conjugation such as ubiquitination or sumoylation; NEDD8, the mammalian homolog of Rub1, can likewise become covalently linked

N-glycosylation

glycosylation of an N atom of a free alpha amino terminal of a peptide [1]

nitration

nitrosylation

covalent transfer of NO to a substrate; S-nitrosylation (transfer of NO to Cys-residues) of a protein controls its activity similar to O-phosphorylation

O-glycosylation

oligomerization

assembly of a complex; resumes the terms homo- and heteromerization and di-/trimerization as well

palmitoylation

covalent or non-covalent attachment of a palmitoyl moiety to a protein amino acid [1]

phosphorylation

catalyzed condensation of a phosphate group to a molecule (for example mediated by protein kinases)

predicted

in silico predicted interaction of a miRNA with a complementary site of a target mRNA

prenylation

class of lipid modification involving covalent addition of either farnesyl (15-carbon) or geranylgeranyl (20-carbon) isoprenoids to conserved cysteine residues at or near the C-terminus of proteins [2]

processing

catalyzed maturation of proteins (usually associated with proteolytic cutting of the signal peptide or more)

redox reaction

a reaction in which one or more electrons are transferred

regulation of abundance

abundance of a protein is regulated in an undefined way

regulation of acetylation

acetylation status of a protein is regulated in an undefined way

regulation of activity

activity of a protein is regulated in an undefined way

regulation of amidation

amidation status of a protein is regulated in an undefined way

regulation of carboxylation

carboxylation status of a protein is regulated in an undefined way

regulation of cleavage

cleavage of a protein is regulated in an undefined way

regulation of degradation

degradation of a protein is regulated in an undefined way

regulation of folding

folding of a protein is regulated in an undefined way

regulation of glycosylation

glycosylation status of a protein is regulated in an undefined way

regulation of localization

localization of a protein is regulated in an undefined way

regulation of methylation

methylation status of a protein is regulated in an undefined way

regulation of N-glycosylation

N-glycosylation status of a protein is regulated in an undefined way

regulation of nitration

nitration status of a protein is regulated in an undefined way

regulation of oxidation

oxidation status of a protein is regulated in an undefined way

regulation of phosphorylation

phosphorylation status of a protein is regulated in an undefined way

regulation of processing

processing of a protein is regulated in an undefined way

regulation of secretion

secretion of a protein is regulated in an undefined way

regulation of stability

stability of a protein is regulated in an undefined way

regulation of sulfation

sulfation status of a protein is regulated in an undefined way

regulation of transport

transport of a protein is regulated in an undefined way

regulation of ubiquitination

ubiquitination status of a protein is regulated in an undefined way

stabilization

additional association of a protein enhances the stability of an already existing complex

sulfation

catalyzed transfer of sulfate from PAPS (adenosine 3'-phosphate 5'-phosphosulfate) to the hydroxyl group of a peptidyltyrosine residue to form a tyrosine O4-sulfate ester and 3',5'-ADP

sumoylation

process by which a SUMO protein (small ubiquitin-related modifier) is conjugated to a target protein via an isopeptide bond between the carboxyl terminus of SUMO with an epsilon-amino group of a lysine residue of the target protein [1]

transactivation

general term for gene activation, involves transcription details

translational repression

mRNA translation repression by a miRNA

translocation

indicates the regulated transfer of a signaling molecule to another subcellular location (often from cytosol to nucleus)

transregulation

transcriptional regulation by binding of a transcription factor to its cognate site within the gene regulatory region; the term transregulation does not differentiate between activation or inhibition

transrepression

repression of a gene by a transcription factor

ubiquitination

coupling of a molecule with ubiquitin

unknown

for semantic reactions where it's unclear if the signal acceptor is activated or inhibited

[1] Definition taken from GO; Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium (2000) Nature Genet. 25: 25-29.

[2] Zhang F. L., Casey P. J.; Protein prenylation: molecular mechanisms and functional consequences. (1996), Annu. Rev. Biochem. 65, 241-269.

3.6. Fields

It should be noted that in individual entries, some fields may be empty. In this case, these fields are not displayed.

Field	Content and format
AC	Accession number

The accession number is the unique identifier for each entry. Its format is "XN" in capital letters followed by nine digits (e.g. XN000012345). If two entries have to be merged , the AC of the primary name is retained. The other AC will be stored in the secondary accession numbers (AS) field.

AS Accession numbers, secondary

The secondary accession numbers are optional alternate identifiers for each entry. They are of the form defined in AC, separated by semicolons, and are created when two entries are merged.

DT Created by

The name of the curator who created the entry. E-mail your feedback directly to the curator.

Updated by

The name of the curator who last updated the entry. E-mail your feedback directly to the curator.

CO Copyright-Information

TY Type

Marks the kind of reaction the entry tells about. It supports the PathwayBuilderTM in coloring/skipping/filtering. A comparison of the different reaction types is given in Table 1.

Terms used:
* molecular evidence; direct
* molecular evidence; indirect
* pathway step; direct
* pathway step; indirect
* pathway step; metabolic
* decomposition; direct
* semantic; direct
* semantic; indirect

The second term indicates if the described interaction in the reaction involves unknown in-between steps (i.e. is indirect or direct). In rare cases, where this differentiation is not clear in the literature, the term unknown is shown.

QU Quality

Terms used:
* Reliability information (scale 1-5) (established for molecular evidence reactions). Material and methods, which are the basis of the evaluation, are listed under comments. For further information please see our quality evaluation concept, Chapter 8 (quality.txt)
* putative o marks reactions that have no or just partial experimental evidence, but are necessary and in compliance with the current common hypothesis of the scientific community

NA Name

Contains the name of a reaction. Besides it, this field can be used to indicate general physiological effects which a molecule, which is linked to this reaction entry has in a direct or indirect manner. For this kind of reaction entry there is only a signal donor or acceptor, not both; e.g. apoptosis Reaction names have different arrows indicating their type:

*<==> equilibrium reactions (complex formation and dissociation)
* --E--> mass flow catalyzed by enzyme E
* --> semantic activation or unidirectional mechanistic step
* --/ semantic inhibition

Stoichiometry is used like this: 2 A --> (A)2 or 2 A --> A:A; if two identical complexes are binding the syntax is: 2 (A:B:C) <> (A:B:C)2

Semantic reactions with two (or more) signal donors are formulated like: A & B --> C

For specification of the Name syntax and examples, see Section 3.5 (reaction_names.txt)

A list of molecule tags is given in Section 2.6 (nametag_loc.txt).

For a list of species tags, see Chapter 7 (nametag_spec.txt)

EF Effect

An explanation of the reaction semantics. For semantic reactions, this always contains activation or inhibition or unknown first. The use of inhibition (activation) means, that the signal acceptor is down-regulated (up-regulated). The other terms briefly describe reaction mechanisms.

