10. South, T. L., Blake, P. R., Harr, D. R. & Summers, M. F. C-terminal retroviral-type zinc finger domain from the HIV-1 nucleocapsid protein is structurally similar to the N-terminal zinc finger domain. Biochemistry 30, 6342–6349 (1991).
11. Luisi, B. F. et al. Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. Nature 352, 497–505 (1991).
12. Marmorsman, R., Carey, M., Pashine, M. & Harrison, S. C. Protein–DNA recognition by GAL4: structure of a protein–DNA complex. Nature 356, 408–414 (1992).
13. Everett, R. D. et al. A novel arrangement of zinc-binding residues and secondary structure in the C3HC4 motif of an alpha herpes virus protein family. J. Mol. Biol. 234, 1088–1047 (1993).
14. Barlow, P. N., Luisi, B., Milner, A., Elliot, M. & Everett, R. Structure of the C3H4 domain by 1H-nuclear magnetic resonance spectroscopy. J. Mol. Biol. 237, 201–211 (1994).
15. Borden, K. L. B. et al. The solution structure of the RING finger domain from the acute promyelocytic leukaemia proto-oncoprotein PML. EMBO J. 14, 1532–1541 (1995).
16. Phillips, S. E. V. The beta-ribbon DNA recognition motif. Annu. Rev. Biophys. Biomol. Struct. 23, 671–701 (1994).
17. Kim, I. L., Nikolov, D. B. & Burley, S. K. Co-crystal structure of TBP recognizing the minor groove of a TATA element. Nature 365, 520–527 (1993).
18. Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. Crystal structure of a yeast TBP–TATA-box complex. Nature 365, 512–520 (1993).
19. Schumacher, M. A., Choi, K. Y., Zalkin, H. & Brennan, R. G. Crystal structure of Laci member, Purr, bound to DNA: minor groove binding by a-helices. Science 266, 763–770 (1994).
20. Flick, K. E. et al. Crystalization and preliminary X-ray studies of F-Pir: a nuclear, intron-encoded homing endonuclease from Physarum polycephalum. Protein Sci. 6, 1–4 (1997).
21. Otwonowicz, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 276, 307–326 (1997).
22. Leslie, A. G. W. in Joint CCP4 and ESF-EACMB Newsletter on Protein Crystallography (Daresbury Laboratory, Warrington, UK, 1979).
23. CCP4 The SERC (UK) Collaborative Computing Project No. 4, a Suite of Programs for Protein Crystallography (Daresbury Laboratory, Warrington, UK, 1997).
24. QUANTA96 X-ray Structure Analysis User’s Reference (Joint CCP4 and ESF-EACMB Newsletter, 1992).
25. Leslie, A. G. W. in Joint CCP4 and ESF-EACMB Newsletter on Protein Crystallography (Daresbury Laboratory, Warrington, UK, 1992).
26. Leslie, A. G. W. The program X-PLOR: a system for X-ray crystallography and NMR (Scripps Research Institute, La Jolla, CA, 1993).
27. Laskowski, R. J., Macarthur, M., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. 26, 303–321 (1993).

Acknowledgements. We thank D. McHugh, K. Stephens and J. D. Heath for initial subcloning, purification and crystallization studies; R. Strong, K. Zhang and B. Scott for advice during the crystallographic analysis; and the beamline staff at the Advanced Light Source (NLBl laboratories), beamline 5-2, particularly T. Earnest, for assistance. B.L.S. and R.J.M. are funded for this project by the NIH. B.L.S. was supported by an NIH training grant and the American Heart Association. M.R.G. was supported by an NSF fellowship and an NIH training grant.

Correspondence and requests for materials and coordinates should be addressed to B.L.S. (e-mail: bstoddar@fred.fhcrc.org). Coordinates have been deposited in the Brookhaven Protein Data Bank (accession nos lipp, 1a73, 1a74).

letters to nature

Emergence of symbiosis in peptide self-replication through a hypercyclic network

David H. Lee, Kay Severin, Yohei Yokobayashi & M. Reza Ghadiri

Nature 390, 591–594 (1997)

Hypercycles are based on second-order (or higher) autocatalysis and defined by two or more replicators that are connected by another superimposed autocatalytic cycle. Our study describes a mutualistic relationship between two replicators, each catalysing the formation of the other, that are linked by a superimposed catalytic cycle. Although the kinetic data suggest the intermediary of higher-order species in the autocatalytic processes, the present system should not be referred to as an example of a minimal hypercycle in the absence of direct experimental evidence for the autocatalytic cross-coupling between replicators.

The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon Archaeoglobus fulgidus

Hans-Peter Klenk, Rebecca A. Clayton, Jean-Francois Tomb, Owen White, Karen E. Nelson, Karen A. Ketchum, Robert J. Dodson, Michelle Gwinn, Erin K. Hickey, Jeremy D. Peterson, Delwood L. Richardson, Anthony R. Kerlavage, David E. Graham, Nikos C. Kyrpides, Robert D. Fleischmann, John Quackenbush, Norman H. Lee, Granger G. Sutton, Steven Gill, Ewen F. Kirkness, Brian A. Dougherty, Keith McKenney, Mark D. Adams, Brendan Loftus, Scott Peterson, Claudia I. Reich, Leslie K. McNeil, Jonathan H. Badger, Anna Glodek, Lixin Zhou, Ross Overbeek, Jeaninne D. Gocayne, Janice F. Weidman, Lisa McDonald, Teresa Utterback, Matthew D. Cotton, Tracy Spriggs, Patricia Artiach, Brian P. Kaine, Sean M. Sykes, Paul W. Sadow, Kurt P. D’Andrea, Cheryl Bowman, Claire Fuji, Stacey A. Garland, Tanya M. Mason, Gary J. Olsen, Claire M. Fraser, Hamilton O. Smith, Carl R. Woese & J. Craig Venter

