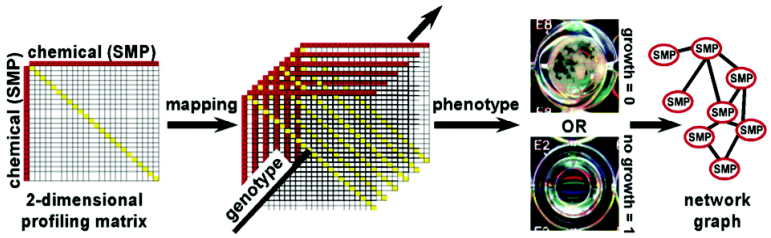
Communication

Chemical Genomic Profiling of Biological Networks Using Graph Theory and Combinations of Small Molecule Perturbations

Stephen J. Haggarty, Paul A. Clemons, and Stuart L. Schreiber

J. Am. Chem. Soc., **2003**, 125 (35), 10543-10545• DOI: 10.1021/ja035413p • Publication Date (Web): 19 August 2003



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 4 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S
COMMUNICATIONS

Published on Web 08/19/2003

Chemical Genomic Profiling of Biological Networks Using Graph Theory and Combinations of Small Molecule Perturbations

Stephen J. Haggarty, Paul A. Clemons, and Stuart L. Schreiber*

Departments of Chemistry and Chemical Biology and of Molecular and Cellular Biology, Howard Hughes Medical Institute, Harvard Institute of Chemistry & Cell Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138

Received April 1, 2003; E-mail: sls@slsiris.harvard.edu

Profiling experiments, where genome-wide measurements are made of multiple experimental samples, can yield rich fingerprints for comparison and interpretation.[1] Differential labeling of mRNA or protein samples and their analyses on microarrays and two-dimensional gels, respectively, are facilitating global views of biological networks.[2] Here, we describe a new type of profiling experiment where the response of genetically similar but not identical cells to pairwise combinations of biologically active small molecules yields a network of chemical genetic interactions. The ability of combinations of small molecules to interact antagonistically or synergistically provides a chemical tool to resolve differences between biological networks; we refer to this as chemical genomic profiling.

Chemical genomic profiling was performed using a wild-type (WT), haploid strain (W303 background) of the budding yeast *Saccharomyces cerevisiae* along with nine otherwise isogenic deletion strains, each missing a component of the spindle assembly/cell polarity network (Table 1).[3] As a model phenotype relevant to the function of these deleted genes, we chose cell cycle progression because of the ease of measuring a change in the optical density of cells cultured in liquid media. To obtain a chemical genomic profile, a two-dimensional matrix of all possible pairwise combinations of 24 small molecules, each with a different structure and known biological activity, was "mapped" onto the WT and nine deletion strains for a total of 5760 perturbations (Figure 1).[4−6]

Besides DMSO as a solvent control, we chose nocodazole (microtubule destabilizer), latrunculin B (actin destabilizer), and rapamycin (inhibitor of TOR proteins), which are biologically active molecules known to reduce yeast growth in a genotype-dependent manner (Figure 2). The other 20 biologically active small molecules were identified as positives from a collection of small molecules[7,8] at the Harvard ICCB using yeast chemical genetic modifier and "synthetic-lethal" screens.[9,10] We refer to these modulators, many with unknown targets and mechanisms of action, as SMPs, for small molecule perturbagens.

After the yeast cultures were incubated for 40 h at room temperature, each well was analyzed visually, and the optical density was compared to the effect of adding DMSO alone. For each strain, the data were encoded into the form of a binary adjacency matrix, ***A***, with one row and one column for each of the 24 small molecules. A value of 0 was used to indicate no observable effect on growth, and a value of 1 was used to indicate no growth or that growth was reduced, in both replicates.[11] Each adjacency matrix was then used to construct a discrete model in the form of a graph $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$ composed of ***V*** nodes, one for each small molecule, and ***E*** edges connecting nodes representing small molecules whose combination resulted in a value of 1 in the adjacency matrix ***A*** (Figure 3).