List of the effect terms used is given in Section 3.5.2 (effect.txt)

The term "reversible" indicates if a reaction is reversible, such as a complex formation.

HP Pathway level

Lists all pathway step reactions that are linked to this entry in the reaction hierarchy.

HS Semantic level

Lists all semantic reactions that are linked to this entry in the reaction hierarchy.

HE Evidence level

Lists all molecular evidence reactions that are linked to this entry in the reaction hierarchy.

DP Decomposed pathway steps

Lists the pathway step reactions to which a decomposition reaction is linked.

DC Decompositions

Lists all decomposition reactions that are linked to a pathway step reaction.

DR External database hyperlink

Database name (e. g. EMBL/GenBank/DDBJ): database accession number; identifier.

CC Comments

A list of annotation categories is given in Chapter 6 (annotation.txt).

CP Location positive and experiment(s)

Lists all locations where the reaction has been shown to take place. In most of the cases these are cell lines used in experiments. Gives experiment abbreviations for material(s) and method(s) used to verify the reaction and cites reference(s).

CN Location negative and experiment(s)

Lists all locations where the reaction has been shown NOT to take place.

MB Molecule/gene upstream

All molecules/genes which the reaction consumes, receives a signal from or is an interaction for.

MA Molecule/gene downstream

All the molecules/genes which the reaction produces or signals to.

MC Catalysts

Lists enzymes and other catalysts such as complexes or allosteric modifiers involved in the reaction.

MI Inhibitors

The inhibitors inhibiting this reaction. But inhibitory situations are principally modeled with reactions. Only where the mechanism of the inhibition is unclear, this field is used

RN Reference number

[consecutive entry reference number]. A list of the papers from which the information in this entry was extracted.

RX Medline database hyperlink

The PMID number in PubMed

RA Reference author(s)

List of authors

RT Reference title

RL Reference publication

4. Table GENE

Gene information in TRANSPATH® was formerly provided in the molecule table. To have a clear separation between genes and gene products, they are now stored in a separate table. All genes are linked to the TRANSFAC® Gene table, where users can find information about the structure of gene regulatory regions, including individual binding sites for transcription factors, and other information as well. Gene entries have hyperlinks to UniGene and Entrez Gene, the human ones also to OMIM and Affymetrix microarray chips. Genes have been included to provide information about the last step in signal transduction pathways: the regulation of target genes by activated transcription factors. Thus, TRANSPATH® presents information about complete signaling pathways: starting with the activation of a receptor at the membrane through a cascade of kinases into the nucleus, where a particular transcription factor is activated and regulates a set of target genes.

The link between a gene and its gene product(s) is given via reactions with the effect 'expression'. Since the expressed proteins can be signaling molecules or transcription factors themselves, this information is an important step to include feedback loops in signaling networks.

4.1. Fields

It should be noted that in individual entries, some fields may be empty. In this case, these fields are not displayed.

Field Content and format

AC Accession number

The accession number is the unique identifier for each entry. Its format is "G" followed by six digits (e.g. G012345). If two entries have to be merged, the AC of the primary name is retained. The other AC will be stored in the secondary accession numbers (AS) field.

AS Accession numbers, secondary

The secondary accession numbers are optional alternate identifiers for each entry. They are of the form defined separated by semicolons, and are created when two entries are merged. The former molecule table accession number is also given here.

DT Created by

The name of the curator who created the entry. E-mail your feedback directly to the curator.

Updated by

The name of the curator who last updated the entry. E-mail your feedback directly to the curator.

CO Copyright-information

NA Name

A short name for the gene, such as the HGNC symbol. It corresponds to the short description for the respective gene in TRANSFAC. A hyperlink with the TRANSFAC accession number is provided in front of the name.

There are tags appended to the name to differentiate the species the gene comes from. This short identifier is useful in reaction names because molecules and genes from different species often interact due to the experiments.

SY Synonyms

Alternative names for the gene. This is needed because the names in biosciences change often. The field lists other names or abbreviations. Different synonyms are separated by a semicolon. The field synonyms is automatically included if you run a query with the search field name.

TY Type

The type of this gene entry. Possible terms are:

*orthogene, group entry for homologous genes in different species
*basic, for taxa-specific genes (mostly species-specific, but also higher taxa e.g. mammalia)

ME Gene product

Proteins encoded by this gene. In addtion, molecule types are shown to allow differentiation between entries from distinct molecule hierarchy levels.

HP Superfamilies

Lists all groups (orthogene entries) this gene belongs directly to.

HC Subfamilies

Lists all basic gene entries.

OS Species

The species that the gene entry belongs to. Given is the common name (if it exists), followed by the Latin denomination, as it is done in TRANSFAC.

DR External database hyperlink

Database name (e.g. EMBL/GenBank/DDBJ): database accession number; identifier.

XB Reaction upstream

A list of reactions that lead to this gene. Here the gene serves as a signal acceptor. Up to now, these are all semantic transregulation reactions

XA Reaction downstream

A list of semantic expression reactions that go out from this molecule.

RN Reference number

[consecutive entry reference number]. A list of the papers from which the information in this entry was extracted.

RX Medline database hyperlink

The PMID number in PubMed

RA Reference author(s)

List of authors

RT Reference title

RL Reference publication

5. Table PATHWAY

The pathway flat file contains two types of entities: pathways and chains. Pathways reflect canonical pathways for specific signaling molecules (mostly ligands or receptors) and are made up of one or more chains. Chains are sets of consequent reactions, i.e. the reactions contain common molecules. Chains that are linked to pathways consist of reactions of the type 'pathway step'. Chains can have bifurcations and even loops (if they carry a regulatory meaning).

The construction of chains and pathways is based on recent (primary and review) literature and on the expert knowledge of the annotator who has worked intensely on the specific topic. In some cases, when a sequence of reactions has been reported by one primary paper, the chain is labeled in the type field as 'evidence chain'. We postulate that the physiological relevance of a reaction sequence is higher when at least one report has proved it in a consistent experimental system.

Chains that are constructed from metabolic pathway step reactions have the term 'metabolic chain' in the type field. Metabolic pathway and chain data in TRANSPATH® reflects canonical metabolic pathway knowledge from most recent textbooks.

It should be noted that in individual entries, some fields might be empty. In this case, these fields are not displayed.

5.1. Fields

Field Content and format

AC Accession number

The accession number is the unique identifier for each entry. Its format is "CH" in capital letters followed by nine digits (e.g. CH000012345). If two entries have to be merged, the AC of the primary name is retained. The other AC will be stored in the secondary accession numbers (AS) field.

