

Human Protein Reference Database

# Human Protein reference Database

A discovery resource for proteomics

W. M. Wong

ISILS 4-11-2004

2

## Overview

- The Human Protein reference database (HPRD)
- Representation of protein data
- Annotation
- An object-oriented database architecture
- Data standardization
- Visualization methods
- Future developments


3

## The Human Protein Reference Database (HPRD)

- Comprehensive protein information resource
- Object-oriented database
- Web based
- Freely available at: <http://www.hprd.org>

4

## HPRD Homepage



Human Protein Reference Database

You are # status

Home Search About Pathways FAQs Register as a user Account Help/Feedback

**Statistics**

Proteins	1032
Protein-Protein Interactions	22514
Disorders	438
PTMs	1429
PubMed Links	119878

**Journal article**

**COMMERCIAL ENTITIES MAY NOT USE THIS SITE WITHOUT PRIOR LICENSING AUTHORIZATION. PLEASE SEND AN EMAIL FOR FURTHER INFORMATION ABOUT LICENSING.**

The Human Protein Reference Database represents a centralized platform to visually depict and integrate information pertaining to domain architecture, post-translational modifications, interaction networks and disease association for each protein in the human proteome. All the information in HPRD has been manually extracted from the literature by expert biologists who track, interpret and analyze the published data. HPRD has been created using an object-oriented database in C++, an open source web application server, that provides versatile in query functions and allows data to be displayed dynamically.

Please cite the following reference for this database  
Pan, S. et al. (2003) Development of human protein reference database as an initial platform for approaching systems biology in humans. *Genome Research*. 13:2363-2371.

Please send any questions or comments about the Human Protein Reference Database to [hspr@biocristalica.org](mailto:hspr@biocristalica.org)

Copyright © 2003-04 Johns Hopkins University and the Institute of Biocristalica

This is a joint project between:

PandeyLab and Institute of Biocristalica

## Query Screen

5

## Representation of Protein Data

6

## Sequence

7

## Interactions

8

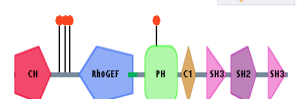
## External Links

9

You are at Home » Proteins » VAV1

**VAV1**

Molecular Class: Guanine nucleotide exchange factor  
 Molecular Function: Guanyl-nucleotide exchange factor activity  
 Biological Process: Signal transduction; Cell communication



ALTERNATE NAMES | DISEASES | PTMs & SUBSTRATES | INTERACTIONS | EXTERNAL LINKS

OMIM: 184875 | Swiss Prot: P15488 | Unigene: 118237 | Locus Link: 7409 | PDB: | Other Resources: |

Credits | Comments

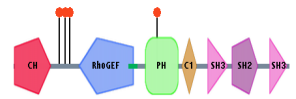
## Alternate Names

10

You are at Home » Proteins » VAV1

**VAV1**

Molecular Class: Guanine nucleotide exchange factor  
 Molecular Function: Guanyl-nucleotide exchange factor activity  
 Biological Process: Signal transduction; Cell communication



SUMMARY | SEQUENCE | INTERACTIONS | PTMs & SUBSTRATES | EXTERNAL LINKS

ALTERNATE NAMES

VAV1 oncogene  
 Oncogene VAV

Credits | Comments

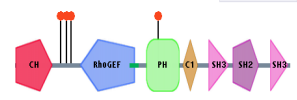
## Diseases

11

You are at Home » Proteins » VAV1

**VAV1**

Molecular Class: Guanine nucleotide exchange factor  
 Molecular Function: Guanyl-nucleotide exchange factor activity  
 Biological Process: Signal transduction; Cell communication



SUMMARY | SEQUENCE | INTERACTIONS | PTMs & SUBSTRATES | EXTERNAL LINKS

ALTERNATE NAMES | DISEASES

OMIM: 184875

DISEASES

Credits | Comments

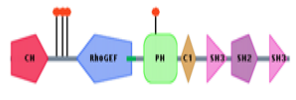
## PTM & Substrates

12

You are at Home » Proteins » VAV1

**VAV1**

Molecular Class: Guanine nucleotide exchange factor  
 Molecular Function: Guanyl-nucleotide exchange factor activity  
 Biological Process: Signal transduction; Cell communication



