eFIP^{online} Help Document

University of Delaware
Computer and Information Sciences &
Center for Bioinformatics and Computational Biology
Newark, DE, USA
December 2013

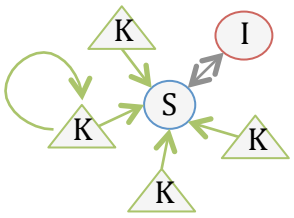


Table of Contents

URL: http://research.bioinformatics.udel.edu/eFIPonline

Introduction

eFIP^{online} is a web service extending two text-mining tools: *RLIMS-P* [1,2,5-7] for the extraction of phosphorylation information, and *eFIP* [3,4] for the detection of the impact of phosphorylation on protein-protein interaction (PPI). **eFIP^{online}** works on both Medline abstracts and PMC full-length articles. It displays information about protein kinases, substrates, and interacting partners. Results are pre-processed and updated on a quarterly basis. These can be queried by searching for proteins in any of the three forms (kinase, substrate, or interacting partner), searching directly for PMIDs/PMCIDs, or by using a PubMed query.

Architecture

eFIP^{online} consists of four main parts: (1) indexing and retrieval of Medline abstracts and PMC full-length articles; (2) extraction of phosphorylation information by the RLIMS-P system; (3) extraction of PPI information concerning phosphorylated proteins, as well as phosphorylation-PPI causal/temporal relations, by the eFIP system; (4) pre-processing of all abstracts and full-length articles and storing the results in the database; and (5) displaying the results online. The pipeline is shown in *Figure 1*.

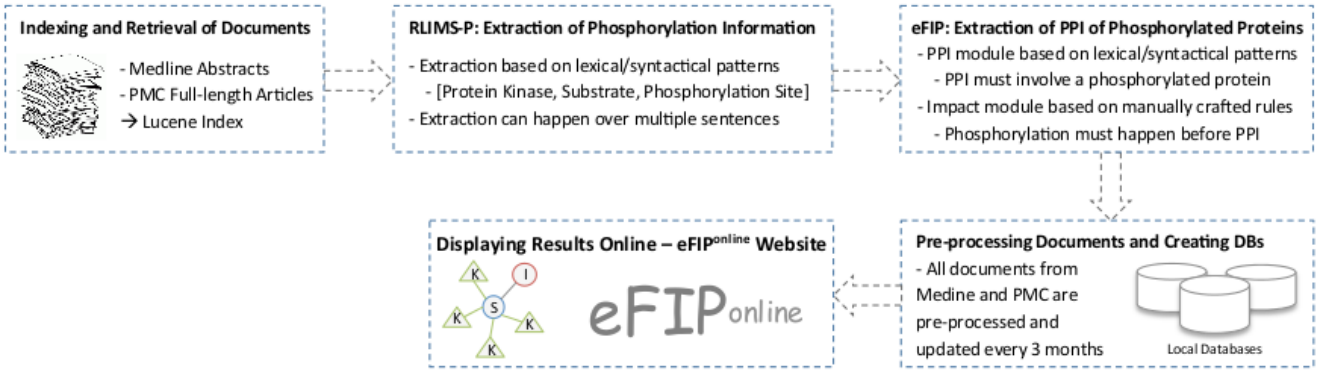


Figure 1. Pipeline of eFIP^{online}

Interface

The website consists of four main pages: (1) the Search Page (or Homepage); (2) the Summary Page (or Results Page); (3) the Text Evidence Page (or Document Page); and (4) the Login Screen facilitating validation of results. We will explain each page in the following sections.

Search Page (Homepage)

The website allows for various search criteria, as can be seen in *Figure 2*.

1. Keywords

The keywords or phrases can be combined using Boolean operators. This input is used to query PubMed documents, and resulting documents are then intersected with the results stored in our local databases. An example query is “apoptosis AND Bax AND phosphorylation”.

2. Protein Names

A protein can also be searched for, together with its type (kinase, substrate, interacting partner, or all of these types). Because a protein is known by many names, we allow for the search of all synonyms, delimited by the Boolean operator OR. An example query is "Bax OR BCL2-associated X protein OR BCL2L4". Note that the user can also choose to search for multiple proteins this way.