AS Accession numbers, secondary

The secondary accession numbers are optional alternate identifiers for each entry. They are of the form defined in AC, separated by semicolons, and are created when two entries are merged.

DT Created by

The name of the curator who created the entry. E-mail your feedback directly to the curator.

Updated by

The name of the curator who last updated the entry. E-mail your feedback directly to the curator.

CO Copyright Information

TY Type

Marks the kind of chain the entry tells about.

Terms used:
* pathway
*chain
*metabolic chain
*evidence chain

NA Name

Names of pathway entries are oriented on characteristic molecules (e.g. ligands, receptors, transcription factors) and are the same as for the respective prepared PathwayBuilder? maps and the hand-drawn maps.

Names of chain entries address start and end molecules (A ---> X), which in case of bifurcations can be several (A, B ---> X, Y, Z). To differentiate between chains with the same start and end molecules, a prominent molecule of the in-between path is indicated (A ---M---> X). In some cases the name is a term for a physiological process such as G1 phase (cell cycle).

HP Pathway level

Lists all pathways that are superordinated.

HC Chain level

Lists all chains that are subordinated.

CC Comments

A list of annotation categories is given in Chapter 6 (annotation.txt).

XN Reactions involved

A cascade view of all reactions the pathway/chain consists of. The level of identation of the red arrows indicates subsequent reactions. Grey arrows and italic reaction names depict the second occurrence of a reaction in a cascade.

MO Molecules involved

A list of all molecules the pathway/chain consists of.

RN Reference number

[consecutive entry reference number]. A list of the papers from which the information in this entry was extracted.

RX Medline database hyperlink

The PMID number in PubMed

RA Reference author(s)

List of authors

RT Reference title

RL Reference publication

6. Annotation

Annotations are additional comments for an entry. Each annotation belongs to a comment category and contains free text. A reference link is given in square brackets after the comment, if available.

Category - a tag to categorize the type of comment.

List of categories used:

* CONFLICT! (text mentions both ACCNOS)
    - If a conflict exists between two papers, this is used to indicate that one exists, and in addition to the annotates of those papers, a third annotate with category "CONFLICT!" is entered, which in its text body mentions the accession numbers of those papers.
* disease
    - Tells about in which disease(s) the molecule is involved. Queries can be executed by using keywords out of a list.
* enzyme classification
    - Mentions the EC No.
* experimental activator
    - Names artificial compounds that can activate reactions or the function of molecules.
* experimental inhibitor
    - Names artificial compounds that can inhibit reactions or the function of molecules.
* general
    - Contains comments that do not fit elsewhere.
* GO: biological process
* GO: molecular function
    - Associated GO-terms (Gene Ontology) and GO evidence code, e.g. "induction of proapoptotic gene products; IEA"; hyperlink containing the corresponding GO ID (here GO:0008633)
    - Associations of TRANSPATH® Molecules with Gene Ontology terms (Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium (2000) Nature Genet. 25: 25-29) were mainly obtained by electronic annotation using a gene association data list from Compugen Inc. (http://www.cgen.com/, http://www.labonweb.com).
* historical
* kinetics/thermodynamics
    - Specification of kinetic or thermodynamic parameters, e.g. KM = 2.4 mmol/l.
* mechanism

- Gives information about the mechanism of a reaction, e.g. signaling by substructures (motifs, part of complexes) or phosphorylation sites, e.g. xxx is phosphorylated at Tyr32.

* methods
  - Contains additional information about methods and material used in experiments that are listed in the fields location positive/negative and experiment(s).

* physiological function
  - Mentions physiological or environmental influences that affect the certain molecule or reaction, or tells something about the effects the entry has; e.g. IL-1 has inflammatory effects.

* sequence

* structure

7. Biological Species

List of species abbreviations

* (v.s.) vertebrate species
* (m.s.) mammalian species

* (a) thale cress, Arabidopsis thaliana
* (aa) American alligator, Alligator mississippiensis
* (ac) Egyptian spiny mouse, Acomys cahirinus
* (AD40) human adenovirus type 40
* (ae) african elephant, Loxodonta africana
* (ai) Amia calva
* (an) Aspergillus nidulans
* (ao) night monkey, Aotus trivirgatus
* (ap) punctuate sea urchin, Arbacia punctulata
* (as) starfish, Asterina pectinifera
* (b) cattle, Bos taurus
* (ba) Batrachoididae sp
* (bc) Baltic cod, Gadus callarias
* (bf) Japanese toad, Bufo japonicus
* (bm) oriental fire-bellied toad, Bombina orientalis
* (bv) yellow-bellied toad, Bombina variegata
* (c) chick, Gallus gallus
* (ca) dog, Canis familiaris
* (cb) chinchilla, Chinchilla brevicaudata
* (cc) common carp, Cyprinus carpio
* (cd) dromedary, Camelus dromedarius
* (ce) Caenorhabditis elegans
* (cg) casiragua, Proechimys guairae
* (ch) green sea-turtle, Chelonia mydas caranigra
* (cj) common marmoset, Callithrix jacchus
* (cl) domestic pigeon, Columba livia
* (cm) elephantfish, Callorhynchus milii
* (cr) Chlamydomonas reinhardtii
* (ct) red-crowned mangabey, Cercocebus torquatus atys
* (cx) Callothrix sp.
* (d) fruit fly, Drosophila melanogaster
* (dd) slime mold, Dictyostelium discoideum
* (de) degu, Octodon degus
* (dr) Russell's viper, Daboia russelli russelli
* (du) domestic duck, Anas platyrhynchos
* (ea) donkey, Equus asinus
* (EBV) Epstein-Barr virus
* (ee) American eel, Anguilla rostrata
* (fc) cat, Felis catus
* (fh) killifish, Fundulus heteroclitus
* (fl) European flounder, Platichthys flesus
* (fr) Japanese pufferfish, Fugu rubripes
* (fw) finback whale, Balaenoptera physalus
* (fy) fission yeast, Schizosaccharomyces pombe
* (gl) golden (Syrian) hamster, Mesocricetus auratus
* (go) goat, Capra hircus
* (gr) gorilla, Gorilla gorilla
* (gp) guinea pig, Cavia porcellus
* (h) human, homo sapiens
* (ha) hamster, Cricetulus griseus
* (ho) horse, Equus caballus

