

PublisherInfo		
PublisherName	:	BioMed Central
PublisherLocation	:	London
PublisherImprintName	:	BioMed Central

Protein database unveiled

ArticleInfo		
ArticleID	:	4852
ArticleDOI	:	10.1186/gb-spotlight-20031006-01
ArticleCitationID	:	spotlight-20031006-01
ArticleSequenceNumber	:	204
ArticleCategory	:	Research news
ArticleFirstPage	:	1
ArticleLastPage	:	3
ArticleHistory	:	RegistrationDate : 2003-10-6 OnlineDate : 2003-10-6
ArticleCopyright	:	BioMed Central Ltd2003
ArticleGrants	:	
ArticleContext	:	130594411

Charles Q Choi

Email: cqchoi@nasw.org

An international team of scientists last week unveiled the [Human Protein Reference Database](#), an online database that currently contains entries for the 3000 most-studied human proteins. Information on a total of 10,000 proteins is expected to be in the database by year's end, freely accessible to noncommercial researchers.

"We think this database is the most user-friendly and comprehensive and annotated resource so far for the proteins in it. Most features of proteins that biologists care about and would want to see are in one place here," said Akhilesh Pandey, the database's principal investigator and assistant professor at Johns Hopkins University.

Ease of use was a high priority for the database, Pandey said. For instance, a biologist looking up information on the breast cancer gene *BRCA1* can search by any of its names and get a single entry containing everything - its alternate names, structure, function, sequence, how it's modified, known interactions with other proteins, where it's found in cells, where it's found in the body.

"We are providing for the first time a comprehensive picture of protein-protein interactions in humans," Pandey said.

The database includes each protein's known roles in health and disease and direct links to related scientific papers. Only experimentally proven or widely accepted facts about a protein are included, without mixing in unproven computer-generated predictions. In the future, the database team hopes the biology community will help to provide updates on proteins as they come.

Pandey is often asked why his team started yet another protein database when others already exist. "Right now, there is information scattered across many different databases with varying levels of accuracy, which does not help the cause of the average biologist," he said. "We feel no gold [standard](#) exists for these databases yet, and we want to develop it."

"I think it's got a lot of potential," said Canada Research Chair in Proteomics Guy Poirier of the Université Laval in Quebec. "It has references to signaling pathways, whereas a lot of other databases just mention the name of the protein and don't mention links to the functions of other proteins."

Dozens of researchers at the [Institute of Bioinformatics](#) in Bangalore, India, created the database with colleagues in the United States, Belgium, Denmark, and Spain. They critically reviewed hundreds of thousands of scientific papers, drawing connections between papers and resolving inconsistencies. Each researcher read an average of at least 10 to 20 papers each every day, with every protein reviewed twice.

"The numbers are closer to 50 a day, but people would tell me they don't believe me," Pandey said.

The Human Protein Reference Database started off the [Online Mendelian Inheritance in Man](#) database and also pulls information from smaller, existing databases to complete every protein's entry. Pandey feels the Human Protein Reference Database's strength is its more accurate and complete entries due to its emphasis on manual curation of entries, as opposed to the automated computer programs most databases employ.

Unveiled Thursday (October 2) in the October [Genome Research](#), the database has been under development since May 2002 and active for 5 months, receiving almost 2 million hits just from word of mouth and presentations at scientific meetings, Pandey said. Johns Hopkins Licensing and Technology Development is currently establishing criteria for companies interested in using the database to pay fees under licensing arrangements.

If researchers searching the database find that a protein of interest has not yet been annotated, they can submit a request for annotation online. Researchers who wish to review a molecule, or even an entire protein family, are welcome to volunteer and receive credit as reviewer for that molecule.

References

1. Human Protein Reference Database, [<http://www.hprd.org>]
2. Spinney L: First Human Proteome Organisation congress *Genome Biology*, November 25, 2002., [http://genomebiology.com/researchnews/default.asp?arx_id=gb-spotlight-20021129-01]
3. Institute of Bioinformatics, [<http://www.ibioinformatics.org>]
4. Online Mendelian Inheritance in Man, [<http://www.ncbi.nlm.nih.gov/omim/>]
5. *Genome Research*, [<http://www.genome.org/>]