3. List of PMIDs/PMCIDs

The user can also choose to provide a list of PMIDs or PMCIDs, delimited by comma, space, or new line. Users can enter up to 200 IDs per run.

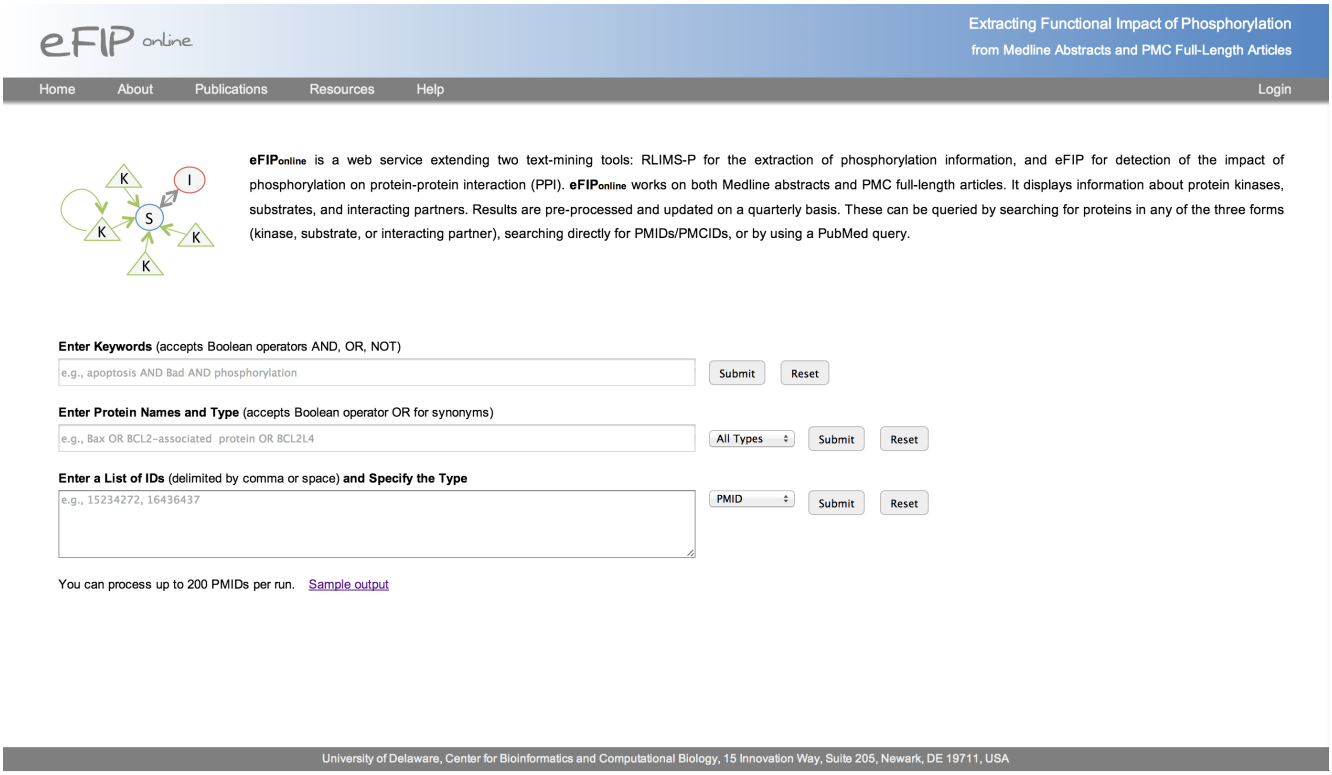


Figure 2. Search Page (Homepage): a search can be conducted using keywords, protein names, or list of PMIDs/PMCIDs.

Summary Page (Results Page)

After the user inputs his/her query in the Search Page, the results are displayed in the Summary Page, as can be seen in *Figure 3*.