* (ie) Indian elephant, Elephas maximus
* (ip) channel catfish, Ictalurus punctatus
* (JCV) JC polyoma virus
* (jq) Japanese quail, Coturnix coturnix japonica
* (lh) long-tailed hamster, Cricetulus longicaudatus
* (ls) Lepisosteus spatula
* (m) mouse, Mus musculus
* (md) muscovy duck, Cairina moschata
* (mf) cynomolgus monkey, Macaca fascicularis
* (mg) turkey, Meleagris gallopavo
* (ml) marbled lungfish, Protopterus aethiopicus
* (mn) pig-tailed macaque, macaca nemestrina
* (mo) monkey, Cercopithecus aethiops (african green monkey)
* (ms) shorthorn sculpin, Myoxocephalus scorpius
* (mu) Mongolian jird, Meriones unquiculatus
* (mv) mink, Mustela vison
* (mx) atlantic hagfish, Myxine glutinosa
* (nj) Chinese cobra, Naja atra
* (nn) Indian cobra, Naja naja
* (nu) nutria, Myocastor coypus
* (ob) olive baboon, Papio anubis
* (oc) coho salmon, Oncorhynchus kisutch
* (ok) chum salmon, Oncorhynchus keta
* (ol) medaka, Oryzias latipes
* (on) Oreochromis niloticus
* (op) opossum, Didelphis marsupialis virginiana
* (ou) orangutan, Pongo pygmaeus
* (pc) yellow baboon, Papio cynocephalus
* (pd) Dumeril's clam worm, Platynereis dumerilii
* (pg) domestic pig, Sus scrofa
* (ph) hamadryas baboon, Papio hamadryas
* (pp) Papio sp.
* (po) porcupine, Hystrix cristata
* (ps) Psammomys obesus
* (pt) chimpanzee, Pan troglodytes
* (r) rat, Rattus norvegicus
* (rb) rabbit, Oryctolagus cuniculus
* (rd) red deer, Cervus elaphus
* (rf) rabbit fish, Chimaera monstrosa
* (rm) rhesus macaque, Macaca mulatta
* (ro) Rodentia sp
* (rr) laughing frog, Rana ridibunda
* (rs) rattlesnake, Crotalus atrox
* (rt) rainbow trout, Oncorhynchus mykiss
* (s) sheep, Ovis aries
* (sa) salmon, Oncorhynchus gorbuscha
* (sb) beechey ground squirrel, Spermophilus beecheyi
* (sc) spotted catshark, Scyliorhinus canicula
* (si) Salmo irideus
* (sl) sea lamprey, Petromyzon marinus
* (sm) common squirrel monkey, Saimiri sciureus
* (sq) spiny dogfish, Squalus acanthias
* (sr) hummingbird, Selasphorus rufus
* (ss) ochre sea star, Pisaster ochraceus
* (sv) SV40, simian virus 40
* (tg) roughskin newt, Taricha granulosa
* (th) bluefin tuna, Thunnus thynnus
* (tm) marbled electric ray, Torpedo marmorata
* (tu) tuna, Katsuwonus pelamis
* (uc) spoonworm, Urechis caupo
* (w) wheat, Triticum aestivum

* (wh) sei whale, Balaenoptera borealis
* (x) clawed frog, Xenopus laevis
* (xb) Kenyan clawed frog, Xenopus borealis
* (xi) Xiphophorus helleri
* (y) yeast, Saccharomyces cerevisiae
* (z) zebra fish, Brachydanio rerio
* (zd) Zaocys dhumnades

8. Quality assessment

TRANSPATH presents more than an extraction of data from the published literature. It also includes a quality evaluation and a reliability code.

The reliability of experimental evidence in the sense of biological relevance is defined by the source of the biological material and the method used for indication (Table 1). Our annotators evaluate the quality of experimental data from primary literature. The reliability code given in the quality matrix (Table 2) is a combination from a list of methods and a category of the biological material.

Biological material can be extracted from a complete organism, for example a knock-out mutant, or recombinant polypeptides. Thus, reactions obtained from experiments with different biological materials have different probabilities for existing in vivo. We provide a reliability value that is defined by each possible combination of category and method on a scale of 1 to 5 (highest to lowest reliability, Table 3). The value is attached to mechanistic reactions. The annotators use only data from peer-reviewed publications. If controls are missing or are not credible in a specific paper, the data will not be included.

Table 1: Biological material and methods: abbreviations and classification

Material -- Definition/Examples	
A1	Wild-type
A2	Natural chromosomal mutants leading to new phenotype
A3	Knock-out
A4	Transgenic animals
A5	Induced, but not defined mutations
B1	Embryonic cells
B2	Stem cells -> pluripotent
B3	Primary cells
B4	Recombinant cells/tissue/organisms (transfected); not permanent cell lines
C1	Permanent tissue culture cells
D1	Permanent or transient transfected or infected cells (tissue culture cells)
D2	Recombinant proteins produced e.g. from Baculo virus vectors in insect cells, transfected/microinjected Xenopus oocytes
E1	Recombinant proteins from Yeast
F1	Recombinant proteins from e.g. E.coli, Bacillus subtilis
F2	In vitro translation, synthetic peptides
X1	Recombinant proteins from unspecified cells
X2	Unspecified material
Method - Description/Examples	
M1	Yeast two-hybrid system

M2	Screening of expression libraries with other (labeled) proteins, e.g. phage display, eukaryotic cDNA library with EXIox vectors
M3	(Co-) localization of proteins in cells and tissues by microscopical analysis, e.g. immunofluorescence, in situ detection, electron microscopy
M4	Affinity precipitation in solution/batch (e.g. with sepharose or magneto beads), e.g. immunoprecipitation, pull-down assay
M5	In vitro protein-protein binding assay (solid phase), e.g. ELISA (Enzyme Linked Immuno Sorbent Assay)
M6	Surface plasmon resonance; detection and quantification of ligands
M7	Far Western or Gel overlay or protein-lipid overlay assay; gel overlay: detection of binding partners on blotted proteins after transfer from a gel
M8	Affinity chromatography with matrices; proteins coupled to e.g. sepharose or agarose matrices
M9	Peptide spot; search for binding partners of defined peptide motifs on a membrane
M10	Crosslinking; detection by SDS-Page and Western blot
M11	EMSA (electrophoretic mobility shift assay)
M12	Binding assay; binding of labeled ligand to receptor-expressing cells, quantitative evaluation, for example Scatchard analysis
M13	Modification of proteins in living cells, e.g. proteinphosphorylation and -dephosphorylation, ubiquitination, farnesylation; detection of the modified proteins, e.g. from cell extracts
M14	Subcellular extraction of proteins; fractionation for localization of proteins
M15	Dot spots; spotted proteins detected with specific antibodies
M16	In vitro modification of a substrate, e.g. phosphorylation or dephosphorylation (e.g. Kinase-assay), ubiquitination
M17	Reporter gene assay, e.g. luciferase assay, beta-galactosidase
M18	Western blot, SDS page
M19	mRNA detection by Northern blot; indirect evidence for protein interaction
M20	Localization of mRNA in tissues by e.g. FISH, ISH (p32, H3); indirect evidence for protein interaction
M21	Detection of mRNA by RT-PCR; reverse transcriptase, thereafter PCR; quantitative mRNA detection; indirect evidence for protein interaction
M22	Mammalian two (or tri) -hybrid assay
M23	Co-sedimentation and -purification by chemico-physical methods; biochemical purification of stable complexes by e.g analytical ultracentrifugation, size-exclusion chromatography (SEC)
M24	Protein cleavage and specific degradation, documented by protein detection of, e.g., S35-labeled or biotinylated protein fragments, immunoblots, SDS-PAGE only,