Nature 390, 364–370 (1997)

The pathway for sulphate reduction is incorrect as published: in Fig. 3 on page 367, adenylyl sulphate 3-phosphotransferase (cysO) is not needed in the pathway as outlined, as adenylyl sulphate reductase (aprAB) catalyses the first step in the reduction of adenylyl sulphate. The correct sequence of reactions is: sulphate is first activated to adenylyl sulphate, then reduced to sulphite and subsequently to sulphide. The enzymes catalysing these reactions are: sulphate adenylyltransferase (aprAB), adenylylsulphate reductase (aprAB), and sulphite reductase (dsrAB). We thank Jens-Dirk Schwenn for bringing this error to our attention.
Al$_{0.165}$Ga$_{0.835}$As spacer layer was grown with two Si (that is, on its left side in Fig. 2a and b), a 231-nm-thick as in quantum cascade lasers$^{13}$. Electrons would be injected from the scheme would use the quantum-well structure of Fig. 2a for the double-quantum-well system, and therefore the electrostatic potentials are identical in both structures. The $\delta$-doping provides a two-dimensional electron gas in the deep well with a calculated sheet electron density of $n_s = 4 \times 10^{11}$ cm$^{-2}$.

For the absorption measurements, we processed our samples in a multipass (six) 45° wedge guide waveguide. This geometry allowed us to couple in linearly polarized radiation with a large component of the polarization normal to the layer (50%) as required by the intersubband absorption selection rule$^{8}$. The absorption was measured with a Fourier-transform infrared spectrometer (FTIR) using a step-scan modulation technique$^{10}$ in which the electron gas in the double well is periodically depopulated by a Ti/Au Shottky barrier contact evaporated on the surface of the sample and the two-dimensional electron gas is contacted by indium balls alloyed into the layer.

The absorption measurements at $T = 10$ K for both structures are compared in Fig. 3 with the results of numerical calculations using the coupled Schrödinger's and Poisson's equations. As predicted, the absorption strength at photon energies between the two resonances is strongly suppressed or enhanced by the interference effect depending on the location of the thin barrier, proving that tunnelling through the latter controls the interference effect when the broadening of the states is dominated by tunnelling. However, the finite broadening introduced by interface disorder prevents full quantum interference; this is the main reason for the departure from the calculated profiles and specifically the reason why the absorption does not vanish in the sample with destructive interference. Indeed, linewidth measurements on samples with the same coupled-well structure but with negligible tunnelling to the continuum showed a full-width at half-maximum of the absorption peaks of $\Delta \lambda = 5$ meV. This structure consists of an identical double quantum well between two 60-nm-thick Al$_{0.33}$Ga$_{0.67}$As barriers. This value is a measure of the non-tunnelling contribution to the broadening of the optical transitions; it is smaller but not negligible compared with the calculated broadening by tunnelling through the 1.5 nm barrier, $\Gamma_T = \Gamma_s = 16$ meV.

Destructive interference in intersubband absorption in a double-well structure coupled by tunnelling to a continuum has recently been inferred from a fit of the absorption lineshape to a model that included the collision broadening in a phenomenological manner$^{13}$. The present experiment gives more direct evidence of tunnelling-induced quantum interference by showing that tunnelling can be used to control the sign of the interference.

It is important to stress the difference between the phenomena described here and the Fano interference in intersubband absorption recently reported by us$^{12}$. In that work a minimum in the absorption arises because of interference between matrix elements for the ground state to the continuum and to a single resonance coupled by tunnelling to the same continuum. This leads to a strongly asymmetric absorption lineshape. In contrast, in the phenomena studied here interference arises between absorption paths through two resonances coupled to a continuum, and the direct matrix element from the ground state to the continuum is negligible.

These findings are relevant for the design of semiconductor lasers without population inversion (LWI). Such lasing action has so far been observed only in gases$^{4,5}$. Essential for LWI is nonreciprocity between emission and absorption. A possible semiconductor LWI scheme would use the quantum-well structure of Fig. 2a for the active regions. The latter would be alternated with electron injectors as in quantum cascade lasers$^{11}$. Electrons would be injected from the thick barrier side at an energy between the two resonances where the absorption cross-section is a minimum, to ensure strong nonreciprocity between intersub-band absorption and emission$^{8,9}$. Although the realization of such a laser would be scientifically important, its implementation would be difficult and its technological impact limited by the very short lifetime (a few tenths of picoseconds) of the excited state which is required to achieve strong interference$^{14}$.

Received 23 June; accepted 6 October 1997.

