9. M. L. Rosenzweig, *Species Diversity in Space and Time* (Cambridge Univ. Press, Cambridge, 1995).
10. R. E. Ricklefs, D. Schluter, in *Species Diversity in Ecological Communities*, R. E. Ricklefs, D. Schluter, Eds. (Chicago Univ. Press, Chicago, 1993) , pp. 350–363.
11. C. Rahbek, G. R. Graves, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4534 (2001).
12. R. K. Colwell, D. C. Lees, *Trends Ecol. Evol.* **15**, 70 (2000).
13. D. C. Lees, C. Kremen, L. Andriamampianina, *Biol. J. Linn. Soc.* **67**, 529 (1999).
14. W. Jetz, C. Rahbek, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 5661 (2001).
15. Materials and methods are available as supporting material on *Science* Online.
16. N. A. C. Cressie, *Statistics for Spatial Data* (Wiley, New York, 1993).
17. J. J. Lennon, *Ecography* **23**, 101 (2000).
18. J. R. G. Turner, J. J. Lennon, J. A. Lawrenson, *Nature* **335**, 539 (1988).
19. J. J. Lennon, J. J. D. Greenwood, J. R. G. Turner, *J. Anim. Ecol.* **69**, 581 (2000).
20. R. K. Colwell, G. C. Hurtt, *Am. Nat.* **144**, 570 (1994).
21. M. R. Willig, S. K. Lyons, *Oikos* **81**, 93 (1998).
22. P. Koleff, K. J. Gaston, *Ecography* **24**, 341 (2001).
23. S. Anderson, *Am. Mus. Novit.* **2833**, 1 (1985).
24. D. J. Currie, V. Paquin, *Nature* **329**, 326 (1987).
25. A. Balmford *et al.*, *Science* **291**, 2616 (2001).
26. J. T. Kerr, L. Packer, *Nature* **385**, 252 (1997).
27. H. Elenga *et al.*, *J. Biogeogr.* **27**, 621 (2000).
28. G. R. Graves, *Auk* **102**, 556 (1985).
29. D. Pomeroy, *Conserv. Biol.* **7**, 901 (1993).
30. T. M. Crowe, A. A. Crowe, *J. Zool.* **198**, 417 (1982).
31. A. J. Stattersfield, D. R. Capper, G. C. L. Dutson, *Threatened Birds of the World* (BirdLife International, International Union for Conservation of Nature and Natural Resources; Lynx Edicions, Barcelona, 2000).

32. We thank D. Rogers, B. Bakker, W. Wint, and S. Hay for Geographic Information Systems advice and discussions; I. Woodward for providing productivity data; L. A. Hansen for assistance with access to the African bird database on behalf of the Zoological Museum University of Copenhagen (ZMUC); and E. Baker, N. Baker, F. Dowsett-Lemaire, R. Dowsett, J. Fjeldså, M. E. Gartshore, H. M. de Klerk, M. Languy, R. B. Payn, COC/BirdLife Cameroon, and BirdLife International for providing data for the ZMUC database. The manuscript benefited tremendously from discussions with R. K. Colwell and comments from J. Brown, J. Fjeldså, R. Freckleton, K. Gaston, N. J. Gotelli, G. R. Graves, P. Harvey, R. Lande, O. Lewis, R. May, I. Owens, and D. Rogers. W.J. was supported by Natural Environment Research Council and German Scholarship Foundation studentships; C.R. was supported by the Danish National Science Foundation (grant J. nr. 21-01-0547).

Supporting Online Material
www.sciencemag.org/cgi/content/full/297/5586/1548/DC1
Materials and Methods
Supporting Text
Tables S1 to S3

11 April 2002; accepted 25 July 2002

Hierarchical Organization of Modularity in Metabolic Networks

E. Ravasz,[1] A. L. Somera,[2] D. A. Mongru,[2] Z. N. Oltvai,[2]* A.-L. Barabási[1]*

Spatially or chemically isolated functional modules composed of several cellular components and carrying discrete functions are considered fundamental building blocks of cellular organization, but their presence in highly integrated biochemical networks lacks quantitative support. Here, we show that the metabolic networks of 43 distinct organisms are organized into many small, highly connected topologic modules that combine in a hierarchical manner into larger, less cohesive units, with their number and degree of clustering following a power law. Within *Escherichia coli*, the uncovered hierarchical modularity closely overlaps with known metabolic functions. The identified network architecture may be generic to system-level cellular organization.

The identification and characterization of system-level features of biological organization is a key issue of postgenomic biology (*1–3*). The concept of modularity assumes that cellular functionality can be seamlessly partitioned into a collection of modules. Each module is a discrete entity of several elementary components and performs an identifiable task, separable from the functions of other modules (*1, 4–8*). Spatially and chemically isolated molecular machines or protein complexes (such as ribosomes and flagella) are prominent examples of such functional units, but more extended modules, such as those achieving their isolation through the initial binding of a signaling molecule (*9*), are also apparent.

Simultaneously, it is recognized that the thousands of components of a living cell are dynamically interconnected, so that the cell's functional properties are ultimately encoded into a complex intracellular web of molecular interactions (*2–6, 8*). This is perhaps most evident with cellular metabolism, a fully connected biochemical network in which hundreds of metabolic substrates are densely integrated through biochemical reactions. Within this network, however, modular organization (i.e., clear boundaries between subnetworks) is not immediately apparent. Indeed, recent studies have demonstrated that the probability that a substrate can react with *k* other substrates [the degree distribution *P*(*k*) of a metabolic network] decays as a power law $P(k) \sim k^{-\gamma}$ with $\gamma \cong 2.2$ in all organisms (*10, 11*), suggesting that metabolic networks have a scale-free topology (*12*). A distinguishing feature of such scale-free networks is the existence of a few highly connected nodes (e.g., pyruvate or coenzyme A), which participate in a very large number of metabolic reactions. With a large number of links, these hubs integrate all substrates into a single, integrated web in which the existence of fully separated modules is prohibited by definition (Fig. 1A).

Yet, the dilemma of a modular versus a highly integrated module-free metabolic network organization remains. A number of approaches for analyzing the functional capabilities of metabolic networks indicate the existence of separable functional elements (*13, 14*). Also, from a purely topologic perspective, the metabolic network of *Escherichia coli* is known to possess a high clustering coefficient (*11*), a property that is suggestive of a modular organization. In itself, this implies that the metabolism of *E. coli* has a modular topology, potentially comprising several densely interconnected functional modules of varying sizes that are connected by few intermodule links (Fig. 1B). However, such clear-cut modularity imposes severe restrictions on the degree distribution, implying that most nodes have approximately the same number of links, which contrasts with the metabolic network's scale-free nature (*10, 11*).

To determine whether such a dichotomy is indeed a generic property of all metabolic networks, we first calculated the average clustering coefficient for 43 different organisms (*10, 15, 16*) as a function of the number of distinct substrates *N* present in their metabolism. The clustering coefficient, defined as $C_i = 2n/k_i(k_i - 1)$, where *n* denotes the number of direct links connecting the k_i nearest neighbors of node *i* (*17*), is equal to 1 for a node at the center of a fully interlinked cluster, and it is 0 for a metabolite that is part of a loosely connected group (Fig. 2A). Therefore, C_i averaged over all nodes *i* of a metabolic network is a measure of the network's potential modularity. We found that, for all 43 organisms, the average clustering coefficient is about an order of magnitude larger than that expected for a scale-free network of similar size (Fig. 2B), suggesting that metabolic networks in all organisms are characterized by a high intrinsic potential modularity. We also observed that, in contrast with

---

[1] Department of Physics, University of Notre Dame, Notre Dame, IN 46556, USA. [2] Department of Pathology, Northwestern University, Chicago, IL 60611, USA.

*To whom correspondence should be addressed. E-mail: zno008@northwestern.edu (Z.N.O.) and alb@nd.edu (A.-L.B.)

the prediction of the scale-free model, for which the clustering coefficient decreases as $N^{-0.75}$ (*18*), the clustering coefficient of metabolic networks is independent of their size (Fig. 2B).

These results demonstrate a fundamental conflict between the predictions of the current models of metabolic organization. The high, size-independent clustering coefficient offers strong evidence for modularity, whereas the power law degree distribution of all metabolic networks (*10, 11*) strongly supports the scale-free model and rules out a manifestly modular topology. To resolve this apparent contradiction, we propose a simple heuristic model of metabolic organization, which we refer to as a "hierarchical" network (Fig. 1C) (*19*). In such a model network, our starting point is a small cluster of four densely linked nodes. Next, we generate three replicas of this hypothetical module and connect the three external nodes of the replicated clusters to the central node of the old cluster, obtaining a large 16-node module. Subsequently, we again generate three replicas of this 16-node module and connect the peripheral nodes to the central node of the old module (Fig. 1C). These replication and connection steps can be repeated indefinitely, in each step quadrupling the number of nodes in the system. The architecture of such a network integrates a scale-free topology with an inherent modular structure. It has a power law degree distribution with degree exponent $\gamma = 1 + (\ln 4)/(\ln 3) = 2.26$, in agreement with $\gamma = 2.2$ observed in metabolic networks. Its clustering coefficient $C \cong 0.6$ is also comparable with that observed for metabolic networks. Most important, the clustering coefficient of the model is independent of the size of the network, in agreement with the results of Fig. 2B.

A unique feature of the proposed network model, not shared by either the scale-free (Fig. 1A) or modular (Fig. 1B) models, is its hierarchical architecture. This hierarchy, which is evident from a visual inspection, is intrinsic to the assembly by repeated quadrupling of the system. The hierarchy can be characterized quantitatively by using the recent observation (*20*) that, in deterministic scale-free networks, the clustering coefficient of a node with k links follows the scaling law $C(k) \sim k^{-1}$. This scaling law quantifies the coexistence of a hierarchy of nodes with different degrees of modularity, as measured by the clustering coefficient, and is directly relevant to our model (Fig. 1C). Indeed, the nodes at the center of the numerous 4-node modules have a clustering coefficient $C = 3/4$, those at the center of a 16-node module have $k = 13$ and $C = 2/13$, and those at the center of the 64-node modules have $k = 40$ and $C = 2/40$, indicating that the higher a node's connectivity, the smaller its clustering coefficient, asymptotically following the $1/k$ law.

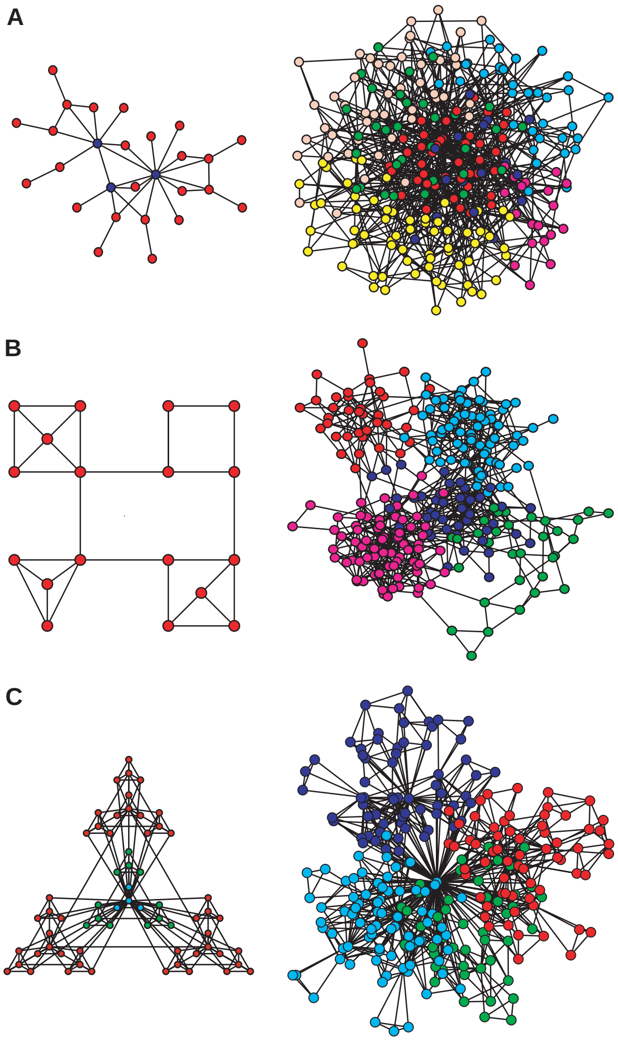

Fig. 1. Complex network models. (**A**) A schematic illustration (left) of a scale-free network, whose degree distribution follows a power law. In such a network, a few highly connected nodes, or hubs (blue circles), play an important role in keeping the whole network together. A typical configuration (right) of a scale-free network with 256 nodes is also shown, obtained using the scale-free model, which requires the addition of a new node at each time such that existing nodes with higher degrees of connectivity have a higher chance of being linked to the new nodes (*12*). The nodes are arranged in space with a standard clustering algorithm (*30*) to illustrate the absence of an underlying modularity. (**B**) Schematic illustration (left) of a manifestly modular network made of four highly interlinked modules connected to each other by a few links. This intuitive topology does not have a scale-free degree distribution, as most of its nodes have a similar number of links, and hubs are absent. A standard clustering algorithm uncovers the network's inherent modularity (right) by partitioning a modular network of $N = 256$ nodes into the four isolated structures built into the system. (**C**) The hierarchical network (left) has a scale-free topology with embedded modularity. The hierarchical levels are represented in increasing order from blue to green to red. Standard clustering algorithms (right) are less successful in uncovering the network's underlying modularity. A detailed quantitative characterization of the three network models is available in (*16*).

To investigate whether such hierarchical organization is present in cellular metabolism, we measured the $C(k)$ function for the metabolic networks of all 43 organisms. As shown in Fig. 2, C through F, for each organism, $C(k)$ is well approximated by $C(k) \sim k^{-1}$, in contrast to the k-independent $C(k)$ predicted by both the scale-free and modular networks. This provides direct evidence for an inherently hierarchical organization. Such hierarchical modularity reconciles within a single framework all the observed properties of metabolic networks: their scale-free topology; high, system size–independent clustering coefficient; and the power law scaling of $C(k)$.

A key issue from a biological perspective is whether the identified hierarchical architecture reflects the true functional organization of cellular metabolism. To uncover potential relations between topological modularity and the functional classification of different metabolites, we concentrated on the metabolic network of *E. coli*, whose metabolic reactions have been exhaustively studied, both biochemically and genetically (*21*). Using a previously established graph-theoretical representation (*10*), we first subjected *E. coli*'s metabolic organization to a three-step reduction process, replacing nonbranching pathways with equivalent links, allowing us to decrease its complexity without altering the network topology (*16*). Next, we calculated the topological overlap matrix $O_T(i, j)$ of the condensed metabolic network (Fig. 3A). A topological overlap of 1 between substrates i and j implies that they are connected to the same substrates, whereas a 0 value indicates that i and j do not share links to common substrates among the metabolites they react with. The metabolites that are part of highly integrated modules have a high topological overlap with their neighbors, and we found that the larger the overlap between two substrates within the *E. coli* metabolic network, the more likely it is that they belong to the same functional class.

As the topological overlap matrix is expected to encode the comprehensive enzyme catalyzed functional relatedness of the substrates forming the metabolic network, we investigated whether potential functional modules encoded in the network topology can be uncovered automatically. Initial application of an average-linkage hierarchical clustering algorithm (*22*) to the overlap matrix of the small hypothetical network shown in Fig. 3A placed those nodes that have a high topological overlap close to each other (Fig. 3B). Also, the method identified the three distinct modules built into the model of Fig. 3A, as illustrated by the fact that the *EFG* and *HIJK* modules are closer to each other in a topological sense, with the *ABC* module being farther from both (Fig. 3B). Application of the same technique on the *E. coli* overlap matrix $O_T(i, j)$ provides a global topologic representation of *E. coli* metabolism (Fig. 4A). Groups of metabolites forming tightly interconnected clusters are visually apparent, and on closer inspection, the hierarchy of nested topologic modules of increasing sizes and decreasing interconnectedness is also evident. To visualize the relation between topological modules and the known functional properties of the metabolites, we color-coded the branches of the derived hierarchical tree according to the predominant biochemical class of the substrates it produces, using the classification of metabolism based on standard, small molecule biochemistry (*15*). As shown in Fig. 4A, and in the three-dimensional representation in Fig. 4B, most substrates of a given small molecule class are distributed on the same branch of the tree (Fig. 4A) and correspond to relatively well delimited regions of the metabolic network (Fig. 4B). Therefore, there are strong correlations between shared biochemical classification of metabolites and the global topological organization of *E. coli* metabolism (Fig. 4A, bottom) (*16*).

To correlate the putative modules obtained from our graph theory–based analysis to actual biochemical pathways, we concentrated on the pathways involving the pyrimidine metabolites. Our method divided these pathways into four putative modules (Fig. 4C), which represent a topologically well-limited area of *E. coli* metabolism (Fig. 4B, blue-shaded region). As shown in Fig. 4D, all highly connected metabolites (Fig. 4D, red-outlined boxes) correspond to their respective biochemical reactions within pyrimidine metabolism, together with those substrates that were removed during the original network reduction procedure, and then added again (Fig. 4D, green-outlined boxes). However, it is also apparent that putative module boundaries do not always overlap with intuitive "biochemistry-based" boundaries. For instance, the synthesis of uridine 5′-mono-

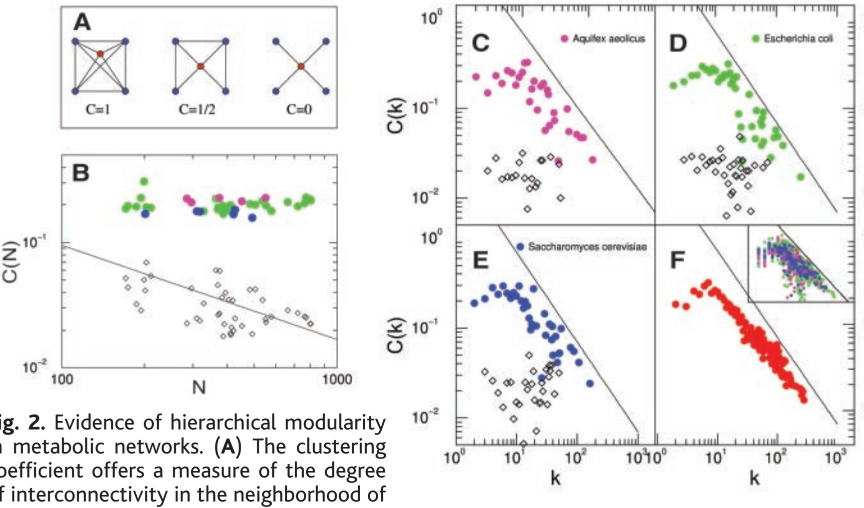

Fig. 2. Evidence of hierarchical modularity in metabolic networks. (**A**) The clustering coefficient offers a measure of the degree of interconnectivity in the neighborhood of a node (*17*). For example, a node whose neighbors are all connected to each other has $C = 1$ (left), whereas a node with no links between its neighbors has $C = 0$ (right). (**B**) The average clustering coefficient $C(N)$ for 43 organisms (*10*) is shown as a function of the number of substrates N present in each of them. Species belonging to archaea (purple), bacteria (green), and eukaryotes (blue) are shown. The dashed line indicates the dependence of the clustering coefficient on the network size for a module-free scale-free network, and the diamonds denote C for a scale-free network with the same parameters (N and number of links) as observed in the 43 organisms. (**C** through **E**) The dependence of the clustering coefficient on the node's degree in three organisms: *Aquidex aeolicus* (archaea) (C), *Escherichia coli* (bacterium) (D), and *Saccharomyces cerevisiae* (eukaryote) (E). (**F**) The $C(k)$ curves averaged over all 43 organisms is shown, and the inset displays all 43 species together. The data points are color coded as in Fig. 2B. In (C through F), the dashed lines correspond to $C(k) \sim k^{-1}$, and in (C through E), the diamonds represent the $C(k)$ value expected for a scale-free network (Fig. 1A) of similar size, indicating the absence of scaling. The wide fluctuations are due to the small size of the network.

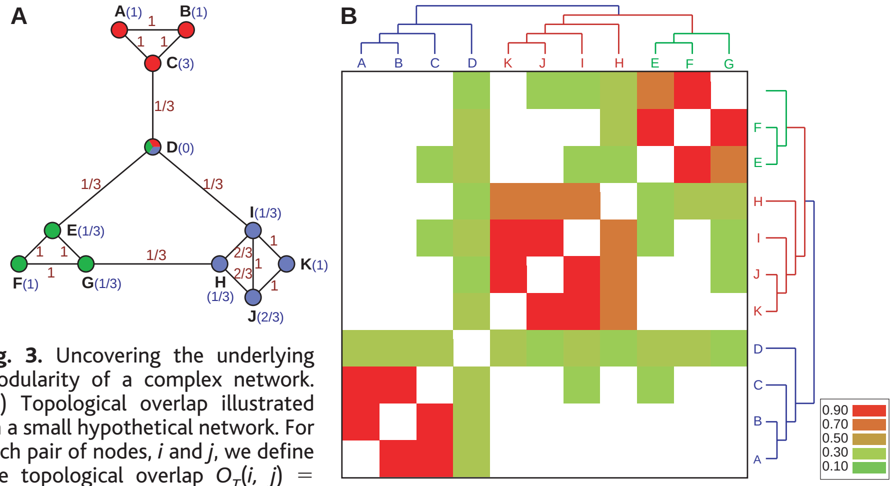

Fig. 3. Uncovering the underlying modularity of a complex network. (**A**) Topological overlap illustrated on a small hypothetical network. For each pair of nodes, i and j, we define the topological overlap $O_T(i, j) = J_n(i, j)/[\min(k_i, k_j)]$, where $J_n(i, j)$ denotes the number of nodes to which both i and j are linked (plus 1 if there is a direct link between i and j) and $[\min(k_i, k_j)]$ is the smaller of the k_i and k_j degrees. On each link, we indicate the topological overlap for the connected nodes, and in parentheses next to each node, we indicate the node's clustering coefficient. (**B**) The topological overlap matrix corresponding to the small network shown in (A). The rows and columns of the matrix were reordered by the application of an average linkage clustering method (*22*) to its elements, allowing us to identify and place close to each other those nodes that have high topological overlap. The color code denotes the degree of topological overlap between the nodes. The associated tree reflects the three distinct modules built into the model of Fig. 3A, as well as the fact that the *EFG* and *HIJK* modules are closer to each other in the topological sense than to the *ABC* module.

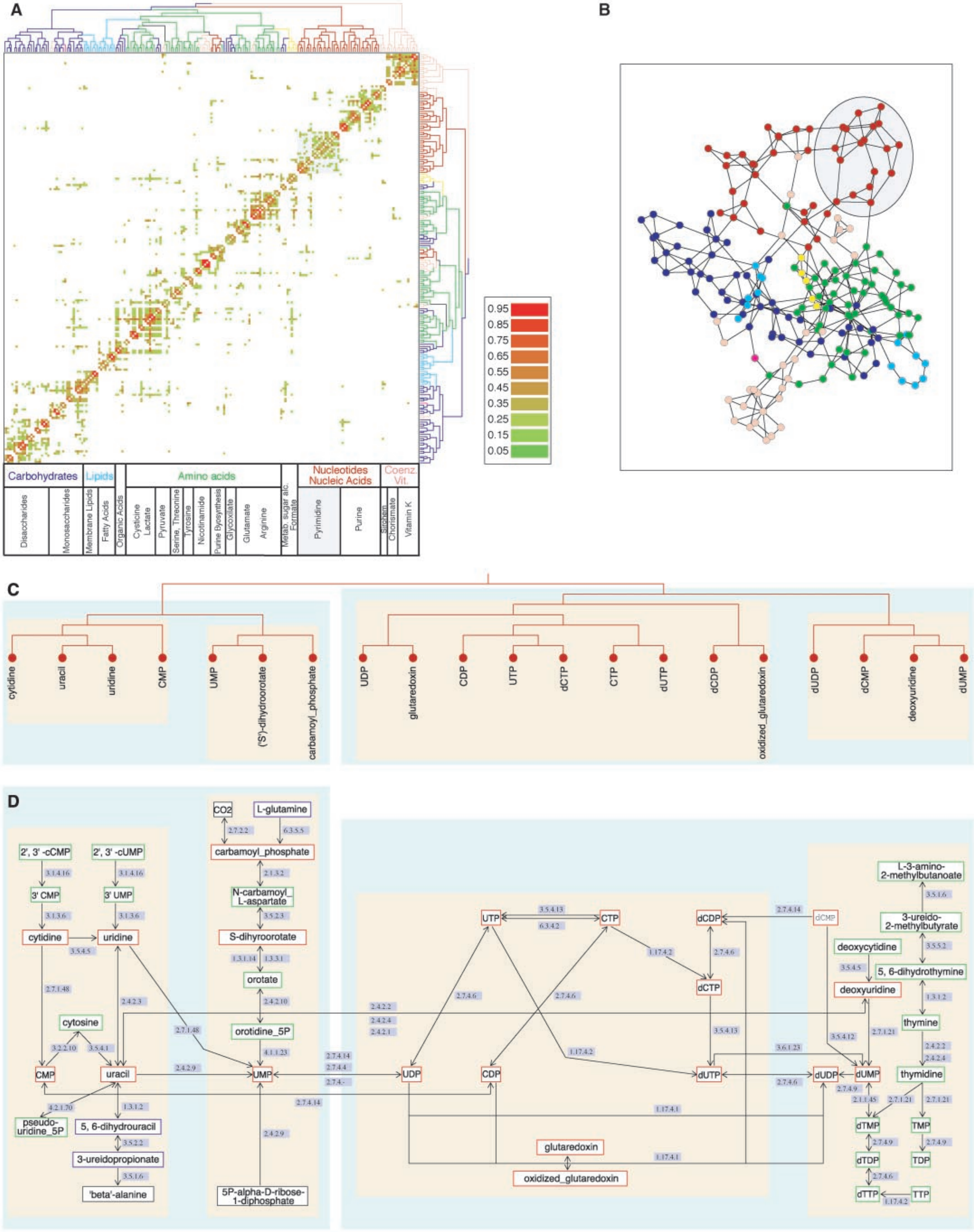
A
B
0.95
0.85
0.75
0.65
0.55
0.45
0.35
0.25
0.15
0.05
Carbohydrates
Lipids
Amino acids
Nucleotides
Nucleic Acids
Coenz.
Vit.
Disaccharides
Monosaccharides
Membrane Lipids
Fatty Acids
Organic Acids
Cysticine
Lactate
Pyruvate
Serine, Threonine
Tyrosine
Nicotinamide
Purine Biosynthesis
Glyoxilate
Glutamate
Arginine
Metab. sugar alc.
Formate
Pyrimidine
Purine
Sirohem
Chorismate
Vitamin K
C
cytidine
uracil
uridine
CMP
UMP
('S')-dihydroorotate
carbamoyl_phosphate
UDP
glutaredoxin
CDP
UTP
dCTP
CTP
dUTP
dCDP
oxidized_glutaredoxin
dUDP
dCMP
deoxyuridine
dUMP
D
2', 3' -cCMP
2', 3' -cUMP
3.1.4.16
3.1.4.16
3' CMP
3' UMP
3.1.3.6
3.1.3.6
cytidine
uridine
3.5.4.5
2.7.1.48
2.4.2.3
cytosine
3.2.2.10
3.5.4.1
CMP
uracil
4.2.1.70
1.3.1.2
pseudo-uridine_5P
5, 6-dihydrouracil
3.5.2.2
3-ureidopropionate
3.5.1.6
'beta'-alanine
CO2
L-glutamine
2.7.2.2
6.3.5.5
carbamoyl_phosphate
2.1.3.2
N-carbamoyl_L-aspartate
3.5.2.3
S-dihyroorotate
1.3.1.14
1.3.3.1
orotate
2.4.2.10
orotidine_5P
2.7.1.48
4.1.1.23
2.4.2.9
UMP
2.4.2.9
5P-alpha-D-ribose-1-diphosphate
2.7.4.14
2.7.4.4
2.7.4.-
2.7.4.14
2.4.2.2
2.4.2.4
2.4.2.1
UTP
CTP
3.5.4.13
6.3.4.2
2.7.4.6
2.7.4.6
1.17.4.2
1.17.4.2
UDP
CDP
dCDP
2.7.4.6
dCTP
3.5.4.13
dUTP
glutaredoxin
oxidized_glutaredoxin
1.17.4.1
1.17.4.1
2.7.4.14
dCMP
3.6.1.23
2.7.4.6
dUDP
dUMP
2.7.4.9
L-3-amino-2-methylbutanoate
3.5.1.6
3-ureido-2-methylbutyrate
3.5.5.2
deoxycytidine
3.5.4.5
5, 6-dihydrothymine
deoxyuridine
1.3.1.2
thymine
3.5.4.12
2.7.1.21
2.4.2.2
2.4.2.4
thymidine
2.1.1.45
2.7.1.21
2.7.1.21
dTMP
TMP
2.7.4.9
2.7.4.9
dTDP
TDP
2.7.4.6
dTTP
TTP
1.17.4.2

Fig. 4. Identifying the topological modules in *E. coli* metabolism. (**A**) The topologic overlap matrix corresponding to *E. coli* metabolism, together with the corresponding hierarchical tree (top) that quantifies the relation between the different modules. The branches of the tree are color coded to reflect the predominant biochemical classification of their substrates. The biochemical classes that we used to group the metabolites represent carbohydrate metabolism (blue); nucleotide and nucleic acid metabolism (red); protein, peptide, and amino acid metabolism (green); lipid metabolism (cyan); aromatic compound metabolism (dark pink); monocarbon compound metabolism (yellow); and coenzyme metabolism (light orange) (*15*). The color code of the matrix denotes the degree of topological overlap shown in the matrix. The large-scale functional map of the metabolism, as suggested by the hierarchical tree, is also shown (bottom). (**B**) Three-dimensional representation of the reduced *E. coli* metabolic network. Each node is color coded by the predominant biochemical class to which it belongs and is identical to the color code applied to the branches of the tree shown in (A). The different functional classes are visibly segregated into topologically distinct regions of metabolism. The blue-shaded region denotes the nodes belonging to pyrimidine metabolism. (**C**) Enlarged view of the substrate module of pyrimidine metabolism. The colored boxes denote the first two levels of the three levels of nested modularity suggested by the hierarchical tree. CDP, cytidine 5′-diphosphate; CMP, cytidine 5′-monophosphate; CTP, cytidine 5′-triphosphate; dCDP, deoxycytidine 5′-diphosphate; dCMP, deoxycytidine 5′-monophosphate; dCTP, deoxycytidine 5′-triphosphate; dUDP, deoxyuridine 5′-diphosphate; dUMP, deoxyuridine 5′-monophosphate; dUTP, deoxyuridine 5′-triphosphate; UTP, uridine 5′-triphosphate. (**D**) A detailed diagram of the metabolic reactions that surround and incorporate the pyrimidine metabolic module. Red-outlined boxes denote the substrates directly appearing in the reduced metabolism and the tree shown in (C). Substrates in green-outlined boxes are internal to pyrimidine metabolism but represent members of nonbranching pathways or end pathways branching from a metabolite with multiple connections (*16*). Blue- and black-outlined boxes show the connections of pyrimidine metabolites to other parts of the metabolic network. Black-outlined boxes denote core substrates belonging to other branches of the metabolic tree (A), and blue-outlined boxes denote nonbranching pathways (if present) leading to those substrates. With the exception of carbamoyl phosphate and *S*-dihydroorotate, all pyrimidine metabolites are connected with a single biochemical reaction. The shaded boxes around the reactions highlight the modules suggested by the hierarchical tree. The shaded blue boxes along the links display the enzymes catalyzing the corresponding reactions, and the arrows show the direction of the reactions according to theWIT metabolic maps (*15*). cCMP, cyclic cytidine 5′-monophosphate; cUMP, cyclic uridine 5′-monophosphate; dTDP, deoxythymidine 5′-diphosphate; dTMP, deoxythymidine 5′-monophosphate; dTTP, deoxythymidine 5′-triphosphate; TDP, thymidine diphosphate; TMP, thymidine monophosphate; TTP, thymidine triphosphate.

phosphate (UMP) from L-glutamine is expected to fall within a single module based on a linear set of biochemical reactions, whereas the synthesis of uridine 5′-diphosphate from UMP leaps putative module boundaries. Thus, further experimental and theoretical analyses will be needed to understand the relation between the decomposition of *E. coli* metabolism offered by our topology-based approach and the biologically relevant subnetworks.

The organization of metabolic networks is likely to combine a capacity for rapid flux reorganization with a dynamic integration with all other cellular function (*11*). Here we show that the system-level structure of cellular metabolism is best approximated by a hierarchical network organization with seamlessly embedded modularity. In contrast to current, intuitive views of modularity (Fig. 1B), which assume the existence of a set of modules with a nonuniform size potentially separated from other modules, we find that the metabolic network has an inherent self-similar property: There are many highly integrated small modules, which group into a few larger modules, which in turn can be integrated into even larger modules. This is supported by visual inspection of the derived hierarchical tree (Fig. 4A), which offers a natural breakdown of metabolism into several large modules, which are further partitioned into smaller, but more integrated submodules.

The mathematical framework proposed here to uncover the presence or absence of such hierarchical modularity and to delineate the modules based on the network topology could apply to other cellular and complex networks as well. As scale-free topology has been found at many different organizational levels, ranging from genetic (*23*) to protein interaction and protein domain (*24*) networks, it is possible that biological networks are always accompanied by a hierarchical modularity. Some nonbiological networks, ranging from the World Wide Web to the Internet, often combine a scale-free topology with a community structure (i.e., modularity) (*25–27*); therefore, these networks are also potential candidates for hierarchical modularity. For biological systems, hierarchical modularity is consistent with the notion that evolution may act at many organizational levels simultaneously: The accumulation of many local changes, which affect the small, highly integrated modules, could slowly impact the properties of the larger, less integrated modules. The emergence of the hierarchical topology through copying and reusing existing modules (*1*) and motifs (*8*), a process reminiscent of the results of gene duplication (*28, 29*), offers a special role to the modules that appeared first in the network. Although the model of Fig. 1C reproduces the large-scale features of the metabolism, understanding the evolutionary mechanism that explains the simultaneous emergence of the observed hierarchical and scale-free topology of the metabolism, as well as its generality to cellular organization, is now a prime challenge.

References and Notes

1. L. H. Hartwell, J. J. Hopfield, S. Leibler, A. W. Murray, *Nature* **402**, C47 (1999).
2. H. Kitano, *Science* **295**, 1662 (2002).
3. Y. I. Wolf, G. Karev, E. V. Koonin, *Bioessays* **24**, 105 (2002).
4. D. A. Lauffenburger, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5031 (2000).
5. C. V. Rao, A. P. Arkin, *Annu. Rev. Biomed. Eng.* **3**, 391 (2001).
6. N. S. Holter, A. Maritan, M. Cieplak, N. V. Fedoroff, J. R. Banavar, *Proc. Natl. Acad. Sci. U.S.A.* **98,** 1693 (2001).
7. J. Hasty, D. McMillen, F. Isaacs, J. J. Collins, *Nature Rev. Genet.* **2**, 268 (2001).
8. S. S. Shen-Orr, R. Milo, S. Mangan, U. Alon, *Nature Genet.* **31**, 64 (2002).
9. U. Alon, M. G. Surette, N. Barkai, S. Leibler, *Nature* **397**, 168 (1999).
10. H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, A.-L. Barabási, *Nature* **407**, 651 (2000).
11. A. Wagner, D. A. Fell, *Proc. R. Soc. London Ser. B* **268**, 1803 (2001).
12. A.-L. Barabási, R. Albert, *Science* **286**, 509 (1999).
13. C. H. Schilling, D. Letscher, B. O. Palsson, *J. Theor. Biol.* **203**, 229 (2000).
14. S. Schuster, D. A. Fell, T. Dandekar, *Nature Biotechnol.* **18**, 326 (2000).
15. R. Overbeek *et al.*, *Nucleic Acids Res.* **28**, 123 (2000).
16. Additional information is available at www.nd.edu/~networks/cell/index.html.
17. D. J. Watts, S. H. Strogatz, *Nature* **393**, 440 (1998).
18. R. Albert, A. L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).
19. A.-L. Barabási, E. Ravasz, T. Vicsek, *Physica A* **299**, 559 (2001).
20. S. N. Dorogovtsev, A. V. Goltsev, J. F. F. Mendes, *Phys. Rev. E* **65**, 066122 (2002).
21. P. D. Karp, M. Riley, S. M. Paley, A. Pellegrini-Toole, M. Krummenacker, *Nucleic Acids Res.* **30**, 56 (2002).
22. M. B. Eisen, P. T. Spellman, P. O. Brown, D. Botstein, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14863 (1998).
23. D. E. Featherstone, K. Broadie, *Bioessays* **24**, 267 (2002).
24. S. Wuchty, *Mol. Biol. Evol.* **18**, 1694 (2001).
25. G. W. Flake, S. Lawrence, C. L. Giles, F. M. Coetzee, *Computer* **53**, 66 (2002).
26. M. Girvan, M. E. J. Newman, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7821 (2002).
27. A. Vazquez, R. Pastor-Satorras, A. Vespignani, *Phys. Rev. E* **65**, 066130 (2002).
28. A. Vazquez, A. Flammini, A. Maritan, A. Vespignani, in press (available at http://arxiv.org/abs/cond-mat/0108043).
29. R. V. Solé, R. Pastor-Satorras, E. D. Smith, T. Kepler, *Adv. Compl. Sys.* **5**, 43 (2002).
30. The algorithm package is available at http://vlado.fmf.uni-lj.si/pub/networks/pajek/.
31. We thank T. Vicsek and I. J. Farkas for discussions on the hierarchical model, C. Waltenbaugh and J. W. Campbell for comments on the manuscript, H. Jeong for help with analyzing the metabolic databases, and the WIT project for making its database available. Research at the University of Notre Dame and at Northwestern University was supported by grants from the U.S. Department of Energy and NIH.

Supporting Online Material
www.sciencemag.org/cgi/content/full/297/5586/1551/DC1
Materials and Methods
SOM Text
Figs. S1 to S15

29 April 2002; accepted 18 July 2002