	HPLC, mass spectral analysis
M25	Interaction measurement by chemico-physical methods, e.g. crystal structure analysis of complexes; spectrofluorometer FRET: fluorescence resonance energy transfer; intensity and/or wavelength of fluorescence of a given interacting compound changes, when another molecule binds to it; quantitative; usually done in vitro; also NMR: looking at changes in nuclear spin, usually done with peptide fragments and in unphysiological solutions, isothermal titration calorimetry (ITC)
M26	mRNA expression profile analysis, e.g. microarray data
M27	Transmembrane potential measurements, e.g. current-voltage experiments
M28	Predicted by binding-site sequence
M29	Co-localization
M30	In vitro binding
M31	In vitro reconstitution of activity
M32	Flow cytometry
M33	Ligand binding
M34	Two-hybrid, cell system not specified
M35	Enzyme activity measurement
M36	DNA footprinting/interference
M37	Chromatin immunoprecipitation
M38	Filter binding assay
M39	One-hybrid - Cell system not specified
M40	Three-hybrid . Cell system not specified
M41	Osmium tetroxide modification
M42	Southwestern blot
M43	Bacterial two-hybrid assay
M44	BiFC (Bimolecular Fluorescence Complementation)
M45	Immunoprecipitation of small molecule complexes with proteins e.g. receptors; the small molecule within the complex could be identified by the following methods: TCL=Thin Layer Chromatography; GC-MS=Gas Chromatography-Mass Spectrometry; H-NMR=Proton Nuclear magnetic resonance; FRET; X-Ray structure analysis; SPA=Scintillaton Proximity Assay
M46	Competitive binding assay of small molecule binding; detection of small molecule binding to a receptor with e.g. Rhodamine-Maleimide Assay, PAN VERA protocol, other physico-chemical methods for small molecule identification, SPA=Scintillaton Proximity Assay

Table 2: Quality matrix

	A1-5	B1-3	B4	C1	D1	D2	E1	F1-2	X1	X2
M1	6	6	6	6	6	6	2	6	6	6
M2	6	6	6	6	6	5	5	5	5	6
M3	2	3	3	3	3	4	4	6	6	6
M4	1	2	2	2	2	3	4	5	5	6
M5	2	2	3	3	4	4	4	5	5	6
M6	3	3	3	3	3	3	3	4	4	6
M7	3	3	3	3	4	4	4	5	5	6
M8	3	3	4	4	4	5	5	5	5	6
M9	4	4	4	4	4	4	4	4	4	6
M10	2	2	3	3	3	4	5	5	5	6
M11	4	4	4	4	4	4	4	4	4	6
M12	3	3	3	3	3	4	4	4	4	6
M13	2	2	2	2	3	4	4	5	5	6
M14	3	3	3	3	3	4	5	5	5	6
M15	3	3	3	3	3	3	4	5	5	6
M16	3	3	3	3	3	4	4	5	5	6
M17	6	3	3	3	3	3	3	4	4	6
M18	2	2	2	2	2	3	3	3	3	6
M19	3	3	4	4	4	4	4	5	5	6
M20	2	6	6	6	6	6	6	6	6	6
M21	4	4	4	4	4	4	4	6	6	6
M22	6	6	6	6	1	2	6	6	6	6
M23	3	3	3	3	3	4	4	5	5	6
M24	3	3	3	3	3	4	4	5	5	6
M25	2	2	2	2	3	3	4	5	5	6
M26	4	4	4	4	5	6	6	6	6	6
M27	3	3	3	4	4	4	4	4	5	6
M28*	6	6	6	6	6	6	6	6	6	6
M29*	2	3	3	3	3	4	4	6	6	6
M30*	3	3	4	4	4	5	5	5	5	6
M31*	6	6	6	6	6	6	6	6	6	6
M32*	6	3	3	3	3	4	6	6	6	6
M33*	2	2	3	3	4	4	4	5	5	6
M34*	6	6	6	6	6	6	6	6	6	6
M35*	3	3	3	3	3	4	4	4	4	6
M36*	6	6	6	6	6	6	6	6	6	6
M37*	1	2	2	2	2	3	4	5	5	6
M38*	4	4	4	4	4	4	4	4	4	6
M39*	6	6	6	6	6	6	6	6	6	6
M40*	6	6	6	6	6	6	6	6	6	6
M41*	6	6	6	6	6	6	6	6	6	6
M42*	3	3	3	3	4	4	4	5	5	6
M43	6	6	6	6	6	6	6	6	6	6
M44	6	6	6	6	2	6	6	6	6	6
M45	6	6	6	6	6	6	6	6	6	6
M46	6	6	6	6	6	6	6	6	6	6

*Methods were added in the course of integration of reactions from Proteome HumanPSD.

Table 3: Quality scale: reliability value

Priority	Reliability
1	highest reliability
2	high reliability
3	moderate (average) reliability
4	modest reliability
5	low reliability
6	quality value not assigned

8.1. TRANSPATH Materials - definitions and examples

We have classified all possible materials used by authors into several major categories, A through F. Category X covers the cases where a lack of knowledge exists and the material could not be identified in detail.

Category A

The whole organism or parts of it (organs, tissues) were used as the material; organism could be of wild type, natural mutant, or genetically manipulated.

Category B

Primary cultures of cells, maybe transfected, but not permanent cell lines.

Category C

Permanent cell lines (permanent cultures of cell)

Category D

Recombinant proteins expressed in vertebrate or insect cells.

Category E

Recombinant proteins expressed in yeast.

Category F

Recombinant proteins expressed in bacterial cells or synthesized *in vitro*.

8.1.1. Category A

A1 Wild-type

Material A with a number always refers to a living organism.

So, in material A1 (wt) signaling is comparable to that in material B3 "primary cells", that have been put into tissue culture media with artificial additives and incubated for at least some hours. A1 materials are mainly used for isolation/purification of large amounts of proteins from whole organism or organs by either immunoprecipitation (M4) or biochemical approaches like gel filtration or anion exchange chromatography (M8, M23 e.g.), or microscopical analysis (M3) *in situ*, e.g.