1. Fano, U. Effect of configuration interaction on intensity and phase shifts. Phys. Rev. 124, 1866–1878 (1961).
2. Harris, S. E. Lasers without inversion: interference of lifetime-broadened resonance. Phys. Rev. Lett. 62, 1033–1035 (1989).
3. Arkhipov, V. G. & Heller, Y. I. Radiation amplification without population at transitions to autoionizing states. Phys. Lett. A 198, 12–14 (1993).
4. Zibrov, S. et al. Experimental demonstration of laser without population inversion via quantum interference in Rh. Phys. Rev. Lett. 75, 1499–1502 (1995).
5. Pudmazandu, G. et al. Laser oscillation without population inversion in a Sodium atomic beam. Phys. Rev. Lett. 76, 2053–2056 (1996).
6. Bolley, K., Imamoglu, A. & Harris, S. E. Observation of electromagnetically induced transparency. Phys. Rev. Lett. 66, 2593–2596 (1991).
7. Imamoglu, A. & Harris, S. E. Lasers without inversion: interference of dressed lifetime-broadened states. Opt. Lett. 13, 1344–1346 (1988).
8. Imamoglu, A. & Ram, R. J. Semiconductor lasers without population inversion. Opt. Lett. 19, 1744–1746 (1994).
9. Levine, B. F., Choi, K. K., Bettha, C. G., Walker, J. & Malik, R. J. New 10 pm infrared detector using intersubband absorption in resonant tunneling GaAs/AlAs superlattices. Appl. Phys. Lett. 50, 1092–1094 (1987).
10. Faist, J. et al. Narrowing of the intersubband electroluminescence spectrum in coupled-quantum-well heterostructures. Appl. Phys. Lett. 65, 94–96 (1994).
11. Smith, H., Cappelleri, M. K., Gossard, A. C. & Imamoglu, A. Tunnel induced transparency: Fano interference in intersubband transitions. Appl. Phys. Lett. 70, 3453–3457 (1997).
12. Faist, J. et al. Tunable Fano interference in intersubband absorption. Opt. Lett. 21, 983–987 (1996).
13. Capasso, F., Faist, J., Stihori, C. & Cho, A. T. Infrared (4–11 pm) quantum cascade lasers. Solid State Commun. 102, 231–236 (1997).
14. Kurghin, J. B. & Rosencher, E. Practical aspects of lasing without inversion in various media. IEEE J. Quantum Electron. QE-32, 1882–1890 (1996).

Acknowledgements. We thank S. Harris, A. Imamoglu, R. J. Ram, A. Tredicucci and C. Gmachl for enlightening discussions, and the authors of ref. 11 for making their manuscript available before publication. The work was supported in part by DARPA/ARO.

Correspondence should be addressed to F.C. (e-mail: fc@bell-labs.com or fc@lucent.com).

Emergence of symbiosis in peptide self-replication through a hypercyclic network

David H. Lee, Kay Severin, Yohei Yokobayashi & M. Reza Ghadiri

Departments of Chemistry and Molecular Biology, and the Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, California 92037, USA

Symbiosis is an association between different organisms that leads to a reciprocal enhancement of their ability to survive. Similar mutually beneficial relationships can operate at the molecular level in the form of a hypercycle, a collective of two or more self-replicating species interlinked through a cyclic catalytic network$^{1-3}$. The superposition of cross-catalysis onto autocatalytic replication integrates the members of the hypercycle into a single system that reproduces through a second-order (or higher) form of nonlinear autocatalysis. The hypercycle population as a whole is therefore able to compete more efficiently for existing resources than any one member on its own. In addition, the effects of beneficial mutations of any one member are spread over the entire population. The formation of hypercycles has been suggested as an important step in the transition from inanimate to living chemistry$^6$, and a large number of hypercycles are expected to be embedded within the complex networks of living systems$^7$. But only one naturally occurring hypercycle has been well documented$^4$, while two autocatalytic chemical systems may contain vestiges of hypercyclic organization$^{9,10}$. Here we report a
The following peptide sequences were employed in this study: replicator 1 (E, peptide fragments (ref. 12 and K. Kumar, D.H.L., M.R.G., unpublished results). The replicator is made from the same electrophilic fragment but a different nucleophilic peptide fragment \( N_2 \). The nucleophilic fragments \( N_1 \) and \( N_2 \) differ in their sequence at the hydrophobic recognition surface—\( N_1 \) is composed of valine and leucine whereas \( N_2 \) is made up of isoleucine and leucine residues. This difference in sequence at the hydrophobic core is known to affect profoundly the recognition surface—

\[ \text{E} \quad \text{N}_1 \quad \text{N}_2 \]

Figure 1 Schematic diagram of a minimal hypercycle based on two self-replicating peptides. Cycles I and III show the self-producing cycles of replicators \( R_1 \) (dark grey/light grey) and \( R_2 \) (dark grey/stripped) respectively, which pre-organize their constituent fragments thereby promoting peptide ligation. Cycle II, where \( R_1 \) promotes \( R_2 \) formation, and cycle IV, where \( R_2 \) promotes \( R_1 \) formation, comprise the catalytic components of the hypercycle and allow the replicators to positively regulate each others’ production. The mechanistic details of the present hypercyclic network may be more complex than the minimal system depicted here. Detailed kinetic analyses of the replicator sequences have shown that the autocatalytically productive intermediates involve, at least in part, quaternary complexes in which two template strands pre-organize the reactive peptide fragments (ref. 12 and K. Kumar, D.H.L., M.R.G., unpublished results). The following peptide sequences were employed in this study: replicator 1 (\( R_1 \), ArCONH-RMKQLEEKYELLSSKA-CLEXEVARLKLQVE-COH-N; replicator 2 (\( R_2 \), ArCONH-RMKQLEEKYELLSSKA-CLEXEVARLKLQVE-COH-N; electrophilic fragment (E), ArCONH-RMKQLEEKYELLSSKA-COSBN; nucleophilic fragment 1 (\( N_1 \), H_N-CLEXEVARLKLQVE-COH-N; nucleophilic fragment 2 (\( N_2 \), H_N-CLEXEVARLKLQVE-COH-N; Bn, benzyl; Ar, 4-acetamidophenyl; and X, lysine-\( \text{NHCO-Ar.} \)

