

Proceedings

Open Access

## PutidaNET: Interactome database service and network analysis of *Pseudomonas putida* KT2440

Seong-Jin Park<sup>†1</sup>, Jong-Soon Choi<sup>†2,3</sup>, Byoung-Chul Kim<sup>1</sup>,  
Seong-Woong Jho<sup>1</sup>, Jea-Woon Ryu<sup>1</sup>, Daeui Park<sup>1</sup>, Kyung-A Lee<sup>1</sup>,  
Jong Bhak<sup>\*1</sup> and Seung Il Kim<sup>\*2</sup>

Addresses: <sup>1</sup>Korean BioInformation Center (KOBIC), KRIBB, Daejeon 305-806, Korea, <sup>2</sup>Proteomics Research Team, Korea Basic Science Institute, Daejeon 305-333, Korea and <sup>3</sup>Graduate School of Analytical Science and Technology, Chungnam National University, Daejeon 305-764, Korea

E-mail: Seong-Jin Park - psj420@kribb.re.kr; Jong-Soon Choi - jschoi@kbsi.re.kr; Byoung-Chul Kim - bckim@kribb.re.kr; Seong-Woong Jho - sucho@kribb.re.kr; Jea-Woon Ryu - jwryu@kribb.re.kr; Daeui Park - daeui@kribb.re.kr; Kyung-A Lee - kalee@kribb.re.kr; Jong Bhak\* - jong@kribb.re.kr; Seung Il Kim\* - ksi@kbsi.re.kr

\*Corresponding author †Equal contributors

from Asia Pacific Bioinformatics Network (APBioNet) Eighth International Conference on Bioinformatics (InCoB2009) Singapore 7-11 September 2009

Published: 3 December 2009

BMC Genomics 2009, 10(Suppl 3):S18 doi: 10.1186/1471-2164-10-S3-S18

This article is available from: <http://www.biomedcentral.com/1471-2164/10/S3/S18>

© 2009 Park et al; licensee BioMed Central Ltd.

This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** *Pseudomonas putida* KT2440 (*P. putida* KT2440) is a highly versatile saprophytic soil bacterium. It is a certified bio-safety host for transferring foreign genes. Therefore, the bacterium is used as a model organism for genetic and physiological studies and for the development of biotechnological applications. In order to provide a more systematic application of the organism, we have constructed a protein-protein interaction (PPI) network analysis system of *P. putida* KT2440.

**Results:** PutidaNET is a comprehensive interaction database and server of *P. putida* KT2440 which is generated from three protein-protein interaction (PPI) methods. We used PSIMAP (Protein Structural Interactome MAP), PEIMAP (Protein Experimental Interactome MAP), and Domain-domain interactions using iPfam. PutidaNET contains 3,254 proteins, and 82,019 possible interactions consisting of 61,011 (PSIMAP), 4,293 (PEIMAP), and 30,043 (iPfam) interaction pairs except for self interaction. Also, we performed a case study by integrating a protein interaction network and experimental I-DE/MS-MS analysis data *P. putida*. We found that 1) major functional modules are involved in various metabolic pathways and ribosomes, and 2) existing PPI sub-networks that are specific to succinate or benzoate metabolism are not in the center as predicted.

**Conclusion:** We introduce the PutidaNET which provides predicted interaction partners and functional analyses such as physicochemical properties, KEGG pathway assignment, and Gene Ontology mapping of *P. putida* KT2440. PutidaNET is freely available at <http://sequenceome.kobic.kr/PutidaNET>.