Example: XN000033572
amphiphysin2(r) + FAK1(r) <=> amphiphysin2(r):FAK1(r)
rat brain: (A1, M4)

Example PMID:11826105
"...Immunoprecipitation from rat brain. S-SCAM was immunoprecipitated from the Triton X-100 extract

of rat crude synaptosomes as described previously (Hirao et al., 2000). The immunoprecipitates were immunoblotted with either the anti-S-SCAM or the anti-beta-catenin antibody..."

A3 Knock-out organisms or their parts (organs, tissues)

Example PMID: 12379224
"...Generation of the EphB6 knockout mice. J1 and RW4 embryonic stem (ES) cells were electroporated with the linearized targeting vector and selected with geneticin on embryonal fibroblast feeder cells generated from the embryos of the cholecystokinin-B receptor knockout mice [24].

Fetal thymus organ culture. Fetal thymus organ cultures (FTOC) were done as described previously [27]. Briefly, the thymic lobes of anembryonic day 15.5 (E15.5) were organ cultured on the surface of Nucleopore filters (0.8 lm pore size, Costar, MA) supported on collagen hemostatic sponges (Integra Life Sciences, NJ) in a 24-well plate containing 0.5 ml RPMI1640 medium supplemented with 10% fetal calf serum, 50 lM 2-mercaptoethanol, 100 U/ml penicillin, 100 lg/ml streptomycin, and 1_ non-essential amino acids (Invitrogen, Netherlands) for 5 days at 37°C under 5% CO2 in the air..."

A4 Transgenic animals

Example PMID: 10973986
"...For nuclear matrix preparations from brain, nuclei were isolated from DRPLA transgenic mice as described elsewhere (Schilling et al. 1999), and the nuclear matrix was prepared from the isolated nuclei following the same protocol as for transfected cells, except that the final wash with 2 M NaCl was omitted. The subnuclear fractions were immunoblotted for atrophin-1 and ETO/MTG8 with antibodies AP142 (Wood et al. 1998) and 2174 (Sacchi et al. 1998)..."

Example PMID: 11442350
"...Six HD transgenic mice at 5 months of age and age-matched nontransgenics (or six DRPLA mice at 12 months and age-matched nontransgenics), were tested.... Animals were killed after decapitation, their brains were immediately removed and bisected sagittally. In the analysis of 4-week-old animals, we homogenized entire hemibrains. We also analyzed the forebrains of older, severely affected, HD (at 6 months), and DRPLA (at 12 months) mice. Tissue was frozen rapidly on dry ice, and stored at 270°C until assayed. Frozen tissues were homogenized in 10-20 vol of chilled 50 mM sodium phosphate buffer, pH 7.0. An aliquot was then adjusted to 0.1 M sodium acetate, cooled, and centrifuged to sediment particulate matter prior to HPLC analysis. Monoamines were measured by HPLC with electrochemical detection, with sensors set at 150, 250, 350, and 500 mV (ESA Inc., Chelmsford, MA). Exactly 10 ml of tissue extract was injected onto a C18 reverse phase MD-150 column (ESA Inc.) and eluted at a flow of 0.6 ml/min for 20 min. The mobile phase consisted of 1.7 mM 1-tetraethylammonium and 8% acetonitrile in 75 mM sodium phosphate buffer, pH 2.9..."

8.1.2. Category B

Materials B1-4 are cells from a living organism, that are not treated too much. Materials are cells that have been purified from an organism and put into culture media, which is an artificial environment for cells that change the signaling events within the cells. They might be passaged or not. Most of these cells die relatively fast after cultivation, at least the number of passages is limited (about the stem cells I am not sure, may be they have to be shifted to permanent cell lines).

B1 Primary cultures made of embryonic cells

Example PMID: 11826105

"...Hippocampal neuron culture and hippocampal slice culture. Hippocampal neuron cultures were performed from embryonic day 18 embryos as described previously (Takeuchi et al., 1997; Goslin et al., 1998). Hippocampal slices were obtained from postnatal day 6 (P6) or P8 rats and cultured on Millicell CM culture plate inserts (Millipore, Bedford, MA) in Eagle's MEM containing 25% (v/v) HBSS, 6.5 gm/ l glucose, 100 U/ml penicillin, 100 μg/ml streptomycin, and 25% (v/v) horse serum at 32°C under 5% CO2..."

B3 Primary made of other cells (not embryonic, not stem cells)

Example PMID: 12097473
"...Neuronal staining. Cultured hippocampal neurons were fixed with 4% paraformaldehyde, permeabilized by methanol, and incubated with primary antibodies [KIF1Bα (1189; 3 μg/ml), PSD-95 (SM55; 3 μg/ml), SAP97 (B9591; 3 μg/ml), S-SCAM (3 μg/ml), and MAP2 (1:200)], followed by Cy3- or FITC-conjugated secondary antibodies. Neuronal images were captured by confocal laser scanning microscopy (LSM510; Zeiss, Oberkochen, Germany)..."

B4 Primary cultures, transfected. Recombinant cells/tissue/organisms.

Example PMID: 14764452
"...Cell Culture Pooled human umbilical venous endothelial cells (HUVECs) were purchased from CellSystems and cultured in endothelial basal medium (EBM; CellSystems) supplemented with hydrocortisone (1 μg/mL), bovine brain extract (3 μg/mL), gentamicin (50 μg/mL), amphotericin B (50 μg/mL), epidermal growth factor (10 μg/mL), and 10% fetal calf serum (FCS) until the third passage. After detachment with trypsin, cells ($4x10^5$ cells) were grown in 6-cm cell culture dishes for at least 18 hours. Figure 3. Homeobox A9 regulates EphB4 expression. A and B, HUVECs were transfected with HoxA9 antisense oligonucleotides or with HoxA9 siRNA for 24 hours, lysed, and EphB4 expression was detected by Western blot analysis HUVECs were transfected with a myc-tagged HoxA9 wild-type construct. Cells were lysed 20 or 40 hours after transfection and expression of EphB4 was detected by Western blot analysis..."

Example PMID: 11114890
"...Figure 4. ...(B) ...The natural CYP3A23 promoter DR3 (WT) or its mutant variants (M1 and IR6) were transfected into primary rat hepatocytes in the presence of expression vectors for CAR or SXR. Cells were subsequently mock treated or treated with indicated compounds. Results are shown as fold induction over solvent, and represent the averages and standard error from triplicate assays. The concentrations of compound were the same as in A..."

8.1.3. Categories C to F

Materials C1 + D1 are permanent, highly manipulated cells such as permanent cell lines. Permanent cell lines are immortalized and therefore transformed (many cancer cell lines) and differ in their regulation/signaling in comparison to primary cells of the same healthy tissue.