The present design of a minimal hypercycle is based on two self-replicating coiled coil peptides \( R_1 \) and \( R_2 \) (Fig. 1). The replicator was recently reported and is produced as the ligation product of the electrophilic peptide fragment (ref. 12 and K. Kumar, D.H.L., M.R.G., unpublished results). II, where

\[ \text{H}_2\text{N-CLEXEVARLKLQVE-COH-N} \]

is composed of valine and leucine whereas \( N_2 \) is made up of isoleucine and leucine residues. This difference in sequence at the hydrophobic core is known to affect profoundly the recognition surface—

\[ \text{E} \quad \text{N}_1 \quad \text{N}_2 \]

Figure 1 Schematic diagram of a minimal hypercycle based on two self-replicating peptides. Cycles I and III show the self-producing cycles of replicators \( R_1 \) (dark grey/light grey) and \( R_2 \) (dark grey/stripped) respectively, which pre-organize their constituent fragments thereby promoting peptide ligation. Cycle II, where \( R_1 \) promotes \( R_2 \) formation, and cycle IV, where \( R_2 \) promotes \( R_1 \) formation, comprise the catalytic components of the hypercycle and allow the replicators to positively regulate each others’ production. The mechanistic details of the present hypercyclic network may be more complex than the minimal system depicted here. Detailed kinetic analyses of the replicator sequences have shown that the autocatalytically productive intermediates involve, at least in part, quaternary complexes in which two template strands pre-organize the reactive peptide fragments (ref. 12 and K. Kumar, D.H.L., M.R.G., unpublished results). The following peptide sequences were employed in this study: replicator 1 (\( R_1 \), ArCONH-RMKQLEEKYELLSSKA-CLEXEVARLKLQVE-COH-N; replicator 2 (\( R_2 \), ArCONH-RMKQLEEKYELLSSKA-CLEXEVARLKLQVE-COH-N; electrophilic fragment (E), ArCONH-RMKQLEEKYELLSSKA-COSBN; nucleophilic fragment 1 (\( N_1 \), H_N-CLEXEVARLKLQVE-COH-N; nucleophilic fragment 2 (\( N_2 \), H_N-CLEXEVARLKLQVE-COH-N; Bn, benzyl; Ar, 4-acetamidophenyl; and X, lysine-\( \text{NHCO-Ar.} \)

The ability of \( R_2 \) to self-replicate was determined by observation of characteristics previously established as signatures of self-replication (Fig. 2). Similar to that of \( R_1 \), the new replicator \( R_2 \) also displays a parabolic growth profile. Numerical fitting of the kinetic data obtained for \( R_2 \) to the empirical rate equations of von Kiedrowski equates the background rate constant \( k_0 = 0.072 \pm 0.005 \text{M}^{-1} \text{s}^{-1} \) and an apparent autocatalytic rate constant \( k_3 = 52 \pm 1 \text{M}^{-2} \text{s}^{-1} \), making \( R_2 \) more efficient than its relative \( R_1 \) (\( k_3 = 0.063 \text{M}^{-1} \text{s}^{-1} \) and constant \( k_0 = 29.4 \text{M}^{-2} \text{s}^{-1} \)).

A solution containing all three fragments \( E, N_1 \) and \( N_2 \) gave a combinatorial synthesis of both replicators. A priori, one would expect a survival-of-the-fittest situation where the more efficient replicator \( R_2 \) would overwhelm \( R_1 \) by consuming the common fragment \( E \) more quickly. At first glance, this expectation seemed to be borne out as \( R_2 \) was produced in greater abundance than \( R_1 \) (as expected, when molecular interactions are disrupted in the presence of guanidinium hydrochloride, no kinetic preference for \( R_2 \) over \( R_1 \) was observed). However, the situation is more interesting and complex. When we sought to give \( R_1 \) an advantage in this competition by adding 40% \( R_1 \) (with respect to the nucleophile concentration) at the start of the reaction, to our surprise the rate of \( R_2 \) self-production increased by only 1.7 times over the unseeded reaction but the rate of \( R_1 \) formation was enhanced to a greater extent, by 5.4 times (Table 1, Fig. 3). Thus the two replicators are not mutually exclusive in their growth; \( R_1 \) catalyses the formation of \( R_2 \) as well as itself. Likewise, perturbation of the reaction by seeding it with 45% \( R_2 \) not only increased the rate of \( R_2 \) production 2.9 times but \( R_1 \) as well, by 3.5 times. Thus a cross-catalytic cycle is cooperatively coupled with two self-replicating reactions, making this system one which is hypercyclic in nature. There are four characteristic outcomes expected for such a hypercyclic network, depending on the relative efficiencies of the coupled catalytic and autocatalytic reactions. The observed greater efficiencies of the catalytic reactions over the autocatalytic components of the system are the most desirable outcomes which assure the stability of the hypercycle: production of one species promotes the production of the other to an even greater degree. This particular mode of catalytic coupling prevents one replicator from overwhelming the other and enables the two to reproduce as a single coherent unit.

To verify that \( R_1 \) and \( R_2 \) catalyse each other’s production, the

The data in this table (in units of \( 10^{-5} \text{M} \)) are for reactions containing the three peptide fragments in the absence and presence of added replicators.

| Product | No replicators added | +40% \( R_1 \) | +45% \( R_2 \) |
|---------|----------------------|----------------|----------------|
| \( R_1 \) | 4.8 | 8.2 | 17.0 |
| \( R_2 \) | 5.8 | 31.1 | 18.9 |

The letters to nature

NATURE | VOL 390 | 11 DECEMBER 1997

Nature © Macmillan Publishers Ltd 1997
reaction mixtures were simplified to include E and only one nucleophile, and then seeded with the template that was not produced in situ (Fig. 4). Comparisons with unseeded reactions revealed that even in these simplified systems one template can promote the formation of the other, giving rate enhancements much larger than what would be expected if the reaction mixture were seeded with the autocatalytic template. Reaction mixtures containing E and N₁ that were seeded with 25% R₂ enhanced the initial rate of production of R₁ from $3.9 \times 10^{-8}$ M min⁻¹ to $1.5 \times 10^{-7}$ M min⁻¹, a 3.8 times increase over the unseeded reaction. Seeding of the same reaction mixture with 25% R₁ would improve the rate by only 2.8 times. Similarly, seeding reaction mixtures containing E and N₂ with 35% R₁ gave a 5.4 times rate enhancement over the $5.0 \times 10^{-8}$ M min⁻¹ rate observed for the reaction without added catalyst. The increase is greater than the 3.6 times enhancement expected for the autocatalytic reaction containing 35% R₂.

