Science
AAAS

Three-Dimensional Structural View of the Central Metabolic Network of *Thermotoga maritima*

Ying Zhang, *et al.*

Science **325**, 1544 (2009);
DOI: 10.1126/science.1174671

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005.

associated with deep gas or oil-bearing deposits (*23*). 16*S* rRNA gene sequence comparisons revealed that taxa enriched in our 50°C experiments (Fig. 3) are most closely related to bacteria from subsurface petroleum reservoirs or oil production facilities (94 to 96% similarity; table S3). Another source of thermophiles could be nearby mid-ocean ridge spreading centers (fig. S1). Large volumes of fluids circulating through ocean crust (*25*) could transport cells away from warm anoxic niches in this seafloor habitat (*26*) and suspend them in abyssal currents. The closest relatives to the Arctic thermophiles also include an anaerobic thermophile isolated from deep, hot crustal fluid (94% similarity; table S3) (*27*).

Petroleum-bearing sediments and fractured ocean crust both host anaerobic heterotrophic microbial communities (*26*, *28*). Areas of discharge connecting these habitats to the water column are widespread, and both processes expel large volumes of fluid into the oceans (*23*, *25*). A combination of different point sources could explain the diversity and distribution of thermophilic taxa in Arctic sediments (Fig. 3 and fig. S1). Although our observations suggest that seabed fluid flow governs the biogeography of thermophilic spore formers, these passive dispersal mechanisms are unlikely to act only on these particular bacteria. Permeable conduits through sediments and ocean crust pass through several microbial niches with changing local temperature and geochemistry (*22*, *25*, *26*). Widespread seeding of the oceans by geofluids from deep biosphere habitats may therefore contribute broadly to the high microbial diversity observed in the marine environment.

References and Notes

1. C. Pedrós-Alió, *Trends Microbiol.* **14**, 257 (2006).
2. M. L. Sogin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 12115 (2006).
3. J. A. Huber *et al.*, *Science* **318**, 97 (2007).
4. L. G. M. Baas-Becking, *Geobiologie of inleiding Tot de Milieukunde* (Van Stockum & Zoon, The Hague, Netherlands, 1934).
5. A. A. Egorova, *C. R. (Dokl.). Acad. Sci. USSR* **19**, 649 (1938).
6. R. H. McBee, V. H. McBee, *J. Bacteriol.* **71**, 182 (1956).
7. J. W. Bartholomew, G. Paik, *J. Bacteriol.* **92**, 635 (1966).
8. M. F. Isaksen, F. Bak, B. B. Jørgensen, *FEMS Microbiol. Ecol.* **14**, 1 (1994).
9. Y. J. Lee *et al.*, *Extremophiles* **9**, 375 (2005).
10. V. Vandieken, C. Knoblauch, B. B. Jørgensen, *Int. J. Syst. Evol. Microbiol.* **56**, 687 (2006).
11. K. O. Stetter *et al.*, *Nature* **365**, 743 (1993).
12. K. Sahm, B. J. MacGregor, B. B. Jørgensen, D. A. Stahl, *Environ. Microbiol.* **1**, 65 (1999).
13. Materials and methods are available as supporting material on *Science* Online.
14. J. Sagemann, B. B. Jørgensen, O. Greef, *Geomicrobiol. J.* **15**, 85 (1998).
15. N. Finke, B. B. Jørgensen, *ISME J.* **2**, 815 (2008).
16. C. Knoblauch, B. B. Jørgensen, *Environ. Microbiol.* **1**, 457 (1999).
17. K. Ravenschlag, K. Sahm, R. Amann, *Appl. Environ. Microbiol.* **67**, 387 (2001).
18. P. Setlow, *Curr. Opin. Microbiol.* **6**, 550 (2003).
19. K. Alain *et al.*, *Int. J. Syst. Evol. Microbiol.* **52**, 1621 (2002).
20. J. Detmers, V. Brüchert, K. S. Habicht, J. Kuever, *Appl. Environ. Microbiol.* **67**, 888 (2001).
21. C. Arnosti, B. B. Jørgensen, *Geomicrobiol. J.* **23**, 551 (2006).
22. R. J. Parkes *et al.*, *Org. Geochem.* **38**, 845 (2007).
23. A. G. Judd, M. Hovland, *Seabed Fluid Flow: The Impact of Geology, Biology and the Marine Environment* (Cambridge Univ. Press, Cambridge, 2007).
24. E. Damm, A. Mackensen, G. Budéus, E. Faber, C. Hanfland, *Cont. Shelf Res.* **25**, 1453 (2005).
25. M. Hutnak *et al.*, *Nat. Geosci.* **1**, 611 (2008).
26. J. P. Cowen, *Res. Microbiol.* **155**, 497 (2004).
27. J. A. Huber, P. Johnson, D. A. Butterfield, J. A. Baross, *Environ. Microbiol.* **8**, 88 (2006).
28. I. M. Head, D. M. Jones, S. R. Larter, *Nature* **426**, 324 (2003).

29. We thank S. Henningsen, J. Mortensen, K. Barker, and A. Steen for assistance aboard R/V *Farm* and the Alfred Wegener Institute for kindly providing laboratory space in Ny Ålesund, Svalbard (project KOP56; RIS ID: 3298). We are grateful to K. Imhoff, M. Meyer, and A. Schipper for technical assistance and to A. Judd, M. Holtappels, N. Finke, and W. Bach for valuable discussions. This work was supported by the Natural Sciences and Engineering Research Council of Canada (C.H.), the Austrian Science Fund (P20185-B17; A.L.), the U.S. National Science Foundation (OCE-0323975 and OCE-0848703; C.A.), and the Max Planck Society. Nucleotide sequences have been deposited in the European Molecular Biology Laboratory database under accession numbers FN396615 to FN396795.