C1 Permanent tissue culture cells

Example PMID: 9214622
"...For morphological analysis, HT1080 and Swiss 3T3 cells were seeded onto glass coverslips at a density of 13105 cells per 35 mm dish, cultured overnight and fixed. sMDCK2 cells were seeded onto glass coverslips at a density of 13104 cells per 35 mm culture dish DMEM containing 10% FCS, and incubated for 16 h. The medium was then changed to DMEM without FCS, and the cells were incubated another 24 h. Serum-starved cells were stimulated with or without 10-7 M TPA (Sigma) for 15 min at 37°C, and fixed. For indirect immunofluorescence, cells were fixed in phosphatebuffered saline (PBS) containing 3.7% paraformaldehyde for 20 min at room temperature, and then permeabilized with 0.2% Triton X-100 in agarose gel containing 2.1% formaldehyde, and transferred to a Biodyne A filter (Pall BioSuport, NY). The filter was then hybridized with the PBS for 10 min..."

D1 Permanent transfected or infected mammalian cells (recombinant tissue culture cells)

Example PMID: 11826105
"...Pull-down assay. COS cells were transfected with various constructs ofbeta-catenin, and the extracts were prepared as described above. For each extract, 160 μl was incubated with either 200 pmol of GST-S-SCAM-10 containing PDZ4 and PDZ5 or 200 pmol of GST fixed on 14.5 μl of glutathione beads. After the beads were washed, the proteins on the beads were immunoblotted with the anti-Myc antibody..."

Materials D2, E1, F1 are proteins produced by recombinant expression.

D2 Recombinant proteins

D2 recombinant proteins produced from cDNAs incorporated in plasmids (e.g. Baculo virus vectors) in non-mammalian high eukaryotic cells, for example in insect cells, transfected/microinjected *Xenopus* oocytes.

Example TP database: XN000021776:
MAT1(h) + RAR-gamma1(h) <=> MAT1(h):RAR-gamma1(h), D2/M4

Example PMID: 10748061
"...Baculoviruses allowing the expression of single subunits of the TFIIH core (XPB, XPD, p62, p52, p44, and p34) and of CAK (cdk7, cyclin H, and MAT1) were as described (17). Baculovirus encoding for hRARgDAB as an His-tagged fusion protein was constructed in the pVL 1392 vector (Pharmingen). Baculovirus expressing hRARa1 as a flag fusion protein was constructed in the pSK278 vector (28)..."

E1 Recombinant proteins expressed in Yeast

Example PMID: 15265006
"...The oligonucleotides that encoded the last 7 amino acid residues and a stop codon of rat HCN1-4 were cloned into a bait plasmid, pAS2-1 (BD Biosciences Clontech). The pACT2 prey plasmid containing the full-length rat tamalin was co-transfected with the above bait plasmids into yeast Y190 cells (BD Biosciences Clontech) and ß-galactosidase reporter gene assays were performed as previously described (Kitano et al. 2002)..."

F1 Recombinant proteins produced in bacteria, e.g. *E. coli*, *Bacillus subtilis*

Example PMID: 15265006
"...GST and MAL fusion proteins were expressed in E. coli and purified by glutathione Sepharose 4B beads (Amersham Biosciences) and amylose resin (New England Biolabs), respectively. The purified proteins were dialysed against PBS at 4 °C. GST and MAL fusion proteins were immobilized on glutathione-Sepharose 4B beads and amylose resin, respectively, and incubated with supernatants..."

F2 *In vitro* translation, synthetic peptides

Example PMID: 12464607
"...In Vitro Translation Reaction-Atrophin-1 constructs were transcribed using T7 polymerase and then translated using the TNT system (Promega) in the presence of [35S]methionine. Translations were incubated with 10-50 ng of each of the caspases for 2 h in the following buffer: 20 mM PIPES, 100 mM NaCl, 1% CHAPS, 10% sucrose, 10 mM dithiothreitol, and 0.1 mM EDTA, pH 7.2, at 37 °C..."

Example PMID: 10973986
"...Antibody APG840 was raised against synthetic peptide DRPLA-425 corresponding to residues 425-439 of human atrophin-1 at Cocalico Biologicals Inc., Reamstown, PA. This peptide is 100% conserved between human and mouse and was previously used to produce antibody AP142 in rabbit (Wood et al., 1998)..."

8.1.4. Category X

X1 Recombinant proteins from unspecified cells

Material X1 is for cases where recombinant proteins have been used but the expressing cell system has not been specified in the reference.

X2 Unspecified material

For cases where no information about the material could be obtained from the reference. This one is used by the curators as last choice when all efforts to specify the material used have failed.

9. Statistics

9.1. General

General	Total Number of Entries	
Table	Release 2011.2	Release 2012.2
Reaction	210,074	302,718
Molecule, total	143,056	228,941
own annotation	141,894	227,777
imported*	1,162	1,164
Gene	40,959	74,502
Pathway	172	180
signal transduction	79	86
metabolic	93	94
Chain	1,381	1,436
metabolic chain	273	275
Reference, total	47,866	50,738
own annotation	46,385	49,257
imported*	1,481	1,481
Clickable Maps	100	105

*imported from SwissProt public Rel.36

9.2. Reactions

Reaction	Total Number of Entries	
Type	Release 2011.2	Release 2012.2
semantic	24,623	36,468
pathway step	5,346	5,590
metabolic	1,251	1,261
decomposition	182	188
molecular evidence	179,923	260,472

9.3. Molecules

Molecule	Total Number of Entries	
Type	Release 2011.2	Release 2012.2
orthofamily	1,536	1,544
family	1,124	1,207
orthogroup	11,696	19,762
isogroup	32,194	54,524
orthobasic	597	606
basic	31,549	70,609
orthocomplex	1,656	1,729
complex	44,388	53,013
orthofamily_mod	53	53
family_mod	601	661
orthogroup_mod	930	958
isogroup_mod	6,085	6,779
orthobasic_mod	21	26
basic_mod	1,946	2,237
orthocomplex_mod	912	1,014
complex_mod	2,984	3,188
miRNA basic	1,265	4,879
pre-miRNA basic	1,362	3,500
mRNA basic	637	1,116
group (XOR)	48	48
smallmolecule	1,023	1,035
other	398	400