1. Statistics

At the top of the Summary Page, the user can see statistics and download the list of relevant document IDs. The user is shown the number of documents with potential phosphorylation events (i.e., mentioning a phosphorylation trigger word), the number of documents that contain phosphorylation events (i.e., documents that were identified by RLIMS-P to be positive), the

number of kinases extracted from these documents, as well as the number of substrates, sites, interacting proteins, and sentences containing impact information.

2. Summary Table

Below the statistics information, the user is shown a summary table containing the extracted information. The data can be saved, by clicking the "Save Table" button. Clicking on the Download button will create a comma-delimited file containing the Document ID, kinase, substrate, site, interacting protein, and phosphorylation-PPI relation information, together with the associated evidence sentences. The order of the information in the file will vary depending on the view from which it was downloaded.

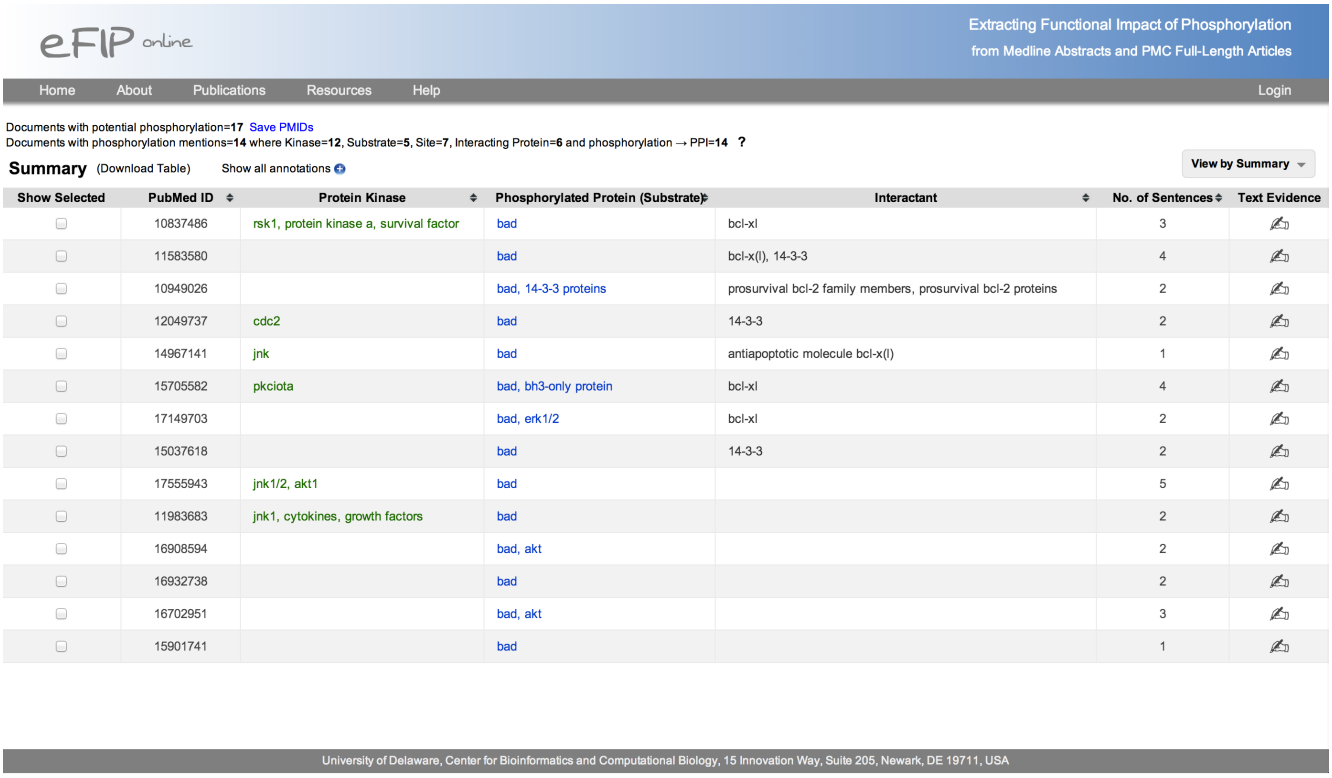


Figure 3. Summary Page (Results Page) showing statistics about the results and a table summarizing the information extracted from the documents relevant to the query search.