Of the possible 276 edges, on average there were 98 ± 12 edges in each network with a maximum of 119 observed in the bik1Δ-network and a minimum of 85 observed in the kip3Δ-network (Table 1). None of the deletion strain networks were identical to each other or the WT-network (Figure 3; see Supporting Information for the complete set of networks). These results reveal that the structure of the genetic network determines the structure of the chemical genetic network. To visualize the global similarities/differences in these latter networks, the Fructherman−Reingold (F−R) algorithm was applied.[12,13] This heuristic graph-drawing method considers the network as a physical system composed of mass particles (nodes) repelling each other and springs (edges) attracting adjacent nodes. Minimizing the "energy" of these systems resulted in the center of each network containing the most highly

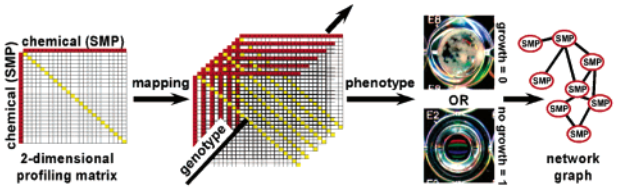


Figure 1. Multidimensional chemical genomic profiling; 276 unique combinations (in duplicate) and 24 single treatments of "small molecule perturbagens" (SMPs) were assayed for an effect on the cell cycle network of *S. cerevisiae*. Each of the 10 strains profiled had a different genotype yielding a three-dimensional matrix of 24 × 24 × 10 observations.

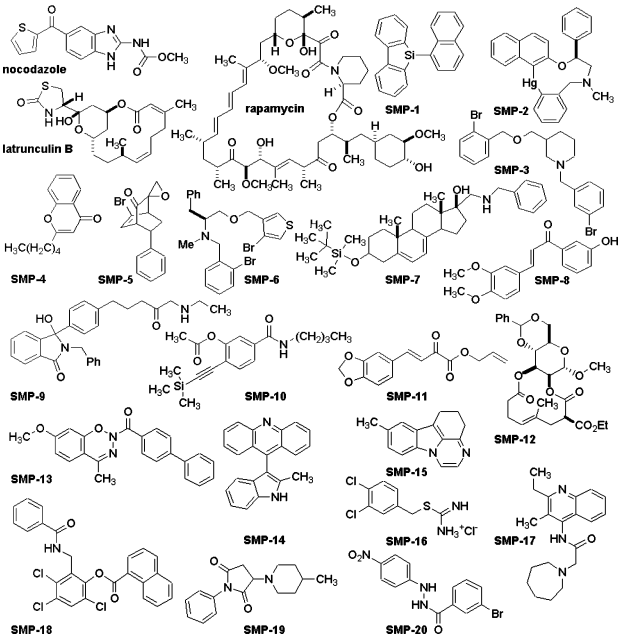


Figure 2. Structures of 23 small molecules (other than DMSO) used to profile 10 yeast genotypes in a three-dimensional matrix.[10]

Table 1. *S. cerevisae* Strains and Graph Theoretic Descriptors Computed for the Networks Obtained by Chemical Genomic Profiling