Molecule	Total Number of Entries	
Species	Release 2011.2	Release 2012.2
human, *Homo sapiens*		
protein, isogroup	9,843	19,280
protein, basic	14,408	34,860
miRNA	338	1,929
pre-miRNA	341	1,527
mRNA	410	781
mouse, *Mus musculus*		
protein, isogroup	7,459	16,066
protein, basic	9,485	23,403
miRNA	267	1,157
pre-miRNA	271	741
mRNA	86	157
rat, *Rattus norvegicus*		
protein, isogroup	4,956	8,810
protein, basic	4,584	8,954
miRNA	237	682
pre-miRNA	231	408
mRNA	13	28
cattle, *Bos taurus*		
protein, isogroup	310	340
protein, basic	283	295
clawed frog; *Xenopus laevis*		
protein, isogroup	177	180
protein, basic	318	320
fruit fly, *Drosophila melanogaster*		
protein, isogroup	48	104

protein, basic	313	380
miRNA	78	430
pre-miRNA	78	240
mRNA	79	79
Caenorhabditis elegans		
protein, isogroup	18	40
protein, basic	170	197
miRNA	115	368
pre-miRNA	113	223
mRNA	17	17

Cross-links to external databases (DB)	Release 2011.2			Release 2012.2		
	Linked TRANSPATH entries	Linked DB entries	Links total number	Linked TRANSPATH entries	Linked DB entries	Links total number
Swiss-Prot	49,806	34,939	50,239	110,788	74,651	111,132
EMBL/GenBank/DDBJ	52,064	272,329	272,329	113,049	443,656	443,656
InterPro	10,342	1,316	29,525	9,881	1,312	28,205
BRENDA	8,876	924	8,876	9,715	1,023	9,715
Affymetrix	49,740	191,768	191,768	110,458	343,716	343,716
miRBase	2,583	2,581	2,585	8,319	8,319	8,321
PubChemCompound	990	990	990	1,003	1,003	1,003
HyperCLDB	289	234	289	289	234	289

Cross-links to BIOBASE databases (DB)	Release 2011.2	Release 2012.2
PROTEOME	51,084	55,011
TRANSFAC® Factor	12,491	13,933

9.4. Genes

Gene	Total Number of Entries	
External database hyperlinks	Release 2011.2	Release 2012.2
Affymetrix	198,655	344,444
EMBL/GenBank/DDBJ	279,831	442,019
Entrez Gene	24,926	47,904
UniGene	40,947	74,418
RefSeq	84,542	158,871
HGNC	11,473	18,108
OMIM	11,026	15,853
Ensembl	19,638	37,240

Cross-links to BIOBASE databases (DB)	Release 2011.2	Release 2012.2
PROTEOME	23,301	41,820
TRANSFAC® Gene, basic	32,534	55,692
TRANSFAC® Gene, orthogene	8,426	18,810

10. History

TRANSPATH® Report 3, 0025 (2012)

Entries that have been removed from TRANSPATH® 2012.1 since the previous release (2012.1).

Reactions

XN000005110	XN000151472	XN000263116	XN000284514
XN000010389	XN000156491	XN000264395	XN000284515
XN000010644	XN000156495	XN000264396	XN000284853
XN000014689	XN000156856	XN000264397	XN000285283
XN000020025	XN000161856	XN000264398	XN000285324
XN000020737	XN000162727	XN000264747	XN000286102
XN000021268	XN000171891	XN000266162	XN000286873
XN000021365	XN000172699	XN000266462	XN000286874
XN000034733	XN000172700	XN000266466	XN000286877
XN000043003	XN000173596	XN000268301	XN000286904
XN000043004	XN000173655	XN000268311	XN000286912
XN000059926	XN000174672	XN000269969	XN000286914
XN000062367	XN000176401	XN000270121	XN000286917
XN000062369	XN000177002	XN000270125	XN000286920
XN000062484	XN000179085	XN000270147	XN000286921
XN000062504	XN000193685	XN000270194	XN000286922
XN000062855	XN000194735	XN000270195	XN000287001
XN000063014	XN000196545	XN000270204	XN000287254
XN000063577	XN000197445	XN000270205	XN000287255
XN000064707	XN000198133	XN000270207	XN000287257
XN000065167	XN000198134	XN000270208	XN000287268
XN000083813	XN000219844	XN000270326	XN000287334
XN000101945	XN000220270	XN000270328	XN000287337
XN000101951	XN000227100	XN000270523	XN000287497
XN000103732	XN000233366	XN000270526	XN000287499
XN000105142	XN000236300	XN000270592	XN000287655
XN000105952	XN000236334	XN000270599	XN000287932
XN000106050	XN000244675	XN000272779	XN000287937
XN000106606	XN000251364	XN000272784	XN000288162
XN000109006	XN000251365	XN000272812	XN000288433
XN000110042	XN000251366	XN000272985	XN000288740
XN000110043	XN000251367	XN000273025	XN000288744
XN000112190	XN000251368	XN000273033	XN000288751
XN000112192	XN000252324	XN000273486	XN000288755
XN000118534	XN000252325	XN000273626	XN000288756
XN000119559	XN000255376	XN000273661	XN000288875
XN000119796	XN000255474	XN000274117	XN000288877
XN000121674	XN000255608	XN000274205	XN000289054
XN000121709	XN000260166	XN000276261	XN000289319
XN000127893	XN000260231	XN000276262	XN000289342
XN000130161	XN000260779	XN000276263	XN000289366
XN000136248	XN000260825	XN000276264	XN000289367
XN000140393	XN000262472	XN000277342	XN000289370
XN000140395	XN000262473	XN000280832	XN000289372
XN000143698	XN000262485	XN000280926	XN000289373
XN000143727	XN000262487	XN000280927	XN000289424
XN000145596	XN000262490	XN000281023	
XN000145971	XN000262515	XN000281479	
XN000146789	XN000263115	XN000283950	

Molecules

MO000032812	MO000170120	MO000184620	MO000197854
MO000080509	MO000170121	MO000186263	MO000197857
MO000083973	MO000170122	MO000188102	MO000197858
MO000094754	MO000170123	MO000193319	MO000197891
MO000096833	MO000170124	MO000193321	MO000197894
MO000097912	MO000176711	MO000193373	MO000197997

MO000131198	MO000176712	MO000193381	MO000198000
MO000137735	MO000177991	MO000193459	MO000198913
MO000137736	MO000177992	MO000193460	MO000198914
MO000138312	MO000177993	MO000196341	MO000198917
MO000139236	MO000177994	MO000196834	MO000199071
MO000140791	MO000179982	MO000196840	MO000199072
MO000143560	MO000180134	MO000197201	MO000199236
MO000155675	MO000180136	MO000197202	MO000199391
MO000156996	MO000180138	MO000197535	MO000199393
MO000156997	MO000180139	MO000197541	MO000199394
MO000158777	MO000180388	MO000197542	MO000199630
MO000163468	MO000180389	MO000197853	

Genes

G006311	G023232	G049751	G060826	G096358
G009850	G037592	G051246	G061437	G096367
G014292	G038506	G052506	G092920	G096369
G018385	G043512	G055256	G094079	