Poster J-12

GEM System version 2 for the analysis of metabolic pathways

Authors:
Kazuharu Arakawa (*Institute for Advanced Biosciences, Keio University, and Japan Society for the Promotion of Science*)
Nobuaki Kono (*Institute for Advanced Biosciences, Keio University*)
Satoshi Tamaki (*Institute for Advanced Biosciences, Keio University*)
Hiromi Yoshida (*Institute for Advanced Biosciences, Keio University*)
Masaru Tomita (*Institute for Advanced Biosciences, Keio University*)

Short Abstract: The Genome-based E-cell Modeling System (GEM System) realizes a automatic conversion of genome sequence data into a quantitative in silico cell-wide metabolic pathway model. Manually curated database of kinetic in silico models maintained for this purpose facilitates the dynamic modeling process.

Long Abstract:
Knowledge in molecular biology is rapidly accumulating in the fields of genome, transcriptome, proteome, and metabolome, demanding for a systems biology approach in order to view the dynamic behavior of a cell as a complex system. However, simulation is a challenging task especially where large-scale modeling is required, due to the necessity for vast amount of accurate parameters and equations that are difficult to be obtained as a complete set. Therefore a large scale modeling of cell in silico demands for a novel high-throughput approach. If successfully integrated, availability of large amount of genome sequence, transcripts and expression data, enzyme reaction data, metabolic pathway maps, and the data of metabolites in cells will create a strong base for a quantitative dynamic cell model.

Here we introduce Genome-based Modeling System (GEM System) application suite developed upon the G-language Genome Analysis Environment for comprehensive study of metabolic pathways. The software system has components for pathway reconstruction based on complete genome sequence data, kinetic simulation model database designated in silico Model Database (ISM) and dynamic model generator with this database, and pathway mapping and visualization tool. Pathway reconstruction is achieved by linking information from major public database such as GenBank, EMBL, SWISS-PROT, KEGG, ARM, Brenda, and WIT, by identifying the cellular proteome through homology and orthology mapping from the genome information, and retrieving the enzyme reaction network thereof. Then a list of stoichiometric reactions of enzymes is compiled, referring the matched protein in the SWISS-PROT database, and a pathway model is generated in Systems Biology Markup Language (SBML) format. The stoichiometric reaction list is then checked for connectivity and isozymes by comparison with reference pathways in KEGG or BioCyc database. Automatically generated dynamic model of cell-wide metabolic pathway includes an Escherichia coli model with 968 reactions and 1195 metabolites, covering 100% of the corresponding metabolic pathway in KEGG database, and 93% of EcoCyc database and 95% of genome-scale metabolic flux model by Palsson et al. (iJR904). Most other bacterial

models achieved more than 90% KEGG coverage. This process is achieved very fast, finishing the entire step for a genome in a few minutes. List of automatically generated bacterial models are available at: http://www.g-language.org/gem/models/static.cgi.

The stoichiometric model is further annotated with the dynamic kinetic equations and parameters available in Brenda or ISM database. ISM is a public online database of kinetic models manually curated from literature, built and maintained for this purpose. There are currently over 100 models in this database, and all models and their components and annotations are searchable. ISM database alone allows model integration by selecting the desired components from the web-site, and the integrated model is generated in SBML and E-Cell ready format. The GEM System automatically retrieves the corresponding information defined in the above stoichiometric model from ISM database and generates a dynamic model. However, since the available kinetic information is limited, large fraction of the parameters had to be estimated since they were not available in the databases, and this feature is currently preliminary. On the other hand, this process is effective for well-studied pathway such as glycolysis in E.coli, the automatically generated in silico model was comparable with the experimental results.

Pathway visualization component maps multiple layers of omics data, including transcriptome, proteome, and metabolome, simultaneously onto KEGG pathways. The mapped pathway image is generated as FLASH vector image for cross-platform viewing. This pathway mapping software is available online as a web application at: http://www.g-language.org/data/marray/. A database of microarray data visualized with this application is also available at the above web site.

Arakawa K, Mori K, Ikeda K, Matsuzaki T, Kobayashi Y, Tomita M, "G-language Genome Analysis Environment: a workbench for nucleotide sequence data mining", Bioinformatics, 2003, 19(2):305-306

Arakawa K, Kono N, Yamada Y, Mori H, and Tomita M, "KEGG-based pathway visualization tool for complex omics data", In Silico Biology, 2005, 5:0039

Arakawa K, Nakayama Y, Tomita M, "GPAC: Benchmarking the sensitivity of genome informatics analysis to genome annotation completeness", In Silico Biology, 2006, 6:0006

Arakawa K, Yamada Y, Shinoda K, Nakayama Y, Tomita M, "GEM System: Automatic prototyping of cell-wide metabolic pathway models from genomes", BMC Bioinformatics, 7:168