network	vertices	edges	density[d]	Zagreb[e]	Randic[e]	Platt[e]	spectra max.[f]	mean EIP[f]	low EIP[f]	high EIP[f]
model A[a]	24	0	0	0	0	0	1	1	1	1
model B[b]	24	276	1.00	12696	12.00	12144	1.04	0.25	0.04	0.50
WT	24	103	0.37	2784	9.77	2578	1.68	0.20	0.06	0.50
kip1Δ	24	88	0.32	2704	9.02	2516	1.75	0.23	0.07	0.50
kip2Δ	24	87	0.32	2168	9.27	1994	1.73	0.19	0.07	0.50
kip3Δ	24	85	0.31	2106	9.22	1936	1.74	0.19	0.07	0.57
bni1Δ	24	116	0.42	3278	10.13	3046	1.65	0.18	0.06	0.32
bim1Δ	24	111	0.40	3108	9.90	2886	1.00	0.24	0.06	0.52
bik1Δ	24	119	0.43	3372	10.20	3134	1.64	0.19	0.06	0.50
kar9Δ	24	87	0.32	2160	9.32	1986	1.73	0.20	0.07	0.55
dyn1Δ	24	104	0.38	2816	9.75	2608	1.74	0.19	0.07	0.52
cin8Δ	24	97	0.35	2568	9.51	2374	1.71	0.23	0.07	0.51
average[c]	24	100	0.36	2721.78	9.62	2520.44	1.63	0.28	0.07	0.50

[a] Model A is a $V = 24$ network with $E = 0$. [b] Model B network is a $V = 24$ network with $E = 276$ (complete without self-loops). [c] Average network calculated from mean of the properties from the 10 yeast genotypes. [d] Density calculated as number of observed edges out of a possible 276. [e] The Zagreb, Randic, and Platt indices are measures of connectivity computed using Pajek.[13] [f] The Laplacian matrix,[15] denoted by L, of a graph $G = (V, E)$ is a V x V real symmetric matrix with one row and one column for each node. L is defined by $L = W - A$, where A is a square adjacency matrix and W is a diagonal weight matrix whose entries are row (or column) sums of A. To normalize L, its diagonal values are divided by W such that all diagonal entries have a value of 1, and all off diagonal values are divided by the square root of the product of the entries of W.[15] The Laplacian spectrum is obtained by solving $L\Psi = \Lambda\Psi$, where Ψ represents the matrix of eigenvector components and Λ represents the corresponding matrix of eigenvalues.[16] "Spectra max." denotes the maximum eigenvalue. The eigenvector inverse participation (EIP) ratio is the sum of the fourth powers of the eigenvector's components and is inversely related to the number of eigenvector components significantly different from zero.[15]

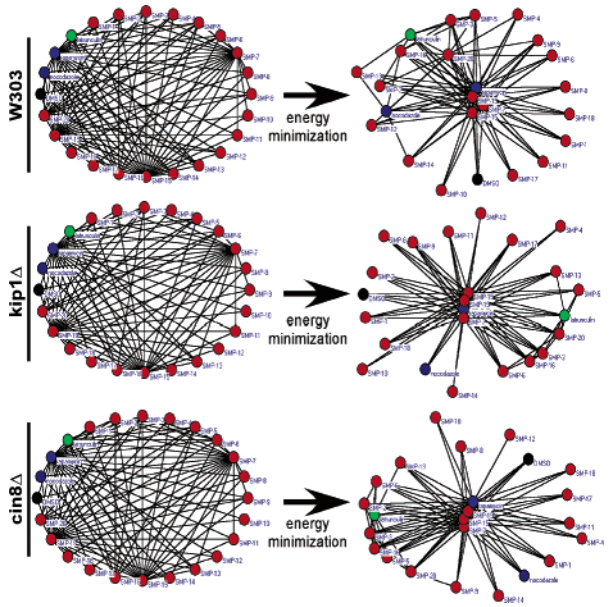


Figure 3. Network graphs G from wild-type (W303) and two yeast deletions strains (cin8Δ and kip1) visualized using the graph-drawing program Pajek.[13] Nodes (colored balls) represent small molecules, and edges (black lines) connect nodes that in combination reduced yeast growth. "Energy" minimization was performed using the F−R algorithm.[12]

