The FetchProt Corpus

Documentation and Annotation Guidelines

SICS Technical Report T2007:11
ISSN 1100-3154
ISRN:SICS-T--2007/11-SE
Authors: Kristofer Franzén & Daniel Oppenheimer
Contact: franzen@sics.se
Date: August 2007
Keywords: text corpus, training data, test data, annotation guidelines, BioNLP

The FetchProt Corpus

Documentation and Annotation Guidelines

Augusti 2007

1 Background

The FetchProt Corpus has been built within the FetchProt project[1] in order to work as training and test data for the FetchProt system. The project, a collaboration between Swedish Institute of Computer Science (SICS), Center for Genomics and Bioinformatics (CGB) at Karolinska Institutet and the company Metamatrix AB, receives partial funding from The Swedish Agency for Innovation Systems – VINNOVA, from November 2003 to November 2006. The aim is to build a system that automatically searches full text scientific articles and identifies proteins for which the function has been experimentally verified. The system is to aid in populating the EXProt database (Ursing et al., 2002) [2].

Since the scope of covering all protein functions is too big, we start by looking at tyrosine kinase as described in Gene Ontology[3] GO:000471 (see also GO:0004714, transmembrane receptor protein tyrosine kinase activity; GO:0004715, non-membrane spanning protein tyrosine kinase activity; GO:0004718, Janus kinase activity).

Experiments that verify the tyrosine kinase activity are listed in (p. 12). In a first instance the goal will be to identify proteins that have undergone one or more of these experiments.

2 Description of the corpus (version 1.0)

The corpus consists of published journal articles freely available on the internet. Most of the selected articles describe experiments on proteins verifying tyrosine kinase activity, but we have also included, as potential false positives, articles that do not have instances of such experiments.
The corpus consists of 200 full text scientific articles.

- 150 of the articles include descriptions of experiments that verify tyrosine kinase activity.
- 50 of the articles do not include such descriptions.
- Both sets contain articles from approximately the same set of journals (Table 1 p. 7.)
- All articles are freely available for anyone to download from the Internet.
- All articles are available on the internet both as PDF-files and in html-versions.
- 771 experiments and 771 results are tagged in the corpus.
- 273 wild types and 92 different mutations of 77 proteins with UNIPROT IDs and 48 proteins without UNIPROT IDs, are subject to experimental validation of tyrosine kinase activity.
- The sentences describing the experiments and their result are tagged, as well as the proteins involved in the experiments, their cell line, species, and mutant description.

2.1 Files

Each journal article is represented by 2 text files in the corpus:

[1] http://fetchprot.sics.se/
[2] http://nar.oupjournals.org/cgi/reprint/30/1/50.pdf
[3] http://www.geneontology.org/ontology/function.ontology

The template file is the xml-file containing a structured template filled with the information annotated for the corresponding article (see 2.3). This file is to be used as a key when evaluating models.
The annotated text file is a file with the content of the article in plain text interspersed with annotations (xml-style tags) surrounding specific semantic elements (see semantic definitions under Tags and their use below) in the text. This file is a reference to where in the files the text in the filled templates is to be found for every annotation.
The text files have been produced by saving the PDF-files as text using Adobe Acrobat Reader v. 6.0.0. This is not an ideal solution, since every way of converting PDF-files to text result in a loss of information. We found that this version of Acrobat Reader provided the best conversion, but it still resulted in corrupt text files. All Greek letters and most mathematical symbols have received incorrect representations and their identity cannot be retraced since they are not converted in a consistent manner. In this version of the corpus we have checked by hand *all text between tags* against the original files and transformed every character that was not converted correctly by Acrobat Reader into a corresponding numeric character reference code (cf. Table 5, page 13). The main purpose of the annotated text file is thus to function as a reference file for the corresponding template file, not to be the canonical text version of the PDF version of the article. In the next release we plan to complement the corpus with text versions derived from the html versions of the articles.
The journal articles were chosen by posing search queries against PubMed[4]. For a list of the queries used, see Table 2. For a list of documents in the corpus for each query c.f. Table 6.

The corpus is downloadable from the FetchProt homepage: http://fetchprot.sics.se/

2.2 The analysis

This is a narrative description of the analysis of the corpus. For a structured description, see 2.3 (Example template) and Figure 1 (Document Type Definition). The focus of the analysis of the corpus is the mentioning of proteins with an experimentally verified tyrosine kinase activity. This is mirrored in that the main entity in the analysis is the specific mutant or wild type of a *protein*. For each *protein*, the *function* slot(s) in the template file consists of *evidence* slot(s) consisting of *experiment* and *result* slots that are filled with the sentence instances from the article. For *experiment* and *result*, the sentence is the minimum level of analysis. The experiments belong to one of the categories listed in (p. 12). The *function*(s), the *evidence*(s), the *experiment*(s), the *result*(s) and the rest of the information about the protein are connected to the specific protein mutant and to each other via identity values in the tags (prid, fun_id, ev_id).
The templates also include information that is not explicit in the articles (and therefore not tagged in the annotated text files). Some of the information in the template file has been inferred from common knowledge and some has been derived from other external knowledge sources. Examples of the first case are when **<mutant_desc>** (mutant description) has been set to **"**wild type" based on the lack of specification of a mutant, and when the **<species>** has been established from the name of the **<cell_line>**. Examples of the latter are **<pubmed_id>, <url>, <wt_uniprot_id>** (wild type protein UniProt identity code), **<wt_ec_number>** (wild type protein EC number), and the **go_id** (Gene Ontology molecular function identity code) feature in the **<function>** tag.