SUMMARY | SEQUENCE | INTERACTIONS | PTMs & SUBSTRATES | EXTERNAL LINKS

PTMs

Residue	Type	Site	Upstream Enzymes
Y	Phosphorylation	152	Lck
Y	Phosphorylation	150	Lck
Y	Phosphorylation	174	Lck
Y	Phosphorylation	411	PKA

Substrates

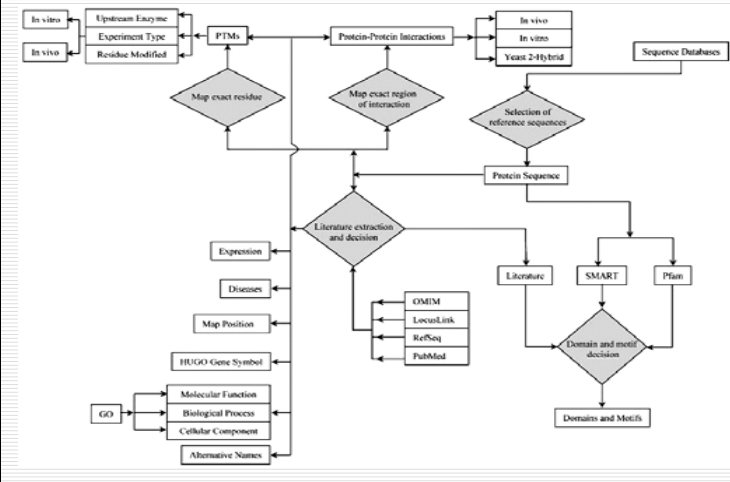
Title	Residue	Type	Site

Credits | Comments

## Annotation

- Manually
    - Reading the published literature
    - Bioinformatics analyses of the protein sequences, e.g. using BLAST
- 

## Annotation Schedule



## An object-oriented database architecture

- Zope – web application server
    - Zope Object Database
      - Transparently stores persistent objects
    - Zcatalog
      - Powerful indexing and searching
  - Python – programming language
- 

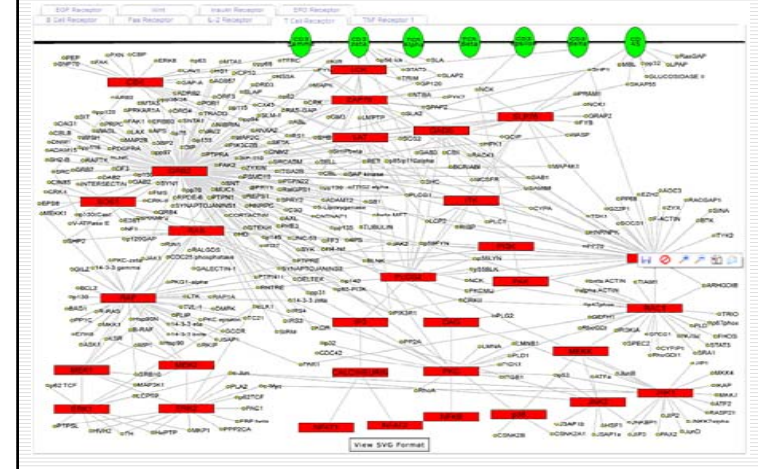
## Data standardization

- Compliant with GO vocabulary
    - Used by other annotation projects
  - XML (eXtensible Markup language)
    - Adopted by the micro array community
    - Easy import and export to other database systems
  - HUGO approved Gene Symbols
    - Easy linking to other databases
-

## Visualization methods

- ★ Pajek
    - ★ Generates protein interaction network pathway diagrams
  - ★ Formats:
    - ★ Jpeg
    - ★ Scalable Vector Graphics (SVG)
      - ★ Zooming without loss of resolution
      - ★ Search capability
      - ★ Link to molecule page of any protein in the network
- 

## Scalable Vector Graphics (SVG)



## Future developments

- ★ Federated Databases
    - ★ Better processing of information
    - ★ Enrich the database
    - ★ Minimizes errors
  - ★ Link with third party annotations
  - ★ Evolution
    - ★ Increase of number of entries
    - ★ Increase of depth of annotation of entries
-