We now consider the sequence selectivity issues in the formation of the hypercyclic peptide network. The operation of the hypercycle is based on complementary, as well as self-complementary, forms of catalysis. As noted below, there is mounting evidence that both processes are strongly sequence selective. Previously we had shown that in the case of replicator R₁ even conservative mutations (Val₉Ala—where a valine has been substituted by an alanine at position 9—and Leu₂₆Ala) in the hydrophobic core residues completely abolish the autocatalytic process¹¹–¹². In this study we have determined that similar replicator R₂ mutations are also autocatalytically infertile. There is also good evidence for high sequence selectivity in the cross-catalytic component of the system. Control studies have indicated that the Leu₂₆Ala R₂ mutant cannot cross-catalyse the formation of replicators R₁ nor R₂. Although in a recent study¹⁵ we have shown that the Val₉Ala R₁ mutant can efficiently cross-catalyse the formation of R₁, we have found it to be ineffective in catalysing R₂ production. Moreover, in a related study we have shown that diminution in the initial rate of peptide fragment condensation of more than 3 orders of magnitude can be caused even by electrostatic substitutions at the solvent-exposed e and g positions of the heptad repeat sequence¹⁷. Although the above studies strongly support high sequence selectivity in the catalytic and autocatalytic components of the hypercyclic network, a significantly large sequence-space must undoubtedly exist that would enable the spontaneous self-organization of even more complex networks. Studies along those lines are under investigation.

The work reported here may have particular relevance to various origin-of-life theories¹–⁴,¹⁸. It has been suggested that at the dawn of life the onset of darwinian evolution must have been marked by

**Figure 3** Replicators R₁ and R₂ self-organize into a two-membered hypercyclic network. a, Production of R₁ (empty circles) and R₂ (empty diamonds) as a function of time for reaction mixtures containing E, N₁ and N₂. b, Formation of R₁ (filled circles) and R₂ (filled diamonds) as a function of time for reaction mixtures containing the three fragments and 40% R₁. c, Formation of R₁ (filled circles) and R₂ (filled diamonds) as a function of time for reaction mixtures containing the three fragments and 45% R₂.

**Figure 4** Replicators R₁ and R₂ are cross-catalytic. a, Formation of R₁ as a function of time for the reaction mixture containing only E and N₂ in the absence (empty circles) and in the presence (filled circles) of 35% R₂. b, Formation of R₂ as a function of time for the reaction mixture containing E and N₁ in the absence (empty diamonds) and in the presence (filled diamonds) of 25% R₁.
Kinetic limitations on droplet formation in clouds

P. Y. Chuang†, R. J. Charlson‡ & J. H. Seinfeld‡
† Department of Environmental Engineering Science, MC 138-78, California Institute of Technology, Pasadena, California 91125, USA
‡ Departments of Atmospheric Sciences and Chemistry, University of Washington, Seattle, Washington 98195, USA
‡ Division of Engineering and Applied Science and Department of Chemical Engineering, MC 210-41, California Institute of Technology, Pasadena, California 91125, USA

The ’indirect’ radiative cooling of climate due to the role of anthropogenic aerosols in cloud droplet formation processes (which affect cloud albedo) is potentially large, up to ~1.5 W m⁻² (ref. 1). It is important to be able to determine the number concentration of cloud droplets to within a few per cent, as radiative forcing as a function of clouds is very sensitive to changes in this quantity, but empirical approaches are problematic. The initial growth of a subset of particles known as condensation nuclei and their subsequent ’activation’ to form droplets are generally calculated with the assumption that cloud droplet activation occurs as an equilibrium process described by classical Köhler theory. Here we show that this assumption can be invalid under certain realistic conditions. We conclude that the poor empirical correlation between cloud droplet and cloud condensation nuclei concentrations is partly a result of kinetically limited growth before droplet activation occurs. Ignoring these considerations in calculations of total cloud radiative forcing based on cloud condensation nuclei concentrations could lead to errors that are of the same order of magnitude as the total anthropogenic greenhouse-gas radiative forcing.

Cloud droplet activation and subsequent treatments of cloud droplet growth in atmospheric models generally rely on the assumption that pre-activation growth is accurately described by an equilibrium model in which the particle diameter is always at equilibrium with the local supersaturation⁴. The equilibrium relationship between supersaturation and particle size for a particle composed of highly soluble inorganic species can be described by the well-known Köhler equation (curve A, Fig. 1). Cloud droplet nuclei (CDN) activate when they grow larger than their critical diameter, Dₚ, after which they can grow spontaneously, limited only by kinetics. The concept of CDN is distinct from that of CCN in that, whereas CCN are defined as those particles that activate to become cloud droplets within a cloud chamber of fixed or prescribed supersaturation, CDN are those particles that actually activate in the atmosphere under conditions of time-varying supersaturation.