Different Views

The table can be customized with different "views" that can be chosen from the drop-down list under the "View by Summary": View by Kinase, View by Substrate, View by Interactant, View by Document ID. All the rows corresponding to the same chosen entity will be displayed together, as shown in *Figure 4* for "View by Interactant".

Show Selected	Interactant	PubMed ID	Phosphorylated Protein (Substrate)	Protein Kinase	Phosphorylation Site	Impact	No. of Sentences	Text Evidence
	bcl-xl	10837486	survival factor	bad	Ser-155	inhibits association	1	
		10837486	protein kinase a	bad	Ser-155	inhibits association	1	
		10837486	rsk1	bad	Ser-155	inhibits association	1	
		17149703		bad	Ser-112	inhibits association	1	
		10837486		bad	Ser-155	has some effect association	1	
		15705582		bad		inhibits association	1	
	14-3-3	12049737		bad	Ser-128	inhibits association	1	
		11583580 ✓		bad	Ser-112	enables association	1	
		15037618		bad		enables association	1	

Figure 4. Summary Page – View by Interactant: showing the results grouped by the Interactant protein.

Columns

Each column in the table can be sorted based on ascending or descending, numerical or alphabetical order, by clicking the arrows next to the column headings. Moreover, the data can be displayed condensed into as few rows as possible (by grouping partially redundant information), or expanded into as many rows as different information is available. This can be done by clicking the +/- signs next to "Show all annotations".

- The first column (Show Selected) allows the user to select which annotation lines to display. To make the selection, the user will have to click on the column name after checking the boxes.
- The second column (Document IDs) displays the PMID and/or PMCID, which links to their corresponding page on the PubMed website.
- The next three columns (Protein Kinase, Substrate, Site) list the phosphorylation information that was extracted by the RLIMS-P system.
- The next two columns (Impact and Interactant) list the PPI information and its relationship to the phosphorylation information (as extracted by eFIP). The combined extracted information from RLIMS-P and eFIP that is shown in one row can be read as follows: "*Protein Kinase* phosphorylation of *Substrate* at *Site Impacts* the association/dissociation of the *Substrate* and the *Interactant*". The user will notice that some of the rows contain only phosphorylation information, and no PPI information. This is because we always display the phosphorylation information, regardless of whether there is a PPI associated with it. The opposite is not true, as we require that the PPI information contain a phosphorylated protein.
- The eighth column (No. of Sentences) shows the number of sentences containing the evidence for the phosphorylation, and the PPI relationship where applicable.
- To last column (Text Evidence) takes the user to the sentences containing the evidence, as well as the entire document from which these sentences were extracted.

Text Evidence Page (Document Page)

Users can choose to see the sentences containing the evidence for a given row of extracted information from the Summary page. Clicking the icon in the Text Evidence column for that row brings the user to the Text Evidence Page of the respective document (see *Figure 5*). The "Back to search" link takes the user back to the Summary Page of results obtained for the initial query.