## Background

*P. putida* KT2440 is a ubiquitous bacterium which can break down a variety of organic materials for food. Because of its versatile metabolic activities, *P. putida* KT2440 is thought to play a pivotal role in the recycling of organic wastes and the degrading of biogenic and xenobiotic pollutants in the environment [1,2]. According to various carbon sources, we want to know the difference of networks according to the substrates. To simplify the culture condition, we selected succinate and benzoate as a sole carbon source. The easy carbon-utilization source, succinate and the required biochemical degradation-requiring benzoate were chosen for the comparison of a network analysis combined with the different proteomic data. An interactome of a species provides important clues about how to interpret metabolic pathways of constituent enzymes and global protein networks, which facilitates understanding the mechanism responsible for the cellular functions. Recently, the genomic-scale identification of protein-protein interaction (PPI) in model organisms, such as *Synechocystis* sp. PCC 6803 and *Xanthomonas oryzae*, have been published to map the whole protein-protein interaction networks [3,4]. Thanks to advanced high-throughput PPI experiments and information technology, many biologists can access large-scale species specific PPI data on the web [5]. Several web sites have been developed to disseminate PPI data such as POINT [6], OPHID [7], and PIANA [8]. POINT and OPHID systems provide predicted PPI information using sequence homology. PIANA integrates several proteins and interaction databases. However, these web services do not include the following methods: structure domain or domain-domain interaction, interaction networks in a graphical network viewer, functional annotation, localization, or the physicochemical properties of PPI data.

We constructed a web-based server, PutidaNET specifically for *P. putida* using major PPI algorithms. Functional and physicochemical annotations are provided using KEGG [9], Gene Ontology [10], amino acid distribution, instability index, isoelectric point, Gravy score, and sub-cellular localization. PutidaNET is designed to be user-friendly and easy to use.

## Methods

### Prediction of protein-protein interaction

The prediction of PPI is based on .PSIMAP [11,12], PEIMAP, and iPfam (domain-domain interaction) [4]. PSIMAP predicts interactions among proteins by using the BLASTP algorithm [13] with a common expectation value (E-value) cut-off of 0.0001. Interactions among domains or proteins for known PDB (Protein Data Bank) <http://www.rcsb.org/pdb> structures are the basis of the

predictions. PEIMAP includes integrating various experimental protein-protein interaction databases such as BIND [14], DIP[15], IntAct [16], MINT [17], HPRD [18], CYGD) [19], and BioGrid [20]. PSIMAP and PEIMAP assume that, in terms of unknown proteins, the query tends to interact with its homolog's partners. The most commonly used concept is 'homologous interaction' [21-23]. In this step, we used to recruit homologous sequences using the PSI-BLAST [14] with a cut-off of 40% sequence identity. Furthermore, we have aligned the Pfam [24] domains of all the *P. putida* KT2440 proteins with hmmpfam by the cut-off of 0.01 (E-value).

In order to select more reliable PPIs, we developed and used a 'combined score' between any pair of proteins which were predicted by PEIMAP, PSIMAP, and iPfam algorithms. This scoring method is also used by the STRING server <http://string.embl.de>[25].

### Protein function annotation

In order to understand the biological function of *P. putida* KT2440 proteome, we searched physicochemical properties and cross-reference databases using KEGG and GO. We used Biopython [26] modules to acquire physicochemical properties, including hydropathy profile, GRAVY score (the average hydropathy score), molecular weight, amino acid distribution, isoelectric point, and protein instability index. In addition, we predicted trans-membrane helices and signal peptides using Phobius [27] and SignalP 3.0 [28] programs for the sub-cellular localization prediction of *P. putida* KT2440 proteome.

PutidaNET provides cross-reference to public database information such as 1) KEGG pathways, 2) GO categories, and 3) GO-slim [29] through protein ID mapping. In order to gain more accurate statistical test results of KEGG and GO assignment, we added Fisher's exact test algorithm (P-value).

### Protein network analysis case study

#### Cell culture and MS/MS analysis

In order to find significant features, we integrated PPI network and proteomic data which were produced as previously described [31]. *P. putida* KT2440 was pre-cultured at 30°C with vigorous shaking in culture media (50 mM potassium phosphate buffer, pH 6.25, 3.4 mM MgSO<sub>4</sub>, 0.3 mM FeSO<sub>4</sub>, 0.2 mM CaCO<sub>3</sub>, 10 mM NH<sub>4</sub>Cl, and 10 mM sodium succinate) and then inoculated into 1 L culture media containing succinate (10 mM) or benzoate (5 mM) as a sole carbon source. The bacteria were harvested at the late exponential phase (absorbance at 600 nm. 0.7-0.8) and suspended in 20 mM Tris-HCl buffer (pH 8.0). Bacteria were disrupted by a French

pressure cell (SLM AMINCO, Urbana, IL) at 20,000 lb/in<sup>2</sup>, and soluble protein mixtures were prepared by centrifugation (15,000 g, 45 min). The protein samples were fractionated by 12% SDS-PAGE. The gel lanes were divided into 42 fractions according to molecular weight, and the sliced gels were digested with trypsin (Promega, Madison, WI). The resulting peptide extracts were pooled and lyophilized in a vacuum concentrator. Tryptic peptides were dissolved with 0.5% TFA (Trifluoroacetic acid) solution prior to further 2D-LC fractionation and used for MS/MS analysis using LTQ linear Ion Trap MS (ThermoFinnigan San Jose, CA). For the database search, the *P. putida* protein database was downloaded from the National Center for Biotechnology Information <http://www.ncbi.nlm.nih.gov/>. Tryptic peptides were identified using SEQUEST (version 3.1 SR1, ThermoFinnigan).

For better accuracy of protein identification by MS/MS analysis, the *P. putida* protein database and the reverse protein database were used to exclude any false-discovered proteins [30].

*sub-network*

We acquired the protein lists in culture media including succinate or benzoate. In order to find regulated sub-network by succinate and benzoate, we analyzed betweenness centrality (BC), the number of shortest paths going through a certain node, and degree, the number of interaction partners, using NetworkAnalyzer, a cytoscape plugin [31]. We used the R software containing some packages and Welch two sample t-test for P-value [32]. Also, we found potential functional modules using MCODE, a cytoscape plugin that finds clusters (highly interconnected regions) in protein networks [33].

**Results**

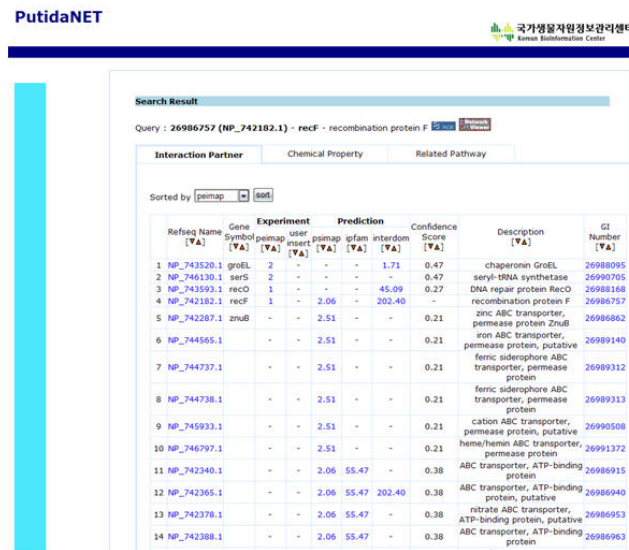
PutidaNET, a free accessible database with 3,254 proteins for *P. putida* KT2440, contains 82,019 PPI partners that have been predicted. Using the PPI algorithms, we obtained 61,011 (PSIMAP), 4,293 (PEIMAP), and 30,043 (iPfam) predicted PPIs except for self interaction. These PPIs were around 74.39% (PSIMAP), 5.23% (PEIMAP), and 36.62% (iPfam) of the *P. putida* KT2440 proteins. Although the total number of predicted interaction targets is very large, as they are ranked by combined score, experimentalists can select high ranking (more probable) ones according to their functional interests.

Figure 1 shows the search interface and the PutidaNET results. If a set of proteins is queried in the web interface, the user can acquire the physicochemical distribution against whole protein distribution, the trans-membrane

a)



b)