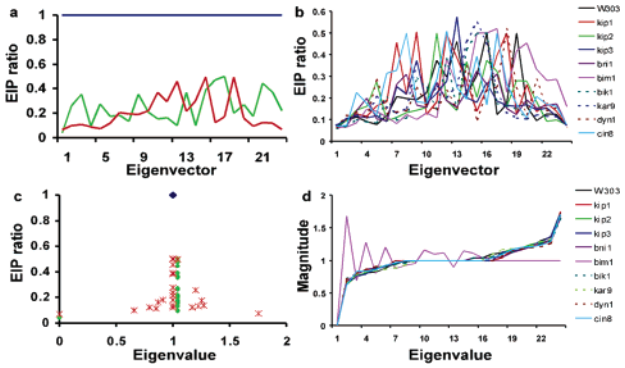


Figure 4. Characterization of chemical networks using properties of the normalized Laplacian matrix (see Table 1 and Supporting Information for details of model graphs and additional graphs).[15,16] (a) The eigenvector inverse participation (EIP) ratios for model network A (green), B (blue), and wild-type yeast (red). (b) EIP ratios for the 10 genotypes showing fluctuations in the degree of localization. (c) EIP ratio as a function of the eigenvalue spectra for model network A (green), B (blue), and WT yeast (red). (d) Genotype-dependent fluctuations in the eigenvalue spectra. Deletion of BIM1, which encodes a microtubule-associated protein, resulted in an eigenvalue spectrum markedly different from those that resulted from the deletion of other components in the same genetic pathway.

connected nodes, which for most, but not all, genotypes included rapamycin, SMP-7, SMP-15, and SMP-19 (Figure 3).

Graph theoretic descriptors that are analogous to molecular descriptors used for the quantitative analysis and comparison of the structures of small molecules[14] were computed for each of the 10 chemical genetic networks (Table 1). The networks were further characterized by deriving the corresponding normalized Laplacian matrix[15] and computing the associated eigenvalues and eigenvectors (Figure 4 and Table 1).[16] For each network, the set of V eigenvalues provides a characteristic "spectrum" with V elements ranging in magnitude from 0 to 2 (Figure 4). Each eigenvector component reveals the contribution of the corresponding node to that particular eigenvector.[15] A value of, or close to, 1 indicates the localization of the eigenvector/eigenvalue to a subset of the graph vertices. Collectively, the numerical values of the descriptors yield a topological fingerprint of each chemical genetic network. This fingerprint provides a higher-level representation of the information inherent in the lower-level relational data obtained from the phenotypic screen. Standard clustering and dimensionality reduction algorithms can then be used to reveal global similarities/differences of the observed chemical genetic networks.[17]

For example, *CIN8* and *KIP1* are genes encoding kinesin-related motor proteins that are known to play an essential, but genetically redundant, role in the organization and function of the mitotic spindle.[3] In agreement with this functional redundancy, hierarchical clustering[17] of the standardized covariance (Pearson correlation) matrix derived from the topological profiles clustered the cin8Δ and kip1Δ genotypes as nearest neighbors (Figure 5a). Similarly, principal component analysis[17] of standardized (Pearson correlation) descriptors positioned the cin8 and kip1Δ deletions as nearest neighbors in a reduced, three-dimensional space maximizing the

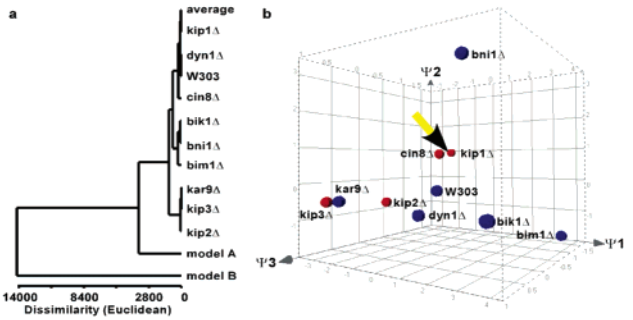


Figure 5. Multidimensional analysis of chemical genomic profiles. (a) Clustering of network graphs in Table 1 using the unweighted pair-group average method and a Euclidean distance metric.[17] (b) Three-dimensional principal component (**Ψ1**−**Ψ3**) model (accounting for 94% of the variance) showing the distances (Euclidean) between networks (red, kinesin and related motor proteins; blue, other gene deletions).[17] The yellow arrow shows the location of cin8Δ and kip1Δ, two motor proteins previously thought to be functionally redundant.

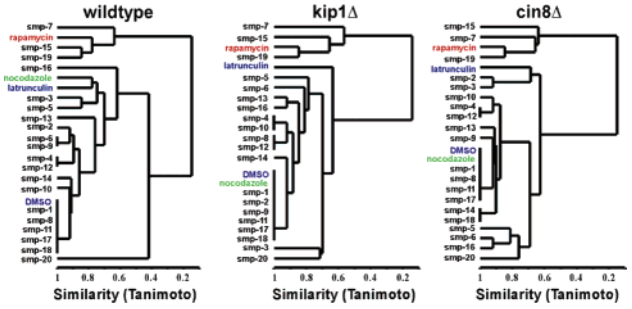


Figure 6. Clustering of small molecules using the unweighted pair-group average method and the Tanimoto distance metric for wild-type and two deletion strains (cin8Δ and kip1Δ).[17] Small molecules connected by a common point of branching are more similar with respect to their effects on proliferation of the particular strain than those not sharing a common branching point.

observed variance (Figure 5b). However, the distance between the cin8Δ/kip1Δ-networks and the kip2-, kip3- (*KIP2* and *KIP3* also encode kinesin-related motor proteins), and WT-networks reveals that the consequences of the different kinesin deletions to pairwise small molecule perturbation are not identical (Figure 5a,b). Further exploration of the mechanistic basis for these differences is now warranted.

The information in ***A*** for each genotype can also be used to cluster the small molecules on the basis of the similarities in their pattern of biological activity. The topologies of the resulting graphs, and thus relationships of the small molecules, were again different for each genotype (Figure 6). However, similar to the results obtained from applying the F−R algorithm (Figure 3), rapamycin, SMP-7, SMP-15, and SMP-19 were clustered closely together in most genotypes. These results reveal that systematically varying the architecture of a genetic network provides a biology-based method for analyzing the diversity of small molecules. The nature of this method renders it complementary to the mapping of chemical space using molecular descriptor analysis.[14] Furthermore, these findings highlight the fact that small genetic differences, here single gene deletions, can translate into significant differences in the effects of many small molecules, as observed with multidrug-resistant tumor cell lines.[2c]

Topological fingerprints obtained by chemical genomic profiling provide computational metrics for analyzing biological networks. Because the outcomes of this method of profiling are dependent upon the interaction of small molecules in the context of an intact genetic network (i.e., perturbations), this method differs fundamentally from profiling methods based upon DNA sequence or mRNA/protein expression patterns (i.e., observations). Besides aiding the characterization of molecular diversity and annotation of chemical space,[8,14] the results herein suggest that chemical genomic profiling may serve as a tool for the characterization of perturbations in biological networks or of the networks themselves (e.g., as a diagnostic tool). These capabilities may lead to new approaches to discern the molecular etiology of highly complex phenotypes, including those involved in human disease.

Acknowledgment. We thank the National Institute for General Medical Sciences for support of this research, the National Cancer Institute (NCI), Merck KGaA, Merck & Co., and the Keck Foundation for support of the ICCB, and the NCI for support of the Initiative for Chemical Genetics. S.L.S. is an Investigator at the Howard Hughes Medical Institute in the Department of Chemistry & Chemical Biology.