Supporting Online Material

www.sciencemag.org/cgi/content/full/325/5947/1541/DC1
Materials and Methods
SOM Text
Figs. S1 to S4
Tables S1 to S3

24 March 2009; accepted 20 July 2009
10.1126/science.1174012

Three-Dimensional Structural View of the Central Metabolic Network of *Thermotoga maritima*

Ying Zhang,[1]* Ines Thiele,[2]*† Dana Weekes,[3] Zhanwen Li,[1] Lukasz Jaroszewski,[3] Krzysztof Ginalski,[4] Ashley M. Deacon,[5] John Wooley,[6] Scott A. Lesley,[7] Ian A. Wilson,[8] Bernhard Palsson,[2] Andrei Osterman,[9] Adam Godzik[1,3,6]‡

Metabolic pathways have traditionally been described in terms of biochemical reactions and metabolites. With the use of structural genomics and systems biology, we generated a three-dimensional reconstruction of the central metabolic network of the bacterium *Thermotoga maritima*. The network encompassed 478 proteins, of which 120 were determined by experiment and 358 were modeled. Structural analysis revealed that proteins forming the network are dominated by a small number (only 182) of basic shapes (folds) performing diverse but mostly related functions. Most of these folds are already present in the essential core (~30%) of the network, and its expansion by nonessential proteins is achieved with relatively few additional folds. Thus, integration of structural data with networks analysis generates insight into the function, mechanism, and evolution of biological networks.

The advent of genome sequencing has enabled development of computational and experimental tools to investigate complete biological systems, but it has also highlighted the difficulty in integrating complex information for the hundreds to thousands of different molecules that compose even the smallest biological networks. Such integration presents many challenges, especially when assembling data from diverse fields, such as biochemistry and structural biology, that use different operational languages and conceptual frameworks. Biochemistry has traditionally focused on individual reactions and pathways, but recent advances in genomics have led to more rapid growth in the reconstruction and modeling of metabolic networks on a genome-wide scale (*1*–*3*). Thus, biochemical reactions, pathways, and networks can now be described in the context of entire cells, thereby enabling more realistic simulations of the behavior of metabolic networks in a growing number of organisms (*4*–*7*). Nevertheless, metabolism is still generally defined in terms of the chemical names and identity of substrates, products, and reactions. It does not explicitly consider the three-dimensional structures of its components, although such knowledge is required for a comprehensive understanding, not only of the individual reactions, but more importantly, of metabolic networks as a whole. Without such knowledge, we cannot rigorously define enzyme mechanisms or

[1]Joint Center for Molecular Modeling (JCMM), Burnham Institute for Medical Research, La Jolla, CA 92037, USA. [2]Department of Bioengineering, University of California at San Diego, La Jolla, CA 92093–0412, USA. [3]Joint Center for Structural Genomics (JCSG), Bioinformatics Core, Burnham Institute for Medical Research, La Jolla, CA 92037, USA. [4]Interdisciplinary Centre for Mathematical and Computational Modelling, Warsaw University, Warsaw, Poland. [5]JCSG, Structure Determination Core, Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA. [6]JCSG, Bioinformatics Core, University of California at San Diego, La Jolla, CA 92093, USA. [7]JCSG, Crystallomics Core, Genomics Institute of the Novartis Research Foundation, San Diego, CA 92121, USA. [8]JCSG, The Scripps Research Institute, La Jolla, CA 92037, USA. [9]Burnham Institute for Medical Research, La Jolla, CA 92037, USA.

*These authors contributed equally to this work.
†Present address: Center of Systems Biology, University of Iceland, IS-101 Reykjavik, Iceland.
‡To whom correspondence should be addressed. E-mail: adam@burnham.org

predict the effects of mutations or drugs; on the global level, we cannot understand the evolutionary relationships between different pathways, how new metabolic capabilities are acquired, and how individual organisms adapt to their particular ecological niches and respond to environmental pressures.

Such an understanding can be provided by structural biology, which has traditionally focused on individual proteins outside of their full, system-level, biological context. The emergence of large-scale structure genomics projects, such as the Protein Structure Initiative (*8*), has provided an exciting new opportunity for structural biology to contribute on a scale similar to that of genomics.

Thermotoga maritima, one of the first discovered hyper-thermophilic bacteria (*9*), represents the deepest known lineage of eubacteria (*9*, *10*), has one of the smallest genomes for a free-living organism (*11*), and has been the subject of extensive experimental analysis (*12*, *13*), making it an ideal model organism for systems biology and for integration of biochemical and structural approaches (*14*).

We constructed a metabolic model of *T. maritima* by a bottom-up approach, which first identified all known biochemical reactions and substrates from almost 150 publications (table S3), providing direct biochemical, genomic, and physiological evidence for more than 50% of the metabolic reactions. We then identified the remaining reactions from high confidence, homology-based annotation databases (*15*, *16*) and from experimental or modeled protein structures (see below). We used flux balance analysis (*17*) to test the completeness of the network, revealing gaps, such as missing enzymes or redundant functional assignments, which were then resolved by manual curation for individual cases. We continued iterative evaluation of the network until its performance reproduced, in silico, the experimentally determined metabolic capabilities of *T. maritima* (tables S9 and S10) (*18*).

Our resulting metabolic reconstruction included 478 metabolic genes, 503 unique metabolites, and 562 intracellular and 83 extracellular metabolic reactions (*18*), and it reproduced *T. maritima's* ability to grow on diverse carbohydrates (table S9) and to produce known metabolic by-products; e.g., acetate and hydrogen. The overall scope, content, and quality of this metabolic reconstruction were comparable with state-of-the-art reconstructions for other model organisms (table S6). Although the current model does not yet provide an exhaustive description of *T. maritima* metabolism, it represents a major step in an iterative process of annotation and modeling of this organism.

The *T. maritima* metabolic reconstruction (mr) defines a specific set of proteins (mrTM) that carry out the biochemical functions that make up a self-sustaining, metabolic network. Of

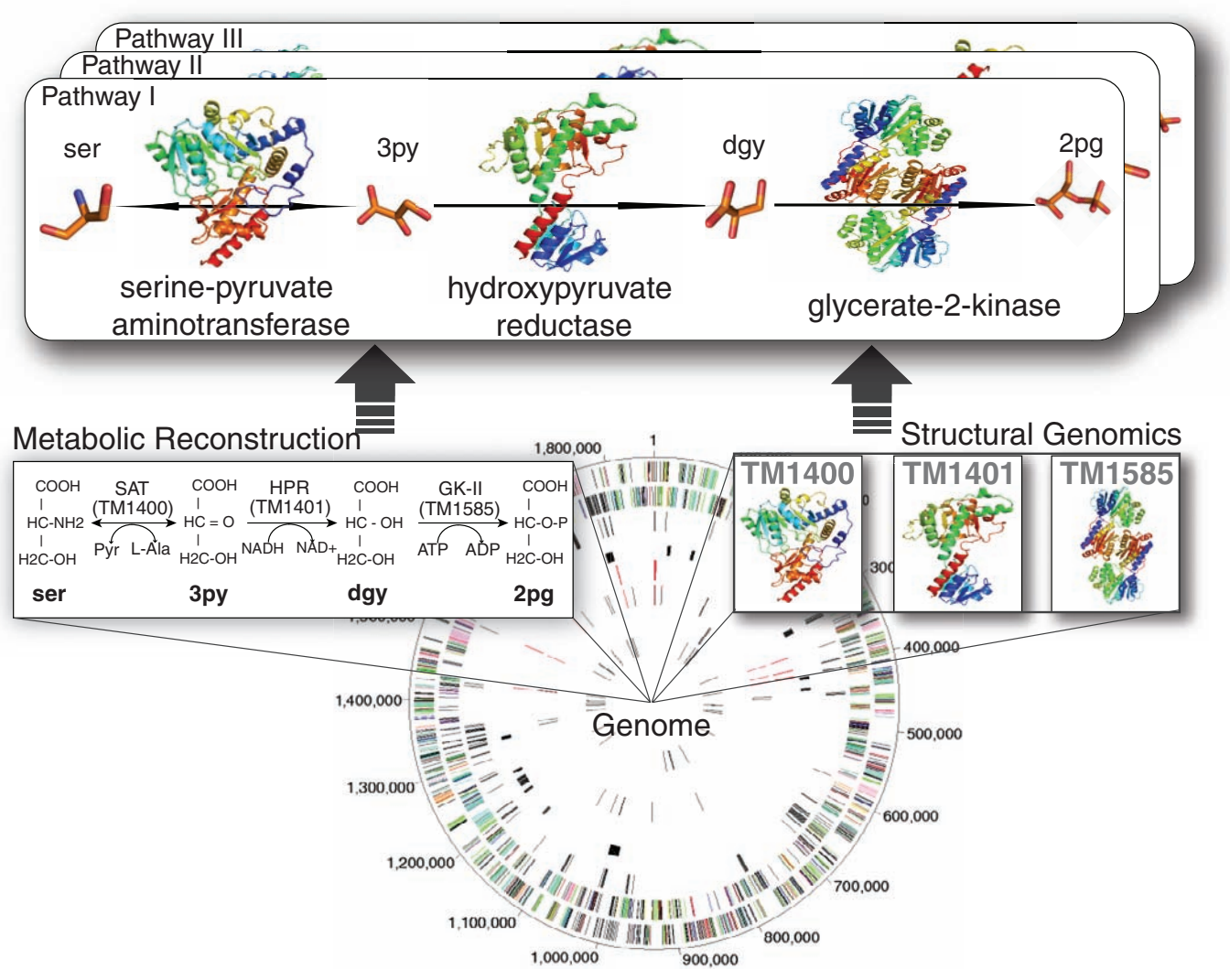
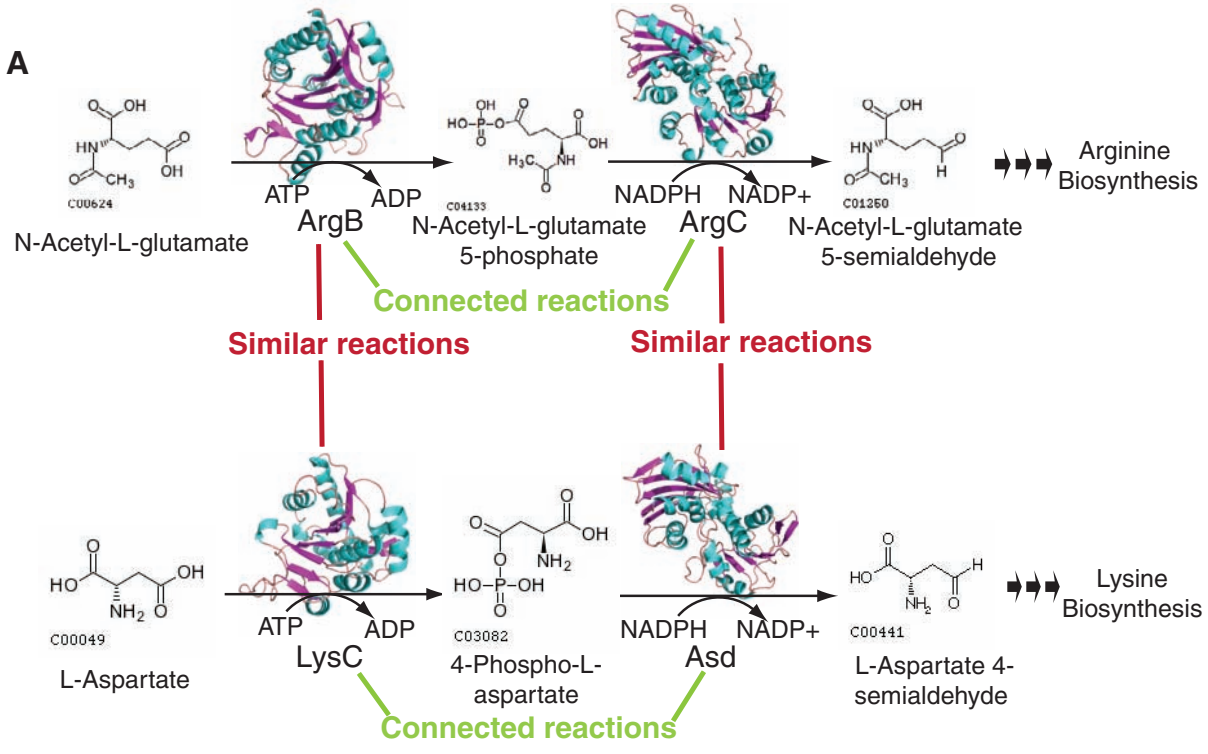


Fig. 1. Combining metabolic reconstruction and structural genomics approaches for an integrated annotation of the *T. maritima* central metabolic network. Underlying genomics information (**Bottom**) enabled both a metabolic reconstruction (**Left**) and an atomic-level structure determination/modeling of *T. maritima* proteins (**Right**). Integration of these two approaches enabled detailed information to be acquired for every reaction in the network (**Top**); an example from the *T. maritima* serine degradation pathway is illustrated (*32*).

478 proteins in this mrTM set, structures of 120 proteins have been determined experimentally (*12*), and 358 were predicted and modeled with a variety of computational approaches (*18*). The quality of the modeled structures spans the spectrum from those comparable to low-resolution, experimental structures (190 were built on templates with more than 30% identity to the targets) to very approximate (52 were based only on fold predictions). For three (TM1444, TM0788, and TM0540), the automated structure prediction approach failed, and approximate structures were modeled by combining several different fold prediction algorithms with manual refinement (*18*). Quality control, as based on public benchmarks in modeling and fold recognition, suggests high confidence that all models are correct at the fold-assignment level (*18*). Thus, these combined approaches allowed us to achieve complete structural coverage for the mrTM set (Fig. 1).

The information from structural determination of *T. maritima* proteins and their homologs provided additional support for functional assignment of 181 individual genes. A total of 41 experimental structures of *T. maritima* proteins contained relevant metabolites, and 140 crystal structures (used as templates for homology modeling) were also determined as complexes with ligands, all of which support the functional assignment in the reconstruction. In at least two cases, TM0449 (*19*–*22*) and TM1643 (*23*), structural analysis was critical for identification of enzymatic function and, in many other cases, substantially contributed to assignment of function.

Metabolic reconstruction not only can be described in a matrix format that can directly be used for metabolic simulations to predict essential genes or growth rates, it can also be represented as a graph. Because the reconstruction represents a fully functional, cell-level model of a metabolic network, analysis of the topology of this graph allows us to answer many interesting questions, especially when combined with knowledge of structures or models for all proteins in the network. For instance, what is the dominant mechanism for expansion of a metabolic network in a single organism? In the "patchwork" hypothesis (*24*), network expansion is driven by recruitment



B

Gene	Fold
ArgB	Carbamate kinase-like
ArgC	*NAD(P)-binding Rossmann-fold*
LysC	Carbamate kinase-like
Asd	*NAD(P)-binding Rossmann-fold*

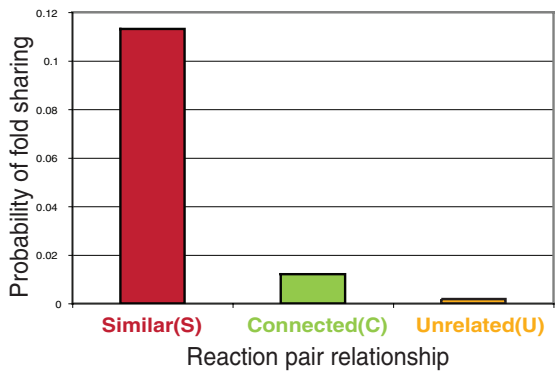


Fig. 2. Classification of metabolic reactions. (**A**) Examples of similar (S), connected (C), and unrelated (U) reactions from the arginine and lysine biosynthesis pathways. ArgB and LysC share a co-substrate [adenosine triphosphate (ATP)] that undergoes the same transformation [to adenosine diphosphate (ADP) + Pi]. Similarly, ArgC and Asd transform the reduced form of $NADP^+$ (NADPH) to nicotinamide adenine dinucleotide phosphate ($NADP^+$). By these criteria, both pairs are classified as similar. At the same time, reaction pairs ArgB/ArgC and LysC/Asd are adjacent in the pathway, because the product of the first reaction is the substrate for the next. These reaction pairs are classified as connected. All other pairs of reactions (ArgB/Asd and ArgC/LysC) are classified as unrelated. In this example, only the enzymes classified as similar (ArgB/LysC and ArgC/Asd) have the same fold. (**B**) Detailed information on the enzymes in (A). (**C**) Bars representing the relative number of pairs with the same fold in each category of reactions.

of proteins that perform similar reactions but are present in distinct pathways. Conversely, in the "retrograde" hypothesis (*25*), proteins evolve, after duplication, to perform dissimilar reactions within the same pathway or neighboring part of the network. Analysis of fold conservation as a function of network topology, therefore, addresses this key issue. Similar analyses have been performed previously on a small set of known pathways (*26*, *27*), but our integrated approach allowed us to analyze the complete set of pathways that form the fully functional, self-sustained metabolic network of a single organism.

We then established an automated protocol to classify metabolic reactions into three categories: (i) similar, (ii) connected, and (iii) unrelated (Fig. 2 and fig. S6). Enzymes that catalyze similar types of reactions have a sixfold higher probability of having the same fold than enzymes catalyzing connected reactions (Fig. 2C), supporting the patchwork hypothesis (*24*). However, it should be noted that proteins catalyzing connected reactions still have a higher chance of having the same fold as those catalyzing unrelated reactions, suggesting a role for gene duplication within pathways during pathway evolution (i.e., the retrograde model). More importantly, the patchwork hypothesis can account for only 11% of the observed structural similarity between mrTM proteins of similar function, indicating that convergent evolution of similar reaction mechanisms [i.e., nonhomologous gene displacement (*28*), where two nonhomologous proteins perform the same or similar metabolic function] is not a rare event and substantially contributes to evolution of the central metabolic network.

Another interesting question is the distribution and frequency of protein folds in this mrTM set. The 478 proteins contain 714 domains, but only 182 distinct folds, which are significantly fewer than would be expected (~300) for an equivalent random set of proteins with known structures (fig. S8). The surprisingly small number of folds arises from the fact that the most popular folds [e.g., the P-loop, triosephosphate isomerase (TIM) barrel, and Rossmann folds] are overrepresented as compared with their frequency in the general protein population (Fig. 3). Some relatively rare folds, abundant in the mrTM set, such as the biotin synthetase and the thiamin diphosphate binding folds, include groups of enzymes that perform specific but essential functions, such as tRNA aminoacylation or carbon metabolism.

The most obvious interpretation of this skewed fold distribution is that the mrTM set, which covers the most fundamental protein functions, consists of the most ancient and, thus, the most abundant protein families. To probe this interpretation further, we analyzed the fold distribution for the core of the *T. maritima* metabolic network, as represented by the set of essential proteins. We identified essential proteins by a reductive evolution simulation approach (*18*, *29*), where iterative simulations are performed to identify a minimal network by randomly eliminating genes from the model until additional elimination would result in a nonviable network. Each simulation led to a different minimal network, of size anywhere between 200 and 300 genes (i.e., corresponding to 42 to 63% of the mrTM set). Statistical analysis of 1000 such minimal networks in independent simulations in glucose minimal medium (*18*) allowed the classification of genes from the mrTM set into three categories: (I) core- or unconditional-essential genes that are always present, (II) nonessential genes that never appear, and (III) "synthetic lethal" or "conditional-essential" genes (*30*) that appear only in some simulations, but not in others, depending on which other genes are removed or retained in a particular network minimization. For example, if two genes have the same essential function, the deletion of either gene would not be lethal, but at least one gene has to be present in the minimal network. The frequency of such genes in multiple simulations reflects the topology of the network and the relative redundancy of gene functions in the network. It is important to emphasize that the core-essential genes would not be sufficient to maintain a viable metabolic network, as all of the many possible minimal networks contain constant (core-essential genes) and variable (subset

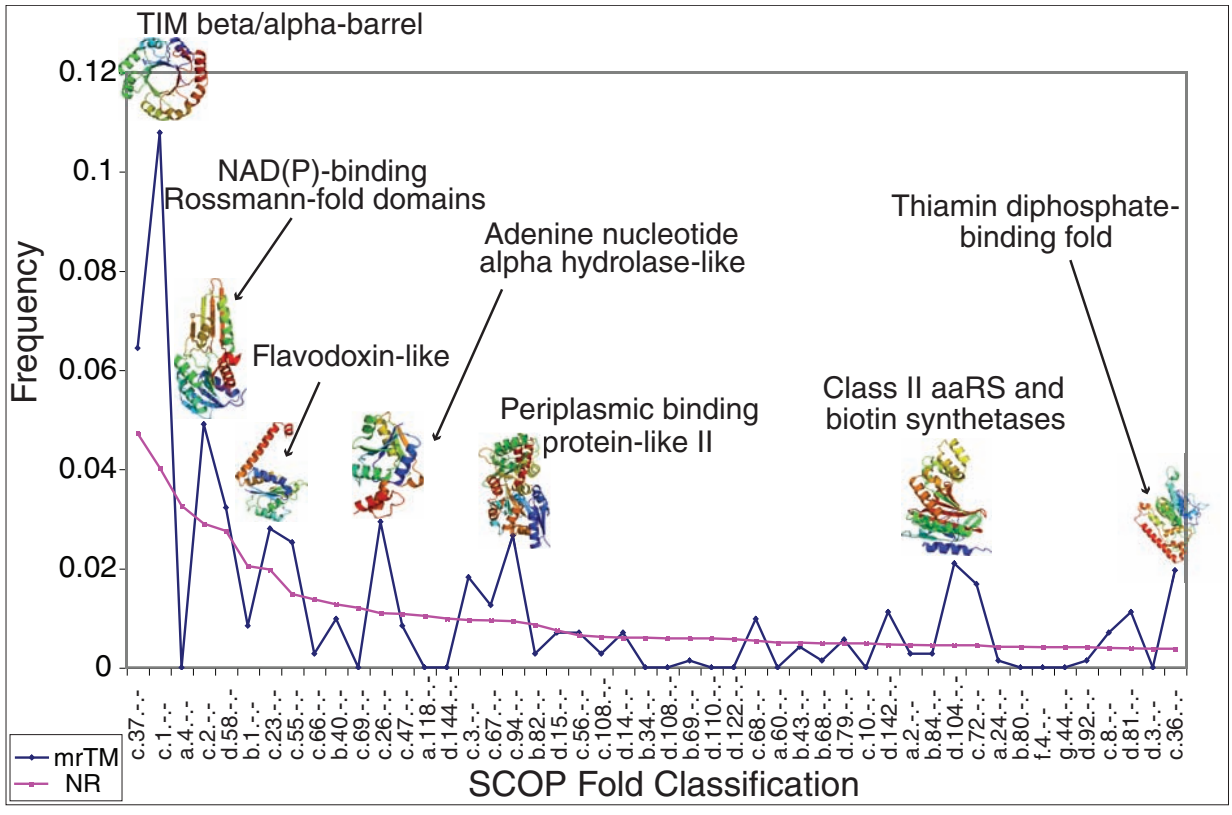


Fig. 3. Distribution of folds in the mrTM protein set with the most overrepresented folds, illustrated by structural ribbon diagrams. Fold codes from the Structural Classification of Proteins (SCOP) database (*33*) are shown on the *x* axis with the observed frequency on the *y* axis. The expected frequency for each fold in the NCBI nonredundant database (*31*) is shown as a magenta trace. TIM, triosephosphate isomerase.

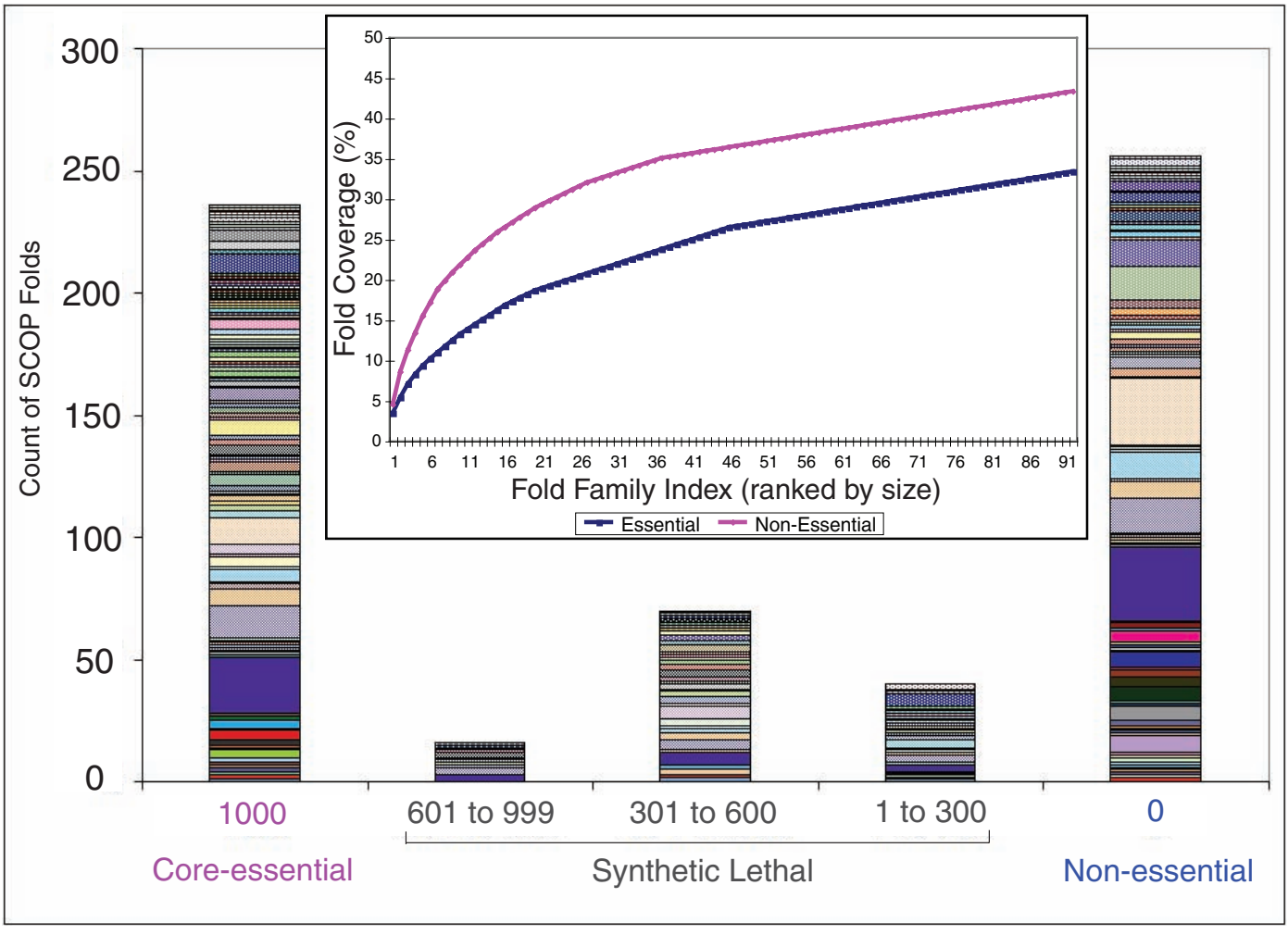


Fig. 4. Fold composition of the nonessential, synthetic lethal, and core-essential protein sets (see text for details) illustrated by colors associated with different folds (see fig. S9 for details). The *x* axis represents the number of simulations that resulted in identification of core-essential (1000 appearances in 1000 simulations), synthetic lethal (from 999 to 1), and nonessential genes (0); and the *y* axis indicates their classification into SCOP fold categories. (**Inset**) Cumulative fold coverage of core-essential and nonessential protein sets (blue, core-essential; magenta, nonessential). The fold distribution in all three groups is different, although core-essential and nonessential sets have some weak similarity, more than either group compared with synthetic lethal sets.

of conditionally essential genes) components. The mrTM set consists of 177 core-essential, 203 nonessential, and 98 conditional-essential genes. Proteins in these three sets have very different fold distributions (Fig. 4). The number of folds in the core-essential group is surprisingly large for its sample size (111 folds for 177 proteins) as compared with the nonessential group, which contains more proteins but a smaller number of folds (92 folds for 203 proteins). This trend is inverse to that observed when mrTM is compared with nonredundant sequences in the National Center for Biotechnology Information (NCBI) database (*31*) (fig. S8), where the mrTM set was more abundant in popular folds. These analyses suggest that core-essential proteins perform unique chemical functions that are strongly associated with specific folds and are so fundamental that their deletion would result in a nonviable network.

We have presented here the integration of a metabolic and structural view of the central metabolic network of the thermophilic bacterium *T. maritima*. Achieving a complete description on these two levels is an important milestone that now enables large-scale analyses, such as the network-scale comparison of correlations between fold conservation and biochemical function. From our study, not only can we provide a quantitative estimate of the dominance of the patchwork model (*24*) versus the retrograde model (*25*) of metabolic evolution, but we can also illustrate the importance of convergent or parallel evolution in proteins carrying out similar biochemical functions. Furthermore, we show that the set of proteins responsible for the central metabolism in *T. maritima* is highly nonrandom and dominated by a small number of folds that significantly exceed their already dominant distribution in the protein universe, suggesting that the central metabolism network has evolved mainly from a set of the most ancient proteins that have had sufficient time to develop divergent functionalities and, hence, expand into the very large and very diverse protein families that we observe today. At the same time, the subset of core-essential proteins reverses this trend and is relatively more diverse than an equivalent subset of nonessential proteins. This counterintuitive situation is attributable to the presence of some specific folds with functions that are so unique that it is impossible to replace them with other existing folds.

References and Notes

1. P. D. Karp, *Science* **293**, 2040 (2001).
2. T. Ideker, D. Lauffenburger, *Trends Biotechnol.* **21**, 255 (2003).
3. A. R. Joyce, B. O. Palsson, *Prog. Drug Res.* **64**, 265 (2007).
4. N. C. Duarte *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 1777 (2007).
5. A. M. Feist *et al.*, *Mol. Syst. Biol.* **3**, 121 (2007).
6. M. Durot *et al.*, *BMC Syst. Biol.* **2**, 85 (2008).
7. R. S. Senger, E. T. Papoutsakis, *Biotechnol. Bioeng.* **101**, 1036 (2008).
8. B. W. Matthews, *Nat. Struct. Mol. Biol.* **14**, 459 (2007).
9. R. Huber *et al.*, *Arch. Microbiol.* **144**, 324 (1986).
10. C. R. Woese, *Microbiol. Rev.* **51**, 221 (1987).

11. K. E. Nelson *et al.*, *Nature* **399**, 323 (1999).
12. S. A. Lesley *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 11664 (2002).
13. S. B. Conners *et al.*, *FEMS Microbiol. Rev.* **30**, 872 (2006).
14. "Integrative Biology of *Thermotoga maritima*" Workshop, San Diego, CA, 9 to 10 July 2007 (http://metagenomics.calit2.net/2007/thermotoga/).
15. S. Okuda *et al.*, *Nucleic Acids Res.* **36**, W423 (2008).
16. R. Overbeek *et al.*, *Nucleic Acids Res.* **33**, 5691 (2005).
17. C. H. Schilling *et al.*, *Biotechnol. Bioeng.* **71**, 286 (2000).
18. Materials and methods are available as supporting material on *Science* Online.
19. P. Kuhn *et al.*, *Proteins* **49**, 142 (2002).
20. TM0449 is a flavin adenine dinucleotide–dependent thymidylate synthase, and our structure has contributed to new developments in functional studies of this and related proteins [see (*21*, *22*) and references therein].
21. A. G. Murzin, *Science* **297**, 61 (2002); published online 23 May 2002 (10.1126/science.1073910).
22. E. M. Koehn *et al.*, *Nature* **458**, 919 (2009).
23. Z. Yang *et al.*, *J. Biol. Chem.* **278**, 8804 (2003).
24. R. A. Jensen, *Annu. Rev. Microbiol.* **30**, 409 (1976).
25. N. H. Horowitz, *Proc. Natl. Acad. Sci. U.S.A.* **31**, 153 (1945).
26. G. L. Holliday *et al.*, *Nat. Prod. Rep.* **24**, 972 (2007).
27. S. C. Rison, J. M. Thornton, *Curr. Opin. Struct. Biol.* **12**, 374 (2002).
28. E. V. Koonin, A. R. Mushegian, P. Bork, *Trends Genet.* **12**, 334 (1996).
29. C. Pal *et al.*, *Nature* **440**, 667 (2006).
30. J. L. Hartman IV, B. Garvik, L. Hartwell, *Science* **291**, 1001 (2001).
31. K. D. Pruitt, T. Tatusova, D. R. Maglott, *Nucleic Acids Res.* **35**, D61 (2007).
32. C. Yang *et al.*, *J. Bacteriol.* **190**, 1773 (2008).
33. A. G. Murzin, S. E. Brenner, T. Hubbard, C. Chothia, *J. Mol. Biol.* **247**, 536 (1995).

34. We specifically acknowledge the invaluable work of individual crystallographers at the JCSG and other Protein Structure Initiative (PSI) centers, as well as individual research groups, who have solved structures analyzed here, either directly or that we used as modeling templates. The full list of these proteins is provided in the supporting online material. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences (NIGMS). This work was supported by the NIH PSI grants P20 GM076221 (JCCM) and U54 GM074898 (JCSG) from the NIGMS; grant DE-FG02-08ER64686 from the Office of Science (Biological and Environmental Research), U.S. Department of Energy; and the Gordon and Betty Moore Foundation CAMERA project.

Supporting Online Material

www.sciencemag.org/cgi/content/full/325/5947/1544/DC1
Materials and Methods
Figs. S1 to S9
Tables S1 to S13
References
Metabolic reconstruction in SMBL and MATLAB formats

8 April 2009; accepted 22 July 2009
10.1126/science.1174671

Details of Insect Wing Design and Deformation Enhance Aerodynamic Function and Flight Efficiency

John Young,[1] Simon M. Walker,[2] Richard J. Bomphrey,[2] Graham K. Taylor,[2] Adrian L. R. Thomas[2]*

Insect wings are complex structures that deform dramatically in flight. We analyzed the aerodynamic consequences of wing deformation in locusts using a three-dimensional computational fluid dynamics simulation based on detailed wing kinematics. We validated the simulation against smoke visualizations and digital particle image velocimetry on real locusts. We then used the validated model to explore the effects of wing topography and deformation, first by removing camber while keeping the same time-varying twist distribution, and second by removing camber and spanwise twist. The full-fidelity model achieved greater power economy than the uncambered model, which performed better than the untwisted model, showing that the details of insect wing topography and deformation are important aerodynamically. Such details are likely to be important in engineering applications of flapping flight.

Insects achieve remarkable flight performance with a diverse range of complex wing designs (*1*, *2*). Computational fluid dynamics (CFD) offers an opportunity to identify the features underpinning the aerodynamic performance of insect wings. By comparing numerical simulations of different designs, it is possible to test the effects of modifications that may be outside the natural range of variation. Unfortunately, a lack of detailed measurements of insect wing kinematics has limited previous numerical studies of insect flight to two-dimensional (2D) models (*3*–*6*) or to 3D models in which the wings are modeled as rigid flat plates (*7*–*11*) or as rigid sections with constant camber and twist (*12*). Such simplifications can dramatically change the conclusions drawn about flow structure (*13*), and no model has yet been validated experimentally against flow visualizations from a real insect. We used the most detailed set of insect wing kinematics published to date (*2*) to develop the first 3D CFD model of an insect with deforming wings. We validated the results of our CFD simulations against qualitative and quantitative flow visualizations of real locusts. We then used progressive simplifications of the wing kinematics to analyze the aerodynamic consequences of the measured twist and camber.

We modeled a typical wingbeat of the desert locust *Schistocerca gregaria* (*14*) by averaging the kinematics of four consecutive wingbeats from one of the individuals described in (*2*). These kinematics were obtained by using four high-speed digital video cameras to track more than 100 natural features and marked points on the wings, which were then used to reconstruct the deforming surface topography of the wings with a mean spatial error of 0.11 mm (*15*). We fitted cubic splines to the wing outline and veins, and we interpolated these spatially to give the surface mesh for the CFD simulations, which we then interpolated temporally to give up to 800 time steps per wingbeat (*14*). We gave the modeled wings a nominal constant thickness of 0.05 mm based on published cross-sections of the wing veins and membrane (*16*). We did not attempt to model variations in thickness due to wing venation. Folding of the hindwing against the thorax could not be modeled exactly, and we instead modeled the hindwing as if it were joined to the thorax along its chord (*14*).

We solved the unsteady incompressible Navier-Stokes equations assuming laminar flow using a commercial CFD package (*14*). We constructed the CFD grid for the locust kinematics in multiple parts by using commercial software, and we incorporated the wing motions via a look-up table prescribing the kinematics (*14*). The wings and body were meshed with a triangular surface grid and surrounded with a thin boundary-layer grid to provide adequate resolution of velocity gradients normal to the surface. These were then surrounded with stationary outer regions representing the wind tunnel, and a deforming inner region in which the wings and boundary layer grids moved. A symmetry plane running through the sagittal plane of the insect was used. Aerodynamic forces on the wings and body were calculated by integrating pressure and viscous shear stress over the surfaces. Starting transients in the calculated aerodynamic forces vanished rapidly within the first wingbeat, with very close agreement between wingbeats thereafter, so we allowed the simulation to run for four repeated wingbeats. Aerodynamic power requirements were calculated by integrating the inner product of the local pressure and viscous forces with the local wing surface velocity in a coordinate system fixed to the insect's body.

We validated our CFD method against an independent CFD algorithm (*17*) by using our method to replicate published force computations for a simple model of a flapping dragonfly wing in hover (*11*, *14*). The predicted instantaneous vertical force coefficients from the two algorithms were in excellent agreement, with a linear

[1]School of Engineering and Information Technology, University of New South Wales, Australian Defence Force Academy, Canberra, Australian Capital Territory 2600, Australia. [2]Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

*To whom correspondence should be addressed. E-mail: adrian.thomas@zoo.ox.ac.uk