To evaluate the conditions under which the equilibrium activation model is valid, two timescales will be defined. One is the timescale for particle growth that would be required for that particle to remain at equilibrium as the ambient supersaturation ratio increases in a rising air parcel, τ₀. The other is the timescale for actual change in the droplet size resulting from condensational growth, τ. Hence, if τ₀ ≥ τ then the equilibrium model is reasonable; otherwise, CDN activation, and hence the cloud droplet size distribution, can be accurately predicted only if the kinetics of droplet growth are considered. To calculate τ₀, the rate of change of the droplet diameter that would be required for that droplet to remain at its equilibrium size, Dₑ/𝑑t, is determined from the combination of two effects. First, the time rate of change of supersaturation, dS/dt, can be determined using a simple one-dimensional adiabatic parcel model. Next, the rate of change of Dₑ with respect to supersaturation, dDₑ/dS, is determined by differentiating

---

Methods

Self-replication of Rₐ. All reactions were done in 0.6 ml Eppendorf tubes at 23°C. A stock solution containing E, N₂ and the internal standard 4-acetamidobenzoic acid (ABA), were seeded with various amounts of Rₐ, benzylmercaptan (1 µl) was then added. Reactions were initiated by adding 3-(N-morpholino)propanesulfonic acid (MOPS) buffer (pH 7.50, 200 mM, 236 µl), giving a total volume of 300 µl and concentrations of [N₁] = 104.5 µM, [E] = 94.2 µM, [R₀] = 0. 4.0, 21.4 or 42.6 µM. [MOPS] = 157 mM, [ABA] = 40.4 µM. Samples (30 µl) were taken at various time points and quenched with 2% trifleuroacetic acid (TFA) in water (70 µl) then stored at −70°C. Samples were analysed by high pressure liquid chromatography on a Zorbax C8 column using an acetonitrile/water/0.1% TFA gradient while monitoring at 270 nm. The identity of all peptides was determined by mass spectrometry and verified by co-injection with authentic standards. Samples were done in duplicate.

Determination of hypercyclic organization in the E/N₁/N₂ mixture. Reactions were done as described above except that the stock solution contained, besides E and ABA, both N₁ and N₂, which was subsequently seeded with either Rₐ, R₀ or water. Reactions were initiated by adding MOPS buffer (pH 7.50, 200 mM, 236 µl), giving a total volume of 300 µl and concentrations of [N₂] = 112.5 µM, [N₁] = 112.7 µM, [E] = 91.1 µM, [MOPS] = 157.7 mM, [ABA] = 97.1 µM, [R₀] = 45.1 µM, [R₁] = 50.4 µM.

Verification of the catalytic components of the hypercycle. Reactions were performed as described above except only one nuclease was present in the reaction mixture and the reaction was seeded with the replicator that was not produced in situ. Initial concentrations are (1) [E] = 88.9 µM, [N₂] = 98.2 µM, [R₁] = 25.2 µM, [ABA] = 50.3 µM; (2) [E] = 80.4 µM, [N₂] = 96.9 µM, [R₁] = 35.3 µM, [ABA] = 36.9 µM.

Received 20 June; accepted 20 October 1997.

1. Eigen, M. & Schuster, P. The hypercycle. A principle of natural self-organization. Part A: emergence of the hypercycle. Naturwissenschaften 64, 541–565 (1977).
2. Eigen, M. & Schuster, P. The hypercycle. A principle of natural self-organization. Part B: the abstract hypercycle. Naturwissenschaften 65, 7–41 (1978).
3. Eigen, M. & Schuster, P. The hypercycle. A principle of natural self-organization. Part C: The realistic hypercycle. Naturwissenschaften 65, 341–369 (1978).
4. Eigen, M. Self-organization of matter and the evolution of biological macromolecules. Naturwissenschaften 58, 483–513 (1971).
5. Müller-Hermel, U. What is a hypercycle? Thyrol. Biol. 102, 589–591 (1983).
6. Lee, D. H., Severin, K. S. & Ghadiri, M. R. Autocatalytic networks: the transition from molecular self-replication to molecular ecosystems. Curr. Opin. Chem. Biol. (in the press).
7. Ricard, J. & Noit, G. Electrostatic effects and the dynamics of enzyme reactions at the surface of plant cells. In A theory of the ionic control of a complex multi-enzyme system. Eur. J. Biochem. 155, 183–190 (1986).
8. Eigen, M., Birnboim, C. K., Gebienoga, M. & Gardner, W. C. The hypercycle. Coupling of RNA and protein biosynthesis in the infection cycle of an RNA bacteriophage. Biochemistry 30, 11005–11018 (1991).
9. Hong, J.-T., Feng, Q., Rotello, V. & Rebek, J. Jr Competition, cooperation and mutation: improving a self-replicating system from three starting materials. Angew. Chem. Int. Edn Engl. 32, 1188–1201 (1993).
10. Lee, D. H., Grania, J. R., Martinez, J. A., Severin, K. S. & Ghadiri, M. R. A self-replicating peptide. Nature 382, 525–528 (1996).
11. Severin, K. S., Lee, D. H., Martinez, J. A. & Ghadiri, M. R. Peptide self-replication via template-directed ligation. Chem. Eur. J. 3, 1017–1024 (1997).
12. Harbury, P. B., Zhang, T., Kim, P. S. & Alber, T. A switch between two-, three- and four-stranded coiled coils in GCN4 leucine zipper mutants. Science 262, 1401–1407 (1993).
13. Hsu, J. C., O’Shea, E. K., Kim, P. S. & Sauer, R. T. Sequence requirements for coiled coils: analysis with λ repressor-GCN4 leucine zipper fusions. Science 250, 1400–1403 (1990).
14. Severin, K., Lee, D. H., Martinez, J. A. & Ghadiri, M. R. Dynamic error-correction in an autocatalytic peptide network. Angew. Chem. Int. Edn. (in press).
15. von Kiedrowski, G. Minimal replicator theory I: parabolic versus exponential growth. Biophys. Chem. Front. 3, 113–116 (1990).
16. Sevemin, K., Lee, D. H., Kennaan, A. J. & Ghadiri, M. R. A synthetic peptide ligase. Nature 398, 706–709 (1999).
17. Kauffman, S. A. The Origins of Order (Oxford Univ. Press, New York, 1993).
18. Kieppers, R.-O. The Origins of Biological Information (MIT Press, Cambridge, MA, 1990).
19. Joyce, G. F. RNA evolution and the origins of life. Nature 338, 217–224 (1989).