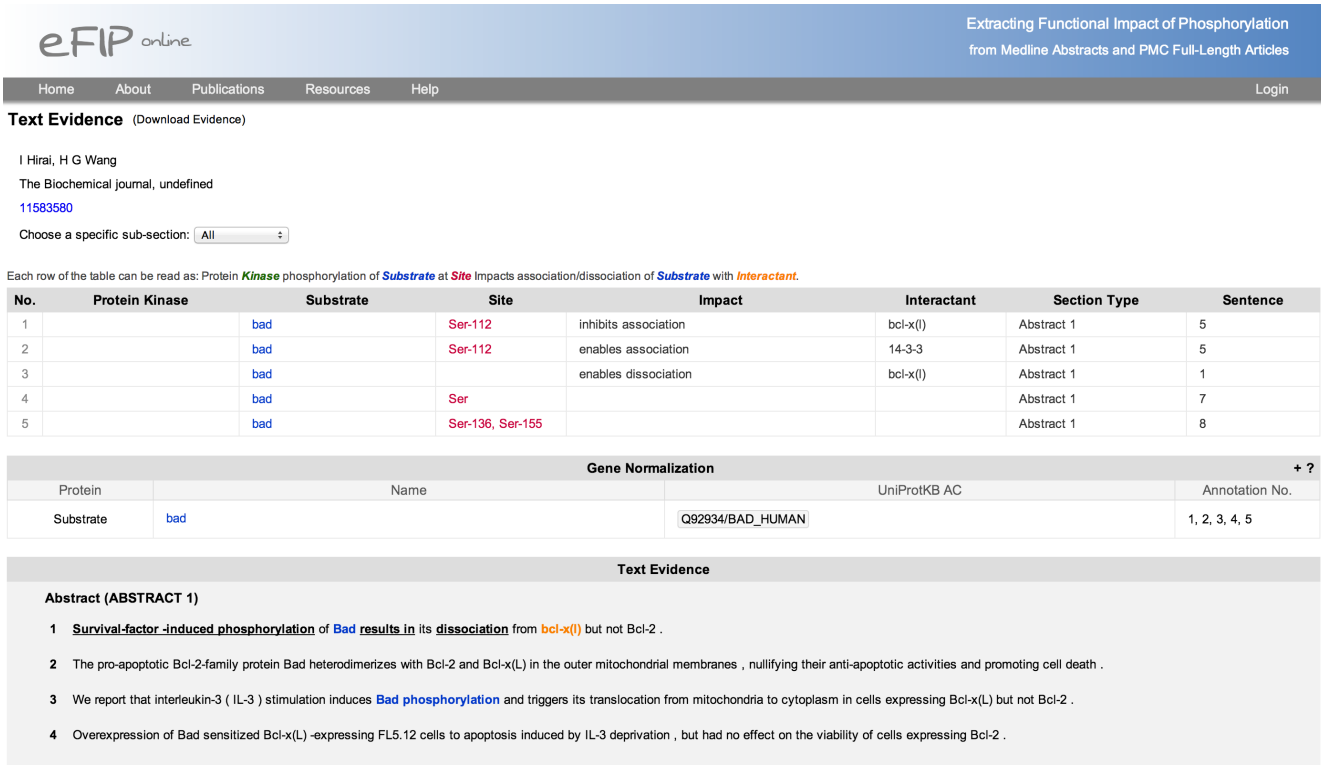


Figure 5. Text Evidence Page (Document Page) showing basic information about the document, a table summarizing the information extracted from this document, and sentences with highlighted evidence.

1. Basic Information

The title, the authors, the journal, and the ID of the document are shown at the top of the Text Evidence Page.

2. Summary Table

A table summarizing the information extracted from this document is shown next. This table is similar to the table shown in the Summary Page. Clicking on any row in the table will highlight the respective sentence(s) in the text from where the information was extracted.

3. Gene Normalization

Below the table, the user can see the kinases, substrates, and interacting partners that were successfully normalized to the UniProtKB. The user can choose to visit the UniProtKB entries by clicking on the name of the protein.