Supporting Information Available: Experimental details for the selection and use of SMPs, and the complete data set containing all 10 chemical genetic networks and diversity analyses (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) (a) Lee, T. I.; et al. *Science* **2002**, *298*, 799−804. (b) Ideker, T.; et al. *Science* **2001**, *292*, 929−934. (c) Scherf, U.; et al. *Nat. Genet.* **2000**, *3*, 208−209. (d) Giaever, G.; et al. *Nature* **2002**, *418*, 387−391.
(2) (a) Jeong, H.; Tombor, B.; Albert, R.; Oltvai, Z. N.; Barabasi, A. L. *Nature* **2000**, *407*, 651−654. (b) Farkas, I. J.; Jeong, H.; Vicsek, T.; Barabasi, A.-L.; Oltvai, Z. N. *Physica* **2003**, *A318*, 601−612. (c) Tong, A. H.; et al. *Science* **2001**, *294*, 2364−2368.
(3) (a) Segal, M.; Bloom, K. *Trends Cell Biol.* **2001**, *11*, 160−166. (b) Hoyt, M. A.; He, L.; Loo, K. K.; Saunders, W. S. *J. Cell Biol.* **1992**, *118*, 109−120.
(4) With an $N \times N$ profiling matrix, where N represents the number of small molecules, there are $(N^2 - N)/2$ unique combinations (in duplicate) and N single treatments (on the diagonal of the matrix).
(5) Stock solutions of small molecules in DMSO were first diluted as appropriate (see Supporting Information, Table S1) in yeast media and then pipetted into the wells of 96-well plates.
(6) Cultures were grown in 1% yeast extract (Difco), 2% peptone (Difco), and 2% dextrose (Sigma) added to autoclaved water.
(7) ICCB Diversity Set-2 collection; ChemBridge DiverSetE.
(8) (a) Schreiber, S. L. *Science* **2000**, *287*, 1964−1969. (b) Schreiber, S. L. *Chem. Eng. News* **2003**, *81*, 51−61.
(9) Haggarty, S. J. Ph.D. Thesis, Harvard University, 2003. Haggarty, S. J.; Blackwell, H. E.; Krishnan, S.; Clemons, P. A.; Schreiber, S. L. **2003**, manuscript in preparation.
(10) SMPs were chosen on the basis of their bioactivity pattern in a variety of yeast screens. See Supporting Information, Table S1, for details about concentrations and solubility.
(11) Replicates within the profiling matrix where only one of the two wells showed no or reduced growth (frequency of 1.3% of total observations) were not considered as edges (see Supporting Information, Table S2).
(12) Fruchterman, T. M.; Reingold, E. M. *Soft. Pract. Exp.* **1991**, *21*, 1129−1164.
(13) Pajek v0.72 (http://vlado.fmf.uni-lj.si/pub/networks/pajek/).
(14) Reviewed in: (a) Agrafiotis, D. K.; Myslik, J. P.; Salemme, F. R. In *Annual Reports in Combinatorial Chemistry and Molecular Diversity*; Pavia, M. Moos, W., Eds.; Kluwer: Dordrecht, 1999; Vol. 2, pp 71−92. (b) Gorse, D.; Lahana, R. *Curr. Opin. Chem. Biol.* **2000**, *4*, 287−294.
(15) (a) Chung, F. R. K. *Spectral Graph Theory*; Am. Math. Soc.: Providence 1997; pp 1−22. (b) Farkas, I.; Derenyi, I.; Barabasi, A.-L.; Vicsek, T.; ArXiv:cond-mat/0102335 v3, 2001.
(16) Eigenvalue spectra and eigenvectors of normalized Laplacian matrices were computed using MATLAB 6 (MathWorks, Inc.) and the commands $\boldsymbol{E}$ = eig(LAP); [$\boldsymbol{V}$,$\boldsymbol{E}$] = eig(LAP), where $\boldsymbol{E}$ represents the matrix of eigenvalues, $\boldsymbol{V}$ represents the matrix of eigenvectors, and ***LAP*** represents the normalized Laplacian matrix.
(17) Hierarchical clustering and principal component analysis was performed using XLStat (Addinsoft) and Excel (Microsoft). Three-dimensional visualizations were plotted using Spotfire DecisionSite (Spotfire, Inc).

JA035413P