Acknowledgements. We thank K. Kumar for discussions. We also thank the Medical Research Council of Canada for a postdoctoral fellowship (D.J.H.L.), and the Deutsche Forschungsgemeinschaft for a postdoctoral fellowship (R.S.).

Correspondence and requests for materials should be addressed to M.R.G. (e-mail: ghadiri@scripps.edu).
The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*

Hans-Peter Klenk**, Rebecca A. Clayton**, Jean-Francois Tomb**, Owen White**, Karen E. Nelson**, Karen A. Ketchum**, Robert J. Dodson**, Michelle Gwinn**, Erin K. Hickey**, Jeremy D. Peterson**, Delwood L. Richardson**, Anthony R. Kerlavage**, David E. Graham**, Nikos C. Kyripes†, Robert D. Fleischmann**, John Quackenbush**, Lisa McDonald**, Jonathan H. Badger**, Keith McKenney**, Lisa McDonald†, Mark D. Adams†, Brendan Loftus†, Scott Peterson†, Claudia I. Reich†, Leslie K. McNeil†, Jonathan H. Badger†, Anna Glodek†, Lixin Zhou†, Ross Overbeek‡, Jeannine D. Gocayne‡, Janice F. Weidman‡, Lisa McDonald‡, Teresa Utterback‡, Matthew D. Cotton‡, Tracy Spriggs‡, Patricia Artiach‡, Brian P. Kaine‡, Sean M. Sykes‡, Paul W. Sadow‡, Kurt P. D’Andrea‡, Cheryl Bowman‡, Claire Fujii‡, Stacey A. Garland‡, Tanya M. Mason‡, Gary J. Olsen‡, Claire M. Fraser‡, Hamilton O. Smith‡, Carl R. Woese† & J. Craig Venter*

* The Institute for Genomic Research (TIGR), Rockville, Maryland 20850, USA
† Department of Microbiology, University of Illinois, Champaign-Urbana, Illinois 61801, USA
‡ Mathematics and Computer Science Division, Argonne National Laboratory, Illinois 60439, USA

**Archaeoglobus fulgidus** is the first sulphur-metabolizing organism to have its genome sequence determined. Its genome of 2,178,400 base pairs contains 2,436 open reading frames (ORFs). The information processing systems and the biosynthetic pathways for essential components (nucleotides, amino acids and cofactors) have extensive correlation with their counterparts in the archaeon *Methanococcus jannaschii*. The genomes of these two Archaea indicate dramatic differences in the way these organisms sense their environment, perform regulatory and transport functions, and gain energy. In contrast to *M. jannaschii*, *A. fulgidus* has fewer restriction–modification systems, and none of its genes appear to contain inteins. A quarter (651 ORFs) of the *A. fulgidus* genome encodes functionally uncharacterized yet conserved proteins, two-thirds of which are shared with *M. jannaschii* (428 ORFs). Another quarter of the genome encodes new proteins indicating substantial archaean gene diversity.

Biological sulphate reduction is part of the global sulphur cycle, ubiquitous in the earth's anaerobic environments, and is essential to the basal workings of the biosphere. Growth by sulphate reduction is restricted to relatively few groups of prokaryotes; all but one of these are Eubacteria, the exception being the archaeal sulphate reducers in the Archaeoglobales. These organisms are unique in that they are unrelated to other sulphate reducers, and because they grow at extremely high temperatures. The known Archaeoglobales are strict anaerobes, most of which are hyperthermophilic marine sulphate reducers found in hydrothermal environments and in subsurface oil fields. High-temperature sulphate reduction by Archaeoglobus species contributes to deep subsurface oil well 'souring' by producing iron sulphide, which causes corrosion of iron and steel in oil- and gas-processing systems.

*Archaeoglobus fulgidus* VC-16 (refs 2, 4) is the type strain of the Archaeoglobales. Cells are irregular spheres with a glycoprotein envelope and monopolar flagella. Growth occurs between 60 and 95°C, with optimum growth at 83°C and a minimum division time of 4 h. The organism grows organoheterotrophically using a variety of both carbon and energy sources, but can grow lithoautotrophically on hydrogen, thiosulphate and carbon dioxide. We sequenced the genome of *A. fulgidus* strain VC-16 as an example of a sulphur-metabolizing organism and to gain further insight into the Archaea through genomic comparison with *Methanococcus jannaschii*.

**General features of the genome**

The genome of *A. fulgidus* consists of a single, circular chromosome of 2,178,400 base pairs (bp) with an average of 48.5% G+C content (Fig. 1). There are three regions with low G+C content (<39%), two rich in genes encoding enzymes for lipopolysaccharide (LPS) biosynthesis, and two regions of high G+C content (>53%), containing genes for large ribosomal RNAs, proteins involved in haem biosynthesis (hemAB), and several transporters (Table 1). Because the origins of replication in Archaea are not characterized, we arbitrarily designated base pair one within a presumed non-coding region upstream of one of three areas containing multiple short repeat elements.