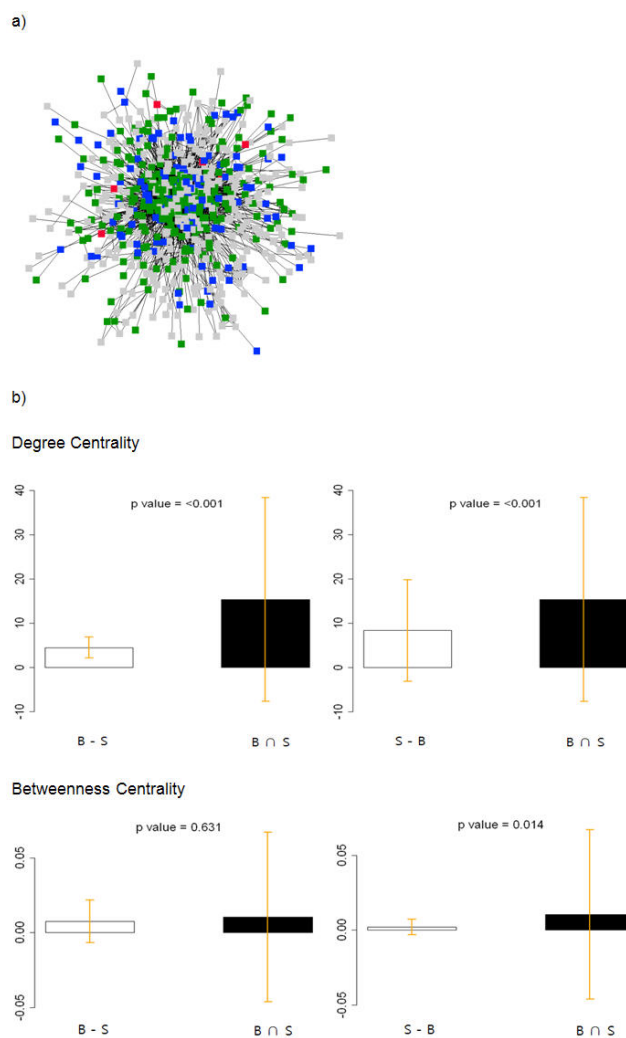


**Figure 1 PutidaNET system and interfaces.** (a) PutidaNET integrates three complementary protein-protein interaction databases including PSIMAP, PEIMAP, and iPfam. It shows three search interfaces: (1) search in high-confidence PPI network, (2) sequence search, (3) categorized tree navigation of gene ontology annotation. (b) A search result showing the list of predicted interacting proteins, confidence score, supporting description, and their synonymous IDs.

protein abundances, and the queried protein set. Therefore, this summarized information can be used to evaluate the input data quality. The user can easily predict protein-protein interaction for queried proteins and examine protein-protein interactions with a network

viewer made by JAVA. As a case study for PutidaNET, we used proteomics experimental data. As a measure of how central each protein is in the PPI network, we calculated two measures of betweenness centrality and degree for all the proteins in *P. putida* KT2440 [34]. And we colored the proteins which have mass abundance values in Figure 2a. From protein network analysis, we acquired some significant features about *P. putida* KT2440. PPIs were regulated specifically by difference sets of benzoate and succinate that tend to occur at the network periphery more than the network center (degree:  $P$ -value =  $< 0.001$  and betweenness centrality:  $P$ -value = 0.014, Figure 2b). Also box plot indicates the each mean by categories. As well as error bar indicates the each confidence interval 95%. This implies that the main protein network of *Pseudomonas putida* KT2440 is regulated by an intersection set of succinate and benzoate. However, PPIs which were detected at the network periphery could be regarded as key regulation factors to use succinate or benzoate by *P. putida* KT2440. We expect that commonly induced proteins in succinate and benzoate media will be included in the essential metabolic pathways, which will be constitutively or continuously expressed regardless of culture conditions. Comparative analysis of 2-DE of *P. putida* KT2440 cultured in minimal medium (succinate) and rich medium (LB) also showed that the major induced protein patterns were very similar (data not shown). Specifically induced proteins in benzoate medium were  $\beta$ -ketoadipate pathway enzymes for benzoate and 4-hydroxybenzoate and stress proteins. On the other hand, enzymes for TCA cycle, pyruvate metabolism, and glycolysis were increased in succinate medium, which will be increased for the utilization of succinate influx.

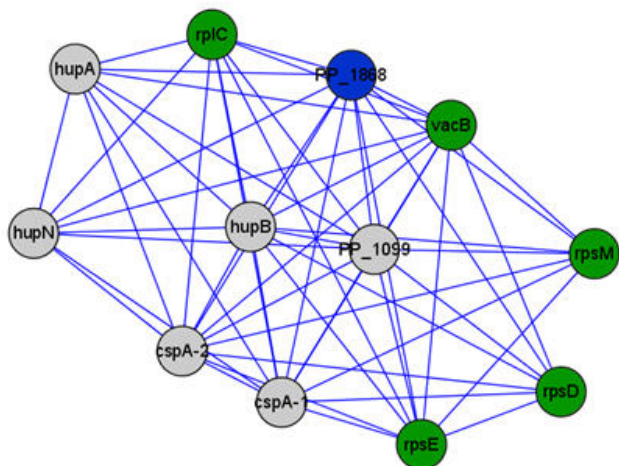
In order to find features in protein networks, we detected functional modules as highly interconnected sub-networks. As a result, we found five functional modules with KEGG pathway information (Figure 3, Additional file 1, Table 1). The functional modules are important PPIs because they represent protein complex or sub-pathway sharing biological functions. The modules which have less than a 0.001  $P$ -value were various metabolic pathways and ribosomes. The metabolic pathway modules describe the characteristics of *P. putida* KT2440 which has a high level of metabolic diversity for biodegradation. This high level of diversity enables the bacterium to utilize a wide range of carbon sources. The ribosome is an organelle that coordinates protein synthesis in all cells. The bacterial ribosome consists of



**Figure 2**  
**Analysis of *P. putida* KT2440 protein interaction network.** (a) *P. putida* KT2440 protein interaction network from the proteomic data as previously described [1]. Red nodes represent the unique proteins when cultured in benzoate, except for certain proteins when cultured in succinate. Blue nodes represent the unique proteins when cultured in succinate, except for certain proteins when cultured in benzoate. Green nodes represent the commonly expressed proteins when cultured in either benzoate or succinate. Grey nodes have no information about Kim's mass information. (b) Average Scores of degree and betweenness centrality in green nodes are higher than those of only unique red and blue. This means that proteins when cultured in either benzoate or succinate regulate specific metabolism. We considered a  $P$ -value less than 0.05 to be statistically significant.



Module 1



**Figure 3**  
**The biological module obtained from MCODE cytoscape plug –in.** This figure shows functional module. For example, module 1 is a functional module about ribosome pathways. The node is a protein reference ID by NCBI. These modules compose cliques. We explained the meaning of node color in Figure 2 legend. The five functional modules show the figure in Additional File 1.

more than 50 ribosomal subunit proteins and three rRNAs. Since bacterial cells contain vast amounts of ribosomes, most ribosomal subunit proteins can be observed as main peaks by mass spectrometry.

**Conclusion**

PutidaNET is the integration of mutually complementary protein-protein interaction information for the systematic analysis of *Pseudomonas putida*. The PutidaNET server is the first web server that provides various kinds of functional information such as a PPI viewer, physico-chemical properties, biological pathways, gene ontology, and protein-protein interaction for *P. putida* KT2440. It can assist researchers to access and obtain the information through an automatic annotation for queried proteins. Using proteomics data from certain medium conditions, we analyzed the characteristics of *P. putida* KT24400 using PutidaNET. Proteomic data gave us the quantitative information of induced proteins at benzoate or succinate culture conditions, which supplements the database. PPI combined with proteomic data can give users more specific information.

**Competing interests**

The authors declare that they have no competing interests.

**Table 1: Pathway analysis of five modules obtained from MCODE**

| <sup>a</sup> Module     | <sup>b</sup> Pathway                      | <sup>c</sup> Matched proteins | Total proteins       | <sup>d</sup> p-value      | <sup>e</sup> Gene list                         |
|-------------------------|---|-------------------------------|----------------------|---------------------------|--|
| <b>Module 1</b>         | Ribosome                                  | 4                             | 12                   | 3.6E <sup>-04</sup>       | rpsM, rpsD, rpsE, rplC                         |
| <b>Module2</b>          | Ribosome                                  | 8                             | 30                   | 5.10E <sup>-08</sup>      | rplV, rplA, rpsT, rplD, rpsC, rplM, rpsB, rpsA |
|                         | 1,2-Dichloroethane degradation            | 3                             |                      | 5.90E <sup>-04</sup>      | PP_2589, PP_3463, PP_2680                      |
|                         | Urea cycle and metabolism of amino groups | 4                             |                      | 7.80E <sup>-04</sup>      | PP_2589, PP_5278, PP_3463, PP_2680             |
|                         | 3-Chloroacrylic acid degradation          | 3                             |                      | 1.00E <sup>-03</sup>      | PP_2589, PP_3463, PP_2680                      |
|                         | Ascorbate and aldarate metabolism         | 3                             |                      | 1.30E <sup>-03</sup>      | PP_2589, PP_3463, PP_2680                      |
|                         | Glycerolipid metabolism                   | 3                             |                      | 2.90E <sup>-03</sup>      | PP_2589, PP_3463, PP_2680                      |
|                         | Butanoate metabolism                      | 4                             |                      | 3.00E <sup>-03</sup>      | PP_2589, PP_3463, PP_2680, ilvB                |
|                         | Bile acid biosynthesis                    | 3                             |                      | 3.70E <sup>-03</sup>      | PP_2589, PP_3463, PP_2680                      |
|                         | Histidine metabolism                      | 3                             |                      | 6.70E <sup>-03</sup>      | PP_2589, PP_3463, PP_2680                      |
|                         | Limonene and pinene degradation           | 3                             |                      | 7.30E <sup>-03</sup>      | PP_2589, PP_3463, PP_2680                      |
| beta-Alanine metabolism | 3   |                               | 8.60E <sup>-03</sup> | PP_2589, PP_3463, PP_2680 |  |
| <b>Module3</b>          | ABC transporters - General                | 4                             | 9                    | 3.8E <sup>-03</sup>       | PP_0225, aapP, PP_1068, PP_5022                |
| <b>Module4</b>          | Ribosome                                  | 8                             | 45                   | 9.9E <sup>-07</sup>       | rplI, rplB, rplQ, rplL, rpmB, rplP, rplE, rplX |
|                         | Glycolysis/Gluconeogenesis                | 4                             |                      | 4.1E <sup>-03</sup>       | aceE, eno, aceF, lpdG                          |
| <b>Module 5</b>         | Ribosome                                  | 6                             | 30                   | 2.8E <sup>-05</sup>       | rpsG, rpsN, rplR, rplS, rplW, rpsH             |

<sup>a</sup>We made five modules using MCODE.

<sup>b</sup>We analyzed pathway analysis using KEGG pathway information.

<sup>c</sup>Count column displays the number of protein containing each pathway in module.

<sup>d</sup>The p-value is confidence score by Fisher's exact test algorithm.

<sup>e</sup>Gene list column shows gene symbols containing each pathway.

## Authors' contributions

SJP, BCK, and JWR constructed the database. SWC and DP developed the web site. JSC designed the overall web site. KAL analyzed the statistics of PPI network results. SJP, JSC, DP, and JWR wrote the main draft of the paper. JB and SIK directed the study and helped with the draft manuscript.

## Note

Other papers from the meeting have been published as part of *BMC Bioinformatics* Volume 10 Supplement 15, 2009: Eighth International Conference on Bioinformatics (InCoB2009): Bioinformatics, available online at <http://www.biomedcentral.com/1471-2105/10?issue=S15>.

## Additional material

### Additional file 1

*Supplementary Figure 1-(a-e) the biological modules obtained from MCODE cytoscape plug-in. This figure shows five functional modules: a) module 1, b) module 2, c) module 3, d) module 4, and e) module 5. For example, module 1 is a functional module about ribosome pathways.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-S3-S18-S1.doc>]

## Acknowledgements

We thank our colleagues at KOBIC and Maryana Bhak for editing the article. This work was supported by a grant from the KRIBB Research Initiative Program of Korea, by a Korea Science and Engineering Foundation (KOSEF) grant funded by the Korean government (MOST) (No. M10508040002-07N0804-00210). In addition, this work was supported by a K-MeP grant (No. T28021) of Korea Basic Science Institute and the Marine and Extreme Genome Research Center Program of the Ministry of Land, Transportation and Maritime.

This article has been published as part of *BMC Genomics* Volume 10 Supplement 3, 2009: Eighth International Conference on Bioinformatics (InCoB2009): Computational Biology. The full contents of the supplement are available online at <http://www.biomedcentral.com/1471-2164/10?issue=S3>.

## References

- Kim YH, Cho K, Yun SH, Kim JY, Kwon KH, Yoo JS and Kim SI: **Analysis of aromatic catabolic pathways in *Pseudomonas putida* KT 2440 using a combined proteomic approach: 2-DE/MS and cleavable isotope-coded affinity tag analysis.** *Proteomics* 2006, **6(4)**:1301-1318.
- Larimer FW, Chain P, Hauser L, Lamerdin J, Malfatti S, Do L, Land ML, Pelletier DA, Beatty JT and Lang AS, et al: **Complete genome sequence of the metabolically versatile photosynthetic bacterium *Rhodospirillum rubrum*.** *Nat Biotechnol* 2004, **22(1)**:55-61.
- Kim WY, Kang S, Kim BC, Oh J, Cho S, Bhak J and Choi JS: **SynechoNET: integrated protein-protein interaction database of a model cyanobacterium *Synechocystis* sp. PCC 6803.** *BMC Bioinformatics* 2008, **9(Suppl 1)**:S20.
- Kim JG, Park D, Kim BC, Cho SW, Kim YT, Park YJ, Cho HJ, Park H, Kim KB and Yoon KO, et al: **Predicting the interactome of *Xanthomonas oryzae pathovar oryzae* for target selection and DB service.** *BMC Bioinformatics* 2008, **9**:41.
- Park D, Kim BC, Cho SW, Park SJ, Choi JS, Kim SI, Bhak J and Lee S: **MassNet: a functional annotation service for protein mass spectrometry data.** *Nucleic Acids Res* 2008, **36** Web Server: W491-495.
- Bowers PM, Pellegrini M, Thompson MJ, Fierro J, Yeates TO and Eisenberg D: **Prolinks: a database of protein functional linkages derived from coevolution.** *Genome Biol* 2004, **5(5)**:R35.
- Brown KR and Jurisica I: **Online predicted human interaction database.** *Bioinformatics* 2005, **21(9)**:2076-2082.
- Aragues R, Jaeggi D and Oliva B: **PIANA: protein interactions and network analysis.** *Bioinformatics* 2006, **22(8)**:1015-1017.
- Kanehisa M and Goto S: **KEGG: kyoto encyclopedia of genes and genomes.** *Nucleic Acids Res* 2000, **28(1)**:27-30.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS and Eppig JT, et al: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.** *Nat Genet* 2000, **25(1)**:25-29.
- Park J, Lappe M and Teichmann SA: **Mapping protein family interactions: intramolecular and intermolecular protein family interaction repertoires in the PDB and yeast.** *Journal of molecular biology* 2001, **307(3)**:929-938.
- Gong S, Yoon G, Jang I, Bolser D, Dafas P, Schroeder M, Choi H, Cho Y, Han K and Lee S, et al: **PSIBase: a database of Protein Structural Interactome map (PSIMAP).** *Bioinformatics (Oxford, England)* 2005, **21(10)**:2541-2543.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W and Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25(17)**:3389-3402.
- Bader GD and Hogue CW: **BIND—a data specification for storing and describing biomolecular interactions, molecular complexes and pathways.** *Bioinformatics (Oxford, England)* 2000, **16(5)**:465-477.
- Xenarios I, Rice DW, Salwinski L, Baron MK, Marcotte EM and Eisenberg D: **DIP: the database of interacting proteins.** *Nucleic acids research* 2000, **28(1)**:289-291.
- Hermjakob H, Montecchi-Palazzi L, Lewington C, Mudali S, Kerrien S, Orchard S, Vingron M, Roechert B, Roepstorff P and Valencia A, et al: **IntAct: an open source molecular interaction database.** *Nucleic acids research* 2004, **32 Database**: D452-455.
- Zanzoni A, Montecchi-Palazzi L, Quondam M, Ausiello G, Helmer-Citterich M and Cesareni G: **MINT: a Molecular INteraction database.** *FEBS letters* 2002, **513(1)**:135-140.
- Peri S, Navarro JD, Kristiansen TZ, Amanchy R, Surendranath V, Muthusamy B, Gandhi TK, Chandrika KN, Deshpande N and Suresh S, et al: **Human protein reference database as a discovery resource for proteomics.** *Nucleic acids research* 2004, **32 Database**: D497-501.
- Guldener U, Munsterkottler M, Kastenmuller G, Strack N, van Helden J, Lemer C, Richelles J, Wodak SJ, Garcia-Martinez J and Perez-Ortin JE, et al: **CYGD: the Comprehensive Yeast Genome Database.** *Nucleic acids research* 2005, **33 Database**: D364-368.
- Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A and Tyers M: **BioGRID: a general repository for interaction datasets.** *Nucleic acids research* 2006, **34 Database**: D535-539.
- Marcotte EM, Pellegrini M, Thompson MJ, Yeates TO and Eisenberg D: **A combined algorithm for genome-wide prediction of protein function.** *Nature* 1999, **402(6757)**:83-86.
- Walhout AJ, Sordella R, Lu X, Hartley JL, Temple GF, Brasch MA, Thierry-Mieg N and Vidal M: **Protein interaction mapping in *C. elegans* using proteins involved in vulval development.** *Science* 2000, **287(5450)**:116-122.
- Deane CM, Salwinski L, Xenarios I and Eisenberg D: **Protein interactions: two methods for assessment of the reliability of high throughput observations.** *Mol Cell Proteomics* 2002, **1(5)**:349-356.
- Sonnhammer EL, Eddy SR, Birney E, Bateman A and Durbin R: **Pfam: multiple sequence alignments and HMM-profiles of protein domains.** *Nucleic Acids Res* 1998, **26(1)**:320-322.
- von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P and Snel B: **STRING: a database of predicted functional associations between proteins.** *Nucleic acids research* 2003, **31(1)**:258-261.
- Chapman B and Chang J: **Biopython: Python tools for computational biology.** *ACM SIGBIO Newsletter* 2000, **20(2)**: 15-19.

27. Kall L, Krogh A and Sonnhammer EL: **A combined transmembrane topology and signal peptide prediction method.** *Journal of molecular biology* 2004, **338(5)**:1027–1036.
28. Emanuelsson O, Brunak S, von Heijne G and Nielsen H: **Locating proteins in the cell using TargetP, SignalP and related tools.** *Nature protocols* 2007, **2(4)**:953–971.
29. Camon E, Magrane M, Barrell D, Lee V, Dimmer E, Maslen J, Binns D, Harte N, Lopez R and Apweiler R: **The Gene Ontology Annotation (GOA) Database: sharing knowledge in Uniprot with Gene Ontology.** *Nucleic acids research* 2004, **32 Database**: D262–266.
30. Park GW, Kwon KH, Kim JY, Lee JH, Yun SH, Kim SI, Park YM, Cho SY, Paik YK and Yoo JS: **Human plasma proteome analysis by reversed sequence database search and molecular weight correlation based on a bacterial proteome analysis.** *Proteomics* 2006, **6(4)**:1121–1132.
31. Assenov Y, Ramirez F, Schelhorn SE, Lengauer T and Albrecht M: **Computing topological parameters of biological networks.** *Bioinformatics* 2008, **24(2)**:282–284.
32. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y and Gentry J, et al: **Bioconductor: open software development for computational biology and bioinformatics.** *Genome Biol* 2004, **5(10)**:R80.
33. Bader GD and Hogue CW: **An automated method for finding molecular complexes in large protein interaction networks.** *BMC Bioinformatics* 2003, **4**:2.
34. Joy MP, Brock A, Ingber DE and Huang S: **High-betweenness proteins in the yeast protein interaction network.** *J Biomed Biotechnol* 2005, **2005(2)**:96–103.

Publish with **BioMed Central** and every scientist can read your work free of charge

*"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."*

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