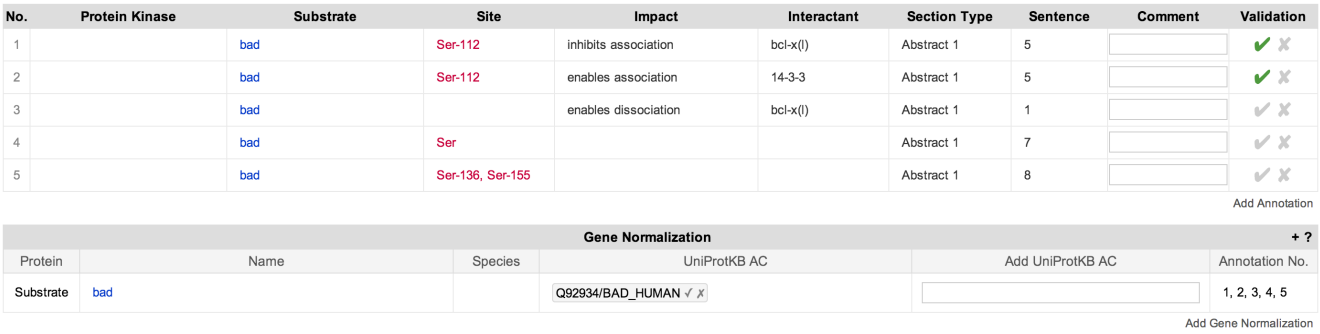
4. Text Evidence

The sentences of the document are displayed next, in the order in which they appear in the document. Kinases are marked in green, substrates in blue, sites in red, and interacting partners in dark orange. Users can customize the highlighting of the information using the check boxes provided at the bottom of the page. Because the document can be quite large, only the sentences

from which information was extracted are shown by default. However, the user can choose to reveal the other sentences, in order to get a better understanding of the facts described in the paper.

Login Screen and Validation of Results

To validate the results and provide additional annotations, users need to log in. In order to log in for the first time, users need to sign up by entering their e-mail, name, and affiliation. Once logged in, the user will be able to see their logged in status and begin validating the results or adding annotations that were missed by the extraction tools. The validation happens in the Text Evidence Page. *Figure 6* shows an example.



No.	Protein Kinase	Substrate	Site	Impact	Interactant	Section Type	Sentence	Comment	Validation
1		bad	Ser-112	inhibits association	bcl-x(l)	Abstract 1	5		✔ ✘
2		bad	Ser-112	enables association	14-3-3	Abstract 1	5		✔ ✘
3		bad		enables dissociation	bcl-x(l)	Abstract 1	1		✔ ✘
4		bad	Ser			Abstract 1	7		✔ ✘
5		bad	Ser-136, Ser-155			Abstract 1	8		✔ ✘

Add Annotation

Gene Normalization + ?

Protein	Name	Species	UniProtKB AC	Add UniProtKB AC	Annotation No.
Substrate	bad		Q92934/BAD_HUMAN ✓ ✗		1, 2, 3, 4, 5

Add Gene Normalization

Figure 6. Validation of Results: once logged in, a user can validate the results and add evidence and gene normalization.

References

1. Torii M, Arighi CN, Wang Q, Wu CH, Vijay-Shanker K. *Text Mining of Protein Phosphorylation Information Using a Generalizable Rule-Based Approach*. In Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedical Informatics (ACM-BCB) 2013:201—210.
2. Torii M, Li G, Li Z, Celen I, Diella F, Oughtred R, Arichi C, Huang H, Vijay-Shanker K, Wu CH. *RLIMS-P: Literature-based curation of protein phosphorylation information.* In Proceedings of BioCreative IV Workshop, vol. 1:247—253.
3. Tudor CO, Arighi CN, Wang Q, Wu CH, Vijay-Shanker K. *The eFIP system for text mining of protein interaction networks of phosphorylated proteins.* Database 2012.
4. Arighi CN, Siu AY, Tudor CO, Nchoutmboube JA, Wu CH, Vijay-Shanker K. *eFIP: A Tool for Mining Functional Impact of Phosphorylation from Literature.* Bioinformatics for Comparative Proteomics Methods in Molecular Biology, 2011, vol. 694:63—75.
5. Yuan X, Zhang-Zhi Hu, Wu HT, Torii M, Narayanaswamy M, Ravikumar KE, Vijay-Shanker K, Wu CH. *An online literature mining tool for protein phosphorylation*. Bioinformatics 2006, 22(13):1668—1669.
6. Hu ZZ, Narayanaswamy M, Ravikumar KE, Vijay-Shanker K, Wu, CH. *Literature mining and database annotation of protein phosphorylation using a rule-based system*. Bioinformatics 2005, 21(11):2759—2765.
7. Narayanaswamy M, Ravikumar KE, Vijay-Shanker K. *Beyond the clause: extraction of phosphorylation information from medline abstracts*. Bioinformatics 21.suppl 1 (2005): i319-i327.