**Open reading frames.** Two independent coding analysis programs and BLAST searches (see Methods) predicted 2,436 ORFs (Figs 1, 2, Tables 1, 2) covering 92.2% of the genome. The average size of the *A. fulgidus* ORFs is 822 bp, similar to that of *M. jannaschii* (856 bp), but smaller than that in the completely sequenced eubacterial genomes (949 bp). All ORFs were searched against a non-redundant protein database, resulting in 1,797 putative identifications that were assigned biological roles within a classification system adapted from ref. 11. Predicted start codons are 76% ATG, 22% GTG and 2% TTG. Unlike *M. jannaschii*, where 18 inteins were found in coding regions, no inteins were identified in *A. fulgidus*. Compared with *M. jannaschii*, *A. fulgidus* contains a large number of gene duplications, contributing to its larger genome size. The average protein relative molecular mass (Mr) in *A. fulgidus* is 29,753, ranging from 1,939 to 266,571, similar to that observed in other prokaryotes. The isoelectric point (pl) of predicted proteins among sequenced prokaryotes exhibits a bimodal distribution with peaks at plS of approximately 5.5 and 10.5. The exceptions to this are *Mycoplasma genitalium* in which the distribution is skewed towards high pl
Figure 1  Circular representation of the *A. fulgidus* genome. The outer circle shows predicted protein-coding regions on the plus strand classified by function according to the colour code in Fig. 2 (except for unknowns and hypotheticals, which are in black). Second circle shows predicted protein-coding regions on the minus strand. Third and fourth circles show IS elements (red) and other repeats (green) on the plus and minus strand. Fifth and sixth circles show tRNAs (blue), rRNAs (red) and sRNAs (green) on the plus and minus strand, respectively.

Table 1  Genome features

| General                  |                       |
|--------------------------|------------------------|
| Chromosome size:         | 2,178,400 bp           |
| Protein coding regions:  | 92.2%                  |
| Stable RNAs:             | 0.4%                   |

| Predicted protein coding sequences: |                       |
|-------------------------------------|------------------------|
| Identified by database match:       | 2,436 (11 per kb)      |
| putative function assigned:         | 1,797                  |
| homologues of *M. jannaschii* ORFs: | 1,096                  |
| conserved hypothetical proteins:    | 916                    |
| No database match:                  | 639                    |
| Members of 242 paralogous families: | 719                    |
| Members of 18 families with known functions: | 475 |

| Stable RNAs | Coordinates              |
|-------------|--------------------------|
| 16S rRNA:   | 1,790,478–1,788,987      |
| 23S rRNA:   | 1,788,751–1,785,820      |
| 5S rRNA:    | 81,144–81,021            |
| 7S RNA:     | 798,067–798,376          |
| RNase P:    | 86,281–86,032            |

| Distinct G+C content regions | Coordinates          |
|------------------------------|----------------------|
| HGC-1, >53% G+C             | 1,786,000–1,797,000  |
| HGC-2, >53% G+C             | 2,158,000–2,159,000  |
| LGC-1, <39% G+C             | 281,000–284,000      |
| LGC-2, <39% G+C             | 544,000–550,000      |
| LGC-3, <39% G+C             | 1,175,000–1,177,000  |

| Short, non-coding repeats | Coordinates          |
|---------------------------|----------------------|
| SR-1A, CTTCCAATCCGATTTGCTGGATATATCACAC | 474–2,113          |
| SR-1B, CTTCCAATCCGATTTGCTGGATATATCACAC | 398,926–401,990     |
| SR-2, CTTCCAATCCGATTTGCTGGATATATCACAC | 1,690,930–1,694,104 |

| Long, coding repeats | Length | Copy number |
|----------------------|--------|-------------|
| LR-01 NADH-flavin oxidoreductase | 1,886 bp | 2 copies |
| LR-02 NisU, NisU + ORF | 1,549 bp | 2 copies |
| LR-03 ISA1214 putative transposase + ISORF2 | 1,214 bp | 6 copies |
| LR-04 ISA1083 putative transposase + ISORF2 | 1,083 bp | 3 copies |
| LR-05 type II secretion system protein | 1,014 bp | 4 copies |
| LR-06 ISA0983 putative transposase | 963 bp | 7 copies |
| LR-07 homologue of MJ0794 | 836 bp | 3 copies |
| LR-08 conserved hypothetical protein | 696 bp | 2 copies |
| LR-09 conserved hypothetical protein | 628 bp | 2 copies |
(median, 9.8) and A. fulgidus where the skew is toward low pl (median, 6.3).

**Multigene families.** In A. fulgidus 719 genes (30% of the total) belong to 242 families with two or more members (Table 1). Of these families, 157 contained genes with biological roles. Most of these families contain genes assigned to the ‘energy metabolism’, ‘transport and binding proteins’, and ‘fatty acid and phospholipid metabolism’ categories (Table 2). The superfamily of ATP-binding subunits of ABC transporters is the largest, containing 40 members. The importance of catabolic degradation and signal recognition systems is reflected by the presence of two large superfamilies: acyl-CoA ligases and signal-transducing histidine kinases. A. fulgidus does not contain a homologue of the large 16-member family found in M. jannaschii.

**Repetitive elements.** Three regions of the A. fulgidus genome contain short (<40 bp) direct repeats (Table 1). Two regions (SR-1A and SR-1B) contain 48 and 60 copies, respectively, of an identical 30-bp repeat interspersed with unique sequences averaging 40 bp. The third region (SR-2) contains 42 copies of a 37-bp repeat similar in sequence to the SR-1 repeat and interspersed with unