

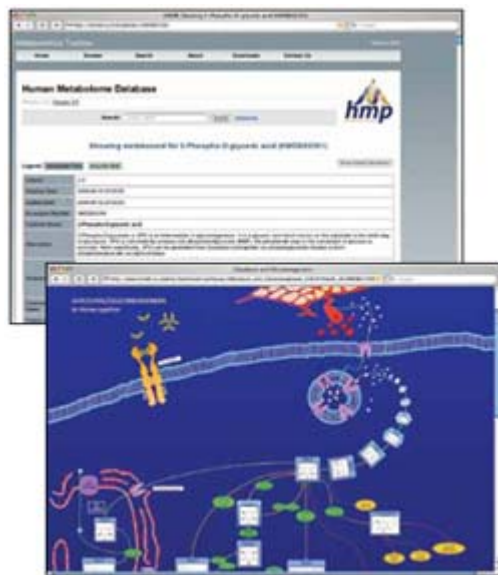
The Scientist

Volume 23 | Issue 5 | Page 51

Charting the Human Metabolome

The success of a database of metabolites has whetted researchers' appetites for more.

By Megan Scudellari



Courtesy of The Human Metabolome Project

After researchers coined the term "metabolomics" in 1998, it appeared in only one or two papers per year. But bolstered by decades of research in analytical chemistry, the field—which focuses on the complete set of small molecule metabolites in a cell, tissue, or organism—rapidly adopted its newfound identity: In 2008, over 500 papers mentioned the word "metabolomics." With growing stacks of literature filled with spectral data, concentrations, chemical structures, and more, the field developed an acute need for ways to process and share information.

"In the world of human metabolomics, you didn't have a database, so if someone found a new metabolite, it stayed there in the paper," says David Wishart, a computational biologist at the University of Alberta. So he decided to build one.

In 2004, with funding from the Canadian Foundation for Innovation and Genome Canada, Wishart began the Human Metabolome Database (HMDB), part of Genome Canada's Human Metabolome Project. A team of spectroscopists, organic chemists, physicians, and bioinformaticians spent two years gathering metabolite data from literature, databases, and their own experimental results. It was a Herculean task. In addition to identifying all detectable metabolites in the human body, the equivalent of the Human Genome Project, the HMDB required structures and concentrations of those metabolites. "It's intrinsically a more difficult process," says Wishart. "It's not something which can be easily streamlined or automated."

The freely-available database made its online debut in late 2006, and in 2007, Wishart and colleagues announced its arrival with this month's Hot Paper, published in *Nucleic Acids Research*, inviting researchers to utilize the database's initial collection of over 2180 metabolite entries. Since its publication, the resource has grown several-fold, with updates and additional databases, and has served as a vital resource for research in cancer, digestive disease, and diabetic nephropathy, among other pathologies. But despite its rapid growth, some researchers find the database too limited for the needs of the field. "The HMDB is a very good first step," says Oliver Fiehn of the University of California, Davis, a member of the Metabolomics Society board of directors, "but it's far from where we have to go."

HMDB 2.0

Today, over \$8.1 million has been poured into the HMDB. It contains over 7100 metabolite entries, and has been cited as a resource in over 100 papers. Monthly hits have doubled since the beginning of 2007, from about 50,000 per month to over 100,000.

In January 2009, HMDB 2.0 was released (<http://www.hmdb.ca/>).¹ Updates include additional data fields, 60 hand-drawn, hyperlinked metabolic pathway maps, and new browsing tools. With Disease Browse, users can scroll through tables of diseases to find associated metabolites. The most popular feature of the database is the ability to download part or all of it, says Wishart.

The HMDB is a "crucial" resource, says Adrian Arakaki of the Georgia Institute of Technology. In a recent experiment, Arakaki and colleagues used the HMDB to determine normal concentrations of human metabolites to compare against predicted levels of those metabolites in leukemia cells.² At the Center for Magnetic Resonance at the University of Florence, Italy, Ivano Bertini and colleagues utilized the database for a recent study of the metabolomics of celiac disease (CD).³ In a study of biomarkers for diabetic kidney disease, Subramaniam Pennathur and colleagues at the University of Michigan isolated elevated metabolites from the urine of diseased mice, then searched the HMDB for their likely identities based on molecular weights.⁴ "We were able to find the majority of metabolites listed," says Pennathur.

More, please

But what the HMDB boasts in depth, averaging 90 different data entries per metabolite, it lacks in breadth, since it is restricted to human-generated metabolites. "There are estimations that the human metabolome is in the order of thousands of metabolites, but there are maybe 20,000 for microbes and 200,000 for plants," says Christoph Steinbeck of the European Bioinformatics

Institute (EBI) in Cambridge. The human gut, for example, is home to over a kilogram of microbes that are not included in the HMDB, adds Fiehn. Researchers need a centralized database that includes those other organisms, he says.

Across the globe, countries have taken up the challenge. In the United States, the major NIH-funded effort is the Metabolomics Network for Drug Response Phenotypes, led by Rima Kaddurah-Daouk, founder and past president of the Metabolomics Society and director of the Pharmacometabolomics Center at Duke University Medical Center. In conjunction with Wishart, the network is working to design and engineer a data management system for clinical drug trial data that connects to the HMDB. "It's a work in progress," says Kaddurah-Daouk.

In Europe, several national initiatives are working to develop broad databases. The largest of these is the Netherlands Metabolomics Centre in Leiden, established in January 2008. Over \$67 million (€53 million) of government, industry, and academic funds have been funneled into the Centre, which plans to establish several databases for storage and retrieval of metabolite information, according to director Thomas Hankemeier.

In addition to national efforts, a fledgling initiative to develop a centralized European database—including microbe, plant, and model organism metabolites—is currently being spearheaded by Steinbeck at EBI. Researchers have "nice local resources," says Steinbeck, "but the big, centralized, well-curated resources are still missing."

Wishart agrees. "The ultimate goal for everyone is to create something that is as deep as the HMDB but that covers all of the other important classes of organisms," he says. "Then we'd have a really comprehensive, really useful resource."

Data derived from the Science Watch/Hot Papers database and the Web of Science (Thomson ISI) show that Hot Papers are cited 50 to 100 times more often than the average paper of the same type and age.

D.S. Wishart et al., "HMDB: the Human Metabolome Database," *Nucleic Acids Res*, 35:D521–26, 2007. (Cited in 105 papers)

References

1. D.S. Wishart et al., "HMDB: a knowledgebase for the human metabolome," *Nucleic Acids Res*, 37:D603–10, 2009.
2. A.K. Arakaki et al., "Identification of metabolites with anticancer properties by computational metabolomics," *Mol Cancer*, 7:57, 2008.
3. I. Bertini et al., "The metabonomic signature of celiac disease," *J Proteome Res.*, 8:170–77, 2009.
4. H. Zang et al., "Rosiglitazone reduces renal and plasma markers of oxidative injury and reverses urinary metabolite abnormalities in the amelioration of diabetic nephropathy," *Am J Physiol Renal Physiol*, 295:F1071–81, 2008.