[4] http://www.ncbi.nlm.nih.gov/entrez/query.fcgi

2.3 Template structure - Example template

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<!DOCTYPE doc SYSTEM "template01.dtd">
<doc>
        <authors>Masahiro Tsuda, Takashi Matozaki, Kaoru Fukunaga, Yohsuke Fujioka, Akira Imamoto,
        Tetsuya Noguchi, Toshiyuki Takada, Takuji Yamao, Hitoshi Takeda, Fukashi Ochi, Tadashi
        Yamamoto, and Masato Kasuga</authors>
        <title>Integrin-mediated Tyrosine Phosphorylation of SHPS-1 and Its Association with SHP-2
        ROLES OF Fak AND Src FAMILY KINASES</title>
        <journal>THE JOURNAL OF BIOLOGICAL CHEMISTRY</journal>
        <year>1998</year>
        <volume>273</volume>
        <pages>13223-13229</pages>
        <pubmed_id>9582366</pubmed_id>
        <url>http://www.jbc.org/cgi/reprint/273/21/13223</url>
        <protein prid="p1">
           <wt_uniprot_id prid="p1">Q9WUD9</wt_uniprot_id>
           <wt_ec_number prid="p1">2.7.1.112</wt_ec_number>
           <protein_name prid="p1">Src</protein_name>
           <organism prid="p1">
                 <species prid="p1">Rat-1</species>
                 <mutant_desc prid="p1">wild-type</mutant_desc>
                 <cell_type prid="p1">fibroblasts</cell_type>
           </organism>
           <function prid="p1" fun_id="f1" go_id="GO:0004715">
                 <evidence prid="p1" fun_id="f1" ev_id="e1" type="006">
                    <experiment prid="p1" fun_id="f1" ev_id="e1">We next examined whether Src or
                    Fak catalyze the tyrosine phosphorylation of SHPS-1 in vitro. Incubation of a GST
                    fusion protein containing the cytoplasmic domain of SHPS-1 (GST-SHPS-1-cyto) with
                    Src kinase immunoprecipitated from Rat-1 cells resulted in the phosphorylation of both
                    GST-SHPS-1-cyto (Fig. 5A) and Src (Fig. 5C).</experiment>
                    <result prid="p1" fun_id="f1" ev_id="e1">Incubation of a GST fusion protein
                    containing the cytoplasmic domain of SHPS-1 (GST-SHPS-1-cyto) with Src kinase
                    immunoprecipitated from Rat-1 cells resulted in the phosphorylation of both GST-
                    SHPS-1-cyto (Fig. 5A) and Src (Fig. 5C).</result>
                 </evidence>
           </function>
        </protein>
        <protein prid="p2">
           <wt_uniprot_id prid="p2">P34152</wt_uniprot_id>
           <wt_ec_number prid="p2">2.7.1.112</wt_ec_number>
           <protein_name prid="p2">Fak</protein_name>
           <organism prid="p2">
                 <species prid="p2">mouse</species>
                 <mutant_desc prid="p2">wild-type</mutant_desc>
           </organism>
           <function prid="p2" fun_id="f2" go_id="GO:0004715">
                 <evidence prid="p2" fun_id="f2" ev_id="e2" type="006">
                    <experiment prid="p2" fun_id="f2" ev_id="e2">In contrast, HA-tagged Fak
                    immunoprecipitated from transfected COS-7 cells with antibodies to HA catalyzed the
                    phosphorylation of poly(Glu, Tyr) (Fig. 5A) and Fak (Fig. 5B), but not that of GST-
                    SHPS-1-cyto (Fig. 5A).</experiment>
                    <result prid="p2" fun_id="f2" ev_id="e2">In contrast, HA-tagged Fak
                    immunoprecipitated from transfected COS-7 cells with antibodies to HA catalyzed the
                    phosphorylation of poly(Glu, Tyr) (Fig. 5A) and Fak (Fig. 5B), but not that of GST-
                    SHPS-1-cyto (Fig. 5A).</result>
                 </evidence>
           </function>
        </protein>
</doc>
```

Table 1 Journal names, abbreviations, and number of relevant and irrelevant articles for each journal.

Journal name	File name starts with …	Rel.	Irrel.
American Journal of Physiology – Renal Physiology	AmericanJournalOfPhysiology	1	
American Journal of Physiology – Cell Physiology	AJP-CellPhysiology	1	
Biochemical Journal	Biochem_J	1	1
Biology of Reproduction	BiologyOfReproduction	1	
Blood	Blood	5	4
Cancer Research	CancerResearch	1	
Cell Death and Differentiation	CellDeathAndDifferentiation	1	
Cell Growth & Differentiation	CellGrowth&Differentiation	4	1
Circulation Research	CirculationResearch	2	1
Clinical Cancer Research	ClinicalCancerResearch	1	1
Diabetes	Diabetes	1	1
Endocrinology	Endocrinology	4	1
European Journal of Biochemistry	EuropeanJournalOfBiochemistry	4	1
Genes & Development	Genes&Development	1	
Journal of Applied Physiology	JournalOfAppliedPhysiology	1	
The Journal of Biological Chemistry	JBC	78	24
Journal of Bacteriology	JBacteriology	1(108)	
The Journal of Cell Biology	JCB	2	1
The Journal of Clinical Investigation	JCI	2	1
The Journal of Clinical Endocrinology & Metabolism	JClinEndocrinology&Metabolism	2	
The Journal of Experimental Medicine	JournalOfExperimentalMedicine	1	
Journal of General Virology	JournalOfGeneralVirology	1	
Journal of Neurophysiology	JNeurophysiol		1
Molecular Biology of the Cell	MBC	2	
Molecular Cancer Research	MolecularCancerResearch	1	
Molecular and Cellular Biology	MCB	5	2
Molecular Endocrinology	MolecularEndocrinology	2	1
Proceedings of the National Academy of Sciences of the United States of America	PNAS	7	3
The American Journal of Pathology	TheAmericanJournalOfPathology	2	
The Journal of Immunology	TheJournalOfImmunology	12	4
The Journal of Neuroscience	TheJournalOfNeuroscience	1	1
The Journal of Pharmacology and Experimental Therapeutics	TheJPETherapeutics	2	1
	Sum	150	50

Table 2 Queries used to find relevant and irrelevant articles, number of included papers from each group.

Query	Relevant papers	Irrelevant papers
alk in vitro	2	1
csk in vitro	5	2
egfr phosphorylation in vitro	1	
fak1 in vitro phosphorylation	1	
fes in vitro	2	1
fgr in vitro	3	1
fyn in vitro	13	6
hck in vitro	5	
insulin receptor phosphorylation in vitro	11	
jak1 in vitro	3	1
jak2 in vitro	7	3
jak3 in vitro	1	4
kns1 in vitro	1	
lck in vitro	10	3
lyn in vitro	4	1
platelet derived growth factor receptor in vitro	4	
pyk2 in vitro phosphorylation	4	
src in vitro phosphorylate	5	
syk in vitro	8	2
tyk2 in vitro	1	
tyrosine kinase activity substrate km	9	6
tyrosine kinase direct substrate	2	9
tyrosine kinase km substrate	1	1
tyrosine kinase substrate in vitro direct	14	3
tyrosine kinase substrate vitro kinetics	6	
tyrosine kinase vitro substrate activity	2	
tyrosine kinase vitro substrate kcat	1	
tyrosine kinase vitro substrate novel	2	
tyrosine phosphorylation in vitro	2	
tyrosine phosphorylation substrate in vitro	11	4
txk in vitro		1
zap-70 in vitro	1	1
yes in vitro	1	
Unknown query	7	
Sum	150	50

3 Annotation Manual

The following instructions have guided the annotator in his work.

- The annotation scheme for the template files follows xml standard.
- The annotated text documents use xml tags, but are not valid xml documents.
- The text that fills the template is the ideal version of the text between tags in the annotated text file. For example, signs for affiliation are removed in the template, and text from figures and tables that show up in the middle of the text due to bad conversion, is not to be transferred to the template. The template file is the key with the correct answer to the information need.
- Tags must not be interlaced, i.e **<a>***hh***** should be **<a>***xx*****
- Under experiment and result
    - Each tagged area consists of at least one full sentences (including period, "."), but can consist of several consecutive sentences
    - The sentence(s) between experiment tags should state which experiment has been performed and, if possible, the expected achievement by doing this (e.g.: *In order to verify X we performed Y*)
    - Only sentences describing relevant and important information is to be tagged
    - A sentence can have more than one tag e.g.

      <experiment prid="p1" ev_id="e1" type="001"><result prid="p1" ev_id="e1">*In order to verify X we performed Y and found that* **<protein_name prid="p1">***protein Z***</protein_name>** *had function W.***</result></experiment>**
    - Consecutive sentences to be included in an experiment or result annotation should be inside the same tags
    - Sentences referring to experiments or results belonging to the same evidence should have the same id
- Only use papers that exist both in PDF and HTML format.

Table 3 Tags and their use.

Typ of tag	Mandatory in template	Tag if in article	Description
<doc>	*		Beginning and end of template. Only one template per article.
<authors>	*	*	All authors of the article as in the article. In the template, numbers or signs for affiliation are removed.
<title>	*	*	Full title of article. In the template, referencing numbers or signs are removed.
<journal>	*	*	Full journal name or abbreviation according to PubMed or as in paper.
<year>	*	*	Year of publication with 4 characters.
<volume>	*	*	Journal volume.
<pages>	*	*	Full page reference start page - end page, full page numbering (NOT 123-5).
<pubmed_id>	*		Pubmed id.
<url>	*		URL of full text article in pdf.
<protein prid="p1">	*		Beginning of description of protein 1 with identity feature (prid="p1")

<wt_uniprot_id prid=”p1”>			Reference to amino acid sequence database (UNIPROT) for the wild type of the protein. If species cannot be derived from text, or if there is no suitable UniProt id in the database, use UNKNOWN.
<wt_ec_number prid=”p1”>			EC number for wild type protein function.
<protein_name prid=“p1”>	*	*	Protein name as stated, and tagged either in evidence or the last instance before the evidence (closest to)
<organism prid=”p1”>	*		Beginning of description of origin of protein.
<species prid="p1">	*	*	Species origin of protein. Organism name as stated in the text, if possible scientific name, but can also be “human”, “murine”, “feline” if unambiguous, but not preferred. The instance closest before the protein name in the text should be tagged. If none before the protein name, the closest instance after should be tagged. If text does not mention species, use UNKNOWN.
<mutant_desc prid=”p1”>		*	Mutant description as it is stated in the text. If not mutant, “WT” or as described in text (e.g., “endogenous”). The instance closest before the protein name in the text should be tagged. If none before the protein name, the closest instance after should be tagged.
<cell_type prid=”p1”>		*	Cell type of origin of tagged protein. The instance closest before the protein name in the text should be tagged. If none before the protein name, the closest instance after should be tagged.
<cell_line prid=”p1”>		*	Cell line of origin of tagged protein. The instance closest before the protein name in the text should be tagged. If none before the protein name, the closest instance after should be tagged.
<function prid=”p1” fun_id=”f1” go_id=”GO:0004713**”>**	*		Beginning of description of function. Can have many evidences. Each **function** has a unique id – **fun_id** as well as a **go_id** identifying the molecular function according to Gene Ontology.
<evidence prid=”p1” ev_id=”e1” fun_id=”f1” type="005">	*		The **evidence** element consists of the **experiment** part and the **result** part. There can only be one experiment and one result per evidence. However, there can be more than one text instance, describing an experiment or a result, per evidence, then describing the same experiment or result. For an evidence annotation, only sentences stating which experiment is performed to verify a specific function is to be tagged. Evidence belonging to the same function should have the same **fun_id**. Each **evidence** has a unique id – **ev_id**. Attribute **type** refers to type of evidence according to list (p. 12).

<experiment prid="p1" fun_id="f1" ev_id="e1">	*	*	Describes the experiment in a sentence(s) like "In order to test X we performed Y". There can be more than one sentence per tag, but only full sentences. All sentences belonging to the same **evidence** should have the same **ev_id**.
<result prid="p1" fun_id="f1" ev_id="e1">	*	*	Describing the results from the above mentioned experiment. Can be more than one sentence. Only full sentences. All sentences belonging to the same **evidence** should have the same **ev_id**.

Table 4 Evidence types in the corpus. A number of experiments used in verifying tyrosine kinase activity.

Type number	Description of evidence
005	PHOSPHOAMINO ANALYSIS
006	KINASE/PHOSPHOTRANSFERASE
024	DENSITOMETRIC ANALYSIS
039	IMMUNOBLOTTING
041	PHOSPHOPEPTIDE MAPPING
042	MASS SPECTROMETRY

Figure 1. The template DTD

```
<!ELEMENT doc (authors,title,journal,year,volume,pages,pubmed_id,url,protein*)>
<!ELEMENT authors (#PCDATA)>
<!ELEMENT title (#PCDATA)>
<!ELEMENT journal (#PCDATA)>
<!ELEMENT year (#PCDATA)>
<!ELEMENT volume (#PCDATA)>
<!ELEMENT pages (#PCDATA)>
<!ELEMENT pubmed_id (#PCDATA)>
<!ELEMENT url (#PCDATA)>
<!ELEMENT protein (wt_uniprot_id, wt_ec_number, protein_name, organism, function+)>
<!ATTLIST protein prid ID #REQUIRED>

<!ELEMENT wt_uniprot_id (#PCDATA)>
<!ATTLIST wt_uniprot_id prid IDREF #REQUIRED>

<!ELEMENT wt_ec_number (#PCDATA)>
<!ATTLIST wt_ec_number prid IDREF #REQUIRED>

<!ELEMENT protein_name (#PCDATA)>
<!ATTLIST protein_name prid IDREF #REQUIRED>

<!ELEMENT organism (species, mutant_desc, cell_type?,cell_line?)>
<!ATTLIST organism prid IDREF #REQUIRED>

<!ELEMENT species (#PCDATA)>
<!ATTLIST species prid IDREF #REQUIRED>

<!ELEMENT mutant_desc (#PCDATA)>
<!ATTLIST mutant_desc prid IDREF #REQUIRED>

<!ELEMENT cell_type (#PCDATA)>
<!ATTLIST cell_type prid IDREF #REQUIRED>

<!ELEMENT cell_line (#PCDATA)>
<!ATTLIST cell_line prid IDREF #REQUIRED>

<!ELEMENT function (evidence+)>
<!ATTLIST function prid IDREF #REQUIRED>
<!ATTLIST function fun_id ID #REQUIRED>
<!ATTLIST function go_id ID #REQUIRED>

<!ELEMENT evidence (experiment+,result+)>
<!ATTLIST evidence prid IDREF #REQUIRED>
<!ATTLIST evidence fun_id IDREF #REQUIRED>
<!ATTLIST evidence ev_id ID #REQUIRED>
<!ATTLIST evidence type CDATA #REQUIRED>

<!ELEMENT experiment (#PCDATA)>
<!ATTLIST experiment prid IDREF #REQUIRED>
<!ATTLIST experiment fun_id IDREF #REQUIRED>
<!ATTLIST experiment ev_id IDREF #REQUIRED>

<!ELEMENT result (#PCDATA)>
<!ATTLIST result prid IDREF #REQUIRED>
<!ATTLIST result fun_id IDREF #REQUIRED>
<!ATTLIST result ev_id IDREF #REQUIRED>
```

Table 5 Numerical character references used.

Character reference	Corresponding character	Name	Unicode
&#34;	"	Quotation Mark	
&#38;	&	Ampersand	
&#60;	<	Less Than Sign	
&#61;	=	Equals Sign	
&#62;	>	Greater Than Sign	
&#126;	~	Tilde	
† (†)	†	Dagger	\u2020
‡ (‡)	‡	Double Dagger	\u2021
• (•)	•	Bullet	\u2022
&#176;	°	Degree Sign	\u00B0
&#177;	±	Plus Minus	\u00B1
&#916;	Δ	Capital Delta	\u0394
&#936;	Ψ	Capital Psi	\u03A8
&#945;	α	Small Alpha	\u03B1
&#946;	β	Small Beta	\u03B2
&#947;	γ	Small Gamma	\u03B3
&#948;	δ	Small Delta	\u03B4
&#949;	ε	Small Epsilon	\u03B5
&#950;	ζ	Small Zeta	\u03B6
&#952;	θ	Small Theta	\u03B8
&#954;	κ	Small Kappa	\u03BA
&#956;	μ	Small Mu	\u03BC
&#963;	σ	Small Sigma	\u03C3

Table 6 Relevant and irrelevant articles in the corpus for each PubMed query

Query	Relevant articles	Irrelevant articles
alk in vitro	Blood_96_1605_1607 TheAmericanJournalOfPathology_154_1657_1663	ClinicalCancerResearch_9_1121_1128
csk in vitro	JBC_272_1355_1362 JBC_273_10004_10010 JBC_274_2291_2297 JBC_274_5422_5428 JBC_275_29183_29186	JBC_279_5975_5983 JBC_271_14989_14994
egfr phosphorylation in vitro	JBC_278_37413_37418	
fak1 in vitro phosphorylation	MolecularEndocrinology_16_1612_1628	
fes in vitro	CellGrowth&Differentiation_11_581_592 MCB_19_8335_8343	Blood_103_912_920
fgr in vitro	JBC_275_41124_41132 TheJournalOfImmunology_172_5034_5040 AJP-CellPhysiology_281_C1385_C1395	PNAS_95_7580_7584
fyn in vitro	EuropeanJournalOfBiochemistry_271_2615_2623 Genes&Development_13_2400_2411 JBC_273_13223_13229 JBC_274_5522_5531 JBC_276_3879_3884 JBC_276_693_699 JBC_277_6775_6778 JCB_149_423_430 PNAS_96_2461_2466 TheJournalOfImmunology_162_7224_7232 TheJournalOfNeuroscience_21_3606_3618 TheJPETherapeutics_303_1325_1333 JBC_276_42389_42400	JBC_277_21561_21566 JBC_278_31574_31583. JBC_278_5894_5901 JCB_155_447_458 CirculationResearch_91_953_960 Jneurophysiol_79_137_142
hck in vitro	Blood_101_690_698 JBC_274_26579_26583 JBC_275_33353_33364 JournalOfGeneralVirology_85_721_729 JBC_276_25605_25611	
insulin receptor phosphorylation in vitro	JBC_277_10698_10703 JClinEndocrinology&Metabolism_89_4678_4684 JournalOfAppliedPhysiology_91_2240_2247 JBC_278_27896_27902 Endocrinology_140_55_62 JBC_273_26962_26968 JBC_273_31890_31900 TheJPETherapeutics_286_569_577 Endocrinology_139_520_526 Endocrinology_139_884_893	

Query	Relevant articles	Irrelevant articles
	JBC_271_7134_7140	
jak1 in vitro	JBC_273_25510_25515 JBC_274_20818_20825 JBC_274_9587_9599	Blood_101_4937_4943
jak2 in vitro	JBC_274_35492_35498 JBC_275_34719_34727 JBC_278_11970_11978 MCB_24_4955_4967 MolecularEndocrinology_17_2268_2282 TheJournalOfImmunology_164_4575_4585 TheJournalOfImmunology_165_2116_2123	JBC_279_44460_44466 MolecularEndocrinology_18_1471_1485 JBC_278_46171_46178
jak3 in vitro	TheJournalOfImmunology_160_2042_2045	TheJournalOfImmunology_166_3724_3732 TheJPETherapeutics_295_912_926 JBC_274_30266_30272 JBC_277_28830_28835
kns1 in vitro	JBC_271_27299_27303	
lck in vitro	CirculationResearch_85_12_22 EuropeanJournalOfBiochemistry_270_2369_2376 JBC_274_20056_20059 JBC_275_1685_1690 JBC_276_17455_17460 JBC_277_4999_5007 JournalOfExperimentalMedicine_193_497_507 PNAS_98_6587_6592 TheJournalOfImmunology_168_1978_1983 MolecularCancerResearch_1_155_163	JBC_279_2937_2944. Blood_103_1002_1010. JBC_278_34854_34863
lyn in vitro	Blood_98_3121_3127 JBC_276_33282_33290 TheJournalOfImmunology_167_6394_6402 TheJournalOfImmunology_163_1321_1326	PNAS_97_6687_6692
platelet derived growth factor receptor in vitro	Biochem_J_327_139_145 AmericanJournalOfPathology_163_277_286 JBC_277_15499_15506 JBC_279_19732_19738	
pyk2 in vitro phosphorylation	AmericanJournalOfPhysiology_275_F447_F451 JBC_274_29196_29201 JBC_274_31261_31271 JBC_277_45203_45210	

Query	Relevant articles	Irrelevant articles
src in vitro phosphorylate	Blood_91_3734_3745 JBC_273_17186_17191 JBC_274_6285_6294 TheJournalOfImmunology_160_1647_1658 MCB_24_2573_2583	
syk in vitro	CellGrowth&Differentiation_12_193_200 ClinicalCancerResearch_5_4264_4272 JBC_275_35442_35447 JBC_276_38595_38601 JBC_276_47982_47992 PNAS_96_11976_11981 TheJournalOfImmunology_165_1344_1351 TheJournalOfImmunology_172_4486_4492	PNAS_101_6158_6163 TheJournalOfImmunology_169_1028_1036
txk in vitro		JBC__277_755_762
tyk2 in vitro	TheJournalOfImmunology_171_2989_2994	
tyrosine kinase activity substrate km	JBacteriology_186_3472_3479 JBC_278_17328_17335 JBC_278_24072_24077 JBC_278_39323_39329 JCB_165_493_503 JClinEndocrinology&Metabolism_88_1323_1332 PNAS_100_14707_14712 PNAS_102_7109_7114 CellGrowth&Differentiation_12_379_386	Diabetes_54_361_366 JBC_279_55348_55354 JBC_279_52500_52516 Biochem_J_385_737_744 JBC_279_44039_44045 EuropeanJournalOfBiochemistry_270_3891_3903
tyrosine kinase direct substrate	PNAS_100_13298_13302 EuropeanJournalOfBiochemistry_268_4158_4168	Blood_102_659_661 Endocrinology_144_2988_2996 JBC_274_27610_27616 JBC_276_42932_42937 JBC_279_33759_33767 JBC_279_49523_49532 MCB_23_6255_6266 MCB_23_7510_7524 TheJournalOfImmunology_172_331_339
tyrosine kinase km substrate	MCB_24_10558_10572	CellGrowth&Differentiation_10_805_812

Query	Relevant articles	Irrelevant articles
tyrosine kinase substrate in vitro direct	Endocrinology_141_621_628 JBC_273_16756_16763 JBC_274_19003_19010 JBC_276_17339_17346 JBC_276_28006_28013 JBC_276_42843_42850 JBC_277_21537_21541 JBC_277_9422_9428 JCI_101_2751_2760 JCI_99_1831_1841 MCB_21_7641_7652 PNAS_94_2295_2300 JBC_274_22151_22154 JBC_278_13740_13746	JBC_278_34854_34863 JBC_277_13620_13627 JBC_276_37747_37753
tyrosine kinase substrate vitro kinetics	JBC_270_29773_29780 JBC_270_30829_30836 JBC_271_15753_15761 JBC_271_4755_4762 JBC_271_7465_7472 JBC_274_7557_7564	
tyrosine kinase vitro substrate activity	BiologyOfRreproduction_67_301_307 MBC_14_1448_1459	
tyrosine kinase vitro substrate kcat	JBC_270_27112_27115	
tyrosine kinase vitro substrate novel	Diabetes_50_824_830 JBC_278_47610_47621	
tyrosine phosphorylation in vitro	JBC_279_55827_55832 CellGrowth&Differentiation_10_387_396	
tyrosine phosphorylation substrate in vitro	CancerResearch_61_8887_8895 CellDeathAndDifferentiation_11_290_300 EuropeanJournalOfBiochemistry_266_1166_1173 JBC_274_33123_33130 JBC_275_41439_41446 JBC_276_16885_16893 JBC_276_47966_47974 JBC_277_25233_25238 JBC_278_5163_5171 JBC_279_14819_14827 CirculationResearch_88_513_519	JBC_2004 Nov 12;279(46):47572-9 JBC_2004 Jun 11;279(24):25755-64 TheJournalOfImmunology_172_2803_2810. CancerResearch_63_7777_7784

Query	Relevant articles	Irrelevant articles
yes in vitro	MBC_11_51_64	
zap-70 in vitro	Blood_101_3534_3542	JCI_115_2287_2295