

Title:**Cellular machine-based Bayesian factor analysis: Exploring drug's mode-of-action in a cell****Authors:**Sangjo Han¹ and Dongsup Kim^{1*}**Address:**¹Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology, Daejeon, South Korea

*To whom correspondence should be addressed. E-mail: kds@kaist.ac.kr

Extended Abstract

In a seminal article in the journal *Cell* entitled 'The cell as a collection of protein machines: preparing the next generation of molecular biologists', Bruce Alberts described a cell as a factory: 'Indeed, the entire cell can be viewed as a factory that contains an elaborated network of interlocking assembly lines, each of which is composed of a set of large protein machine'. In recent genome-wide study of yeast, two independent groups, European Molecular Biology Laboratory (EMBL), Cellzome (a spin-off company from EMBL), and the university of Toronto, have surveyed first comprehensive protein complexes, called protein machines by Bruce Alberts, using tandem affinity purification (TAP). Furthermore, Cellzome's paper suggests molecular rationale of protein complexes for gene-to-phenotype relationship. In our study, those protein complexes were re-defined as 'cellular machine' to emphasize essential roles for cellular process of 'protein complexes' or 'molecular machines', and then cellular machine-based Bayesian factor analysis was developed to relate growth fitness of genome-wide deletion strains to 'hidden activities' of a collection of cellular machines in a cell. In other words, at the organism's phenotype level, the chemical-genetic profiles representing relative growth fitness of systematic deletion strains have been modeled in terms of cellular machines at the molecular level (Figure 1). This is the first and essential approach in order to model growth fitness of pooled deletion strains at the molecular level using real biological entities.

To show that our cellular machine-based model is reasonable and machine activities inferred by our model are reliable, hierarchical clustering analysis and literature survey were performed, which showed predictive power of common mode-of-action of bioactive compounds

as well as grouping of cellular machines with similar biological behavior. We finally presented machine-based way of drug's target-pathway prediction to emphasize practical aspect of cellular machine-based modeling. For example, from machine-based approach, we were able to highlight target-protein, TOR1, of rapamycin (Figure 2) as well as RUB1, UBA3, UBC12, and ULA1 related to protein neddylation as relevant biological pathway for cellular toxicity of camptothecin. We believe that our machine-based approach can provide appropriate framework for combining and modeling genome-wide functional profiling data of several types of pooled deletion strains under different environments (i.e., heterozygous and homozygous deletion strains or between different species), which might especially contribute to predict more precisely relevant pathway including target-proteins that interact directly with bioactive compounds.

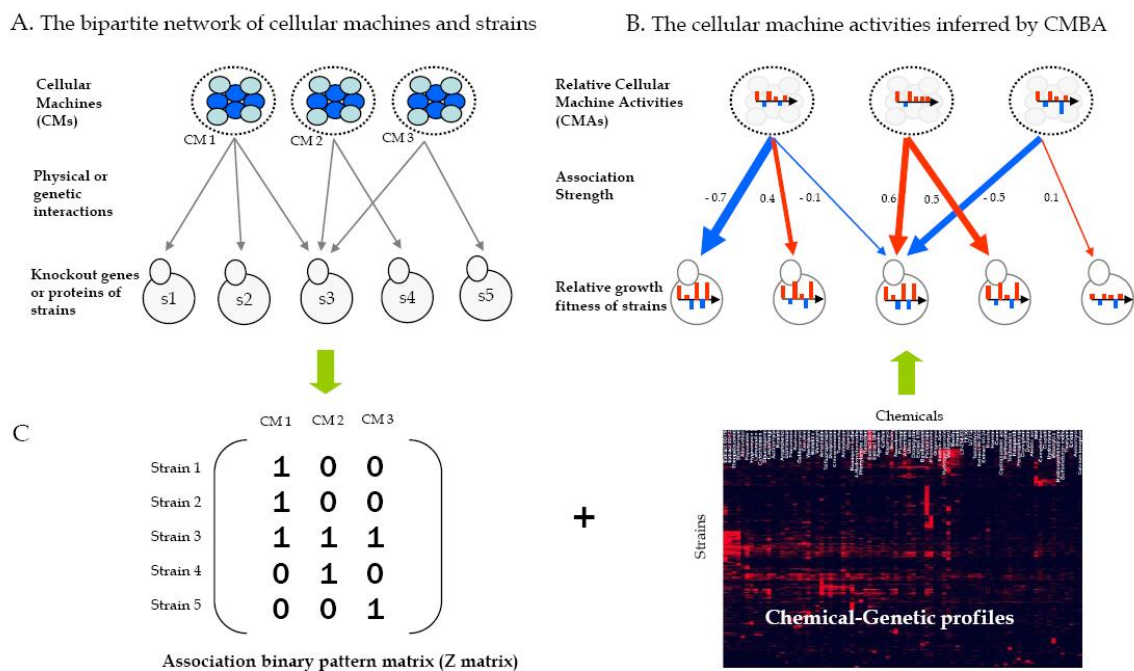


Figure 1. Procedures for inferring the hidden activities of a collection of cellular machines in a cell (A) A bipartite network illustrating the first-order relationships between cellular machines (CMs) and strains they are associated with. The definitions of ‘cellular machine’, ‘strains’, and ‘association’ in the study are as follows: the first yeast comprehensive protein complexes have been re-defined as a collection of ‘cellular machines’ in a cell to emphasize essential roles for cellular process of those ‘protein complexes’. The ‘strains’ are defined as a collection of pooled deletion mutants released from Saccharomyces Genome Deletion Project. The ‘association’ is defined as the existence of physical or genetic interactions between at least one of components in CMs and knockout gene product of a strain. In bipartite network, we

assume that the relative growth fitness of strains under different chemicals is mainly caused by the deleterious associations of CMs and strains. (B) The bipartite CM-strain network reconstructed by applying cellular machine based Bayesian factor analysis (CMBA). The bar charts within dotted circles in the top of panel show the relative activities of CM depending on chemicals inferred from our analysis. The bar charts within each strain in the bottom represent the relative growth fitness under different chemicals, which are used as observed data for our analysis. The thicknesses of arrows in the middle denote the association strength between CMs and strains inferred from our analysis. The colors of red and blue indicate ‘positive’ and ‘negative’ association, respectively. (C) It shows two types of input data for CMBA, one of which is genetic and physical interaction as a *prior* knowledge in the left, which is represented in the form of matrix containing initial association binary patterns of each strain (row) to CMs (column) called Z matrix. The other is the chemical-genetic profiles representing relative growth fitness of pooled deletion strains under different chemicals are shown as known observed data for CMBA in the right, which is called E matrix.

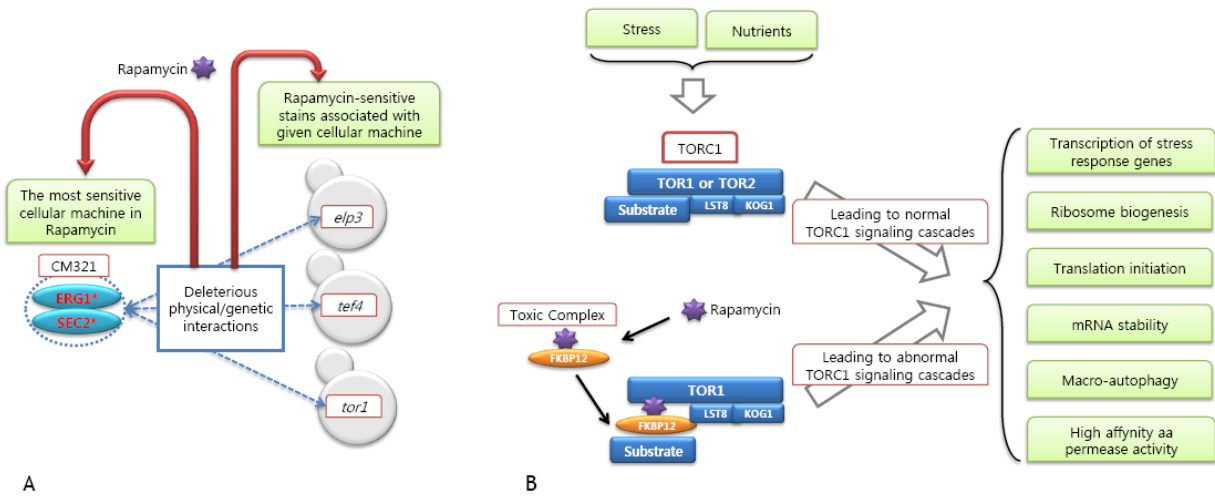


Figure 2. Target pathway of rapamycin (A) The most sensitive cellular machine to rapamycin, CM321, is comprised of ERG1 and SEC2, both of which are essential genes. Any of components in such machine may be physically or genetically associated with ELP3, TEF4, and TOR1 among gene products deleted in all the sensitive strains to rapamycin. According to model assumption, it can be interpreted as follows: Strains of machine-associated gene deletions have deleterious biological interactions with that machine. It may lead to decrease the growth fitness of those strains in rapamycin. (B) Target of Rapamycin (TOR) pathway primarily regulated by Target of Rapamycin Complex 1(TORC1) with/without rapamycin. When the rapamycin is treated in a yeast cell, it binds FKBP12, forming toxic complex, which inhibits specifically TOR1, an essential component of TORC1. It gives rise to abnormal TORC1 signaling cascades related to broad biological functions, transcription, translation, mRNA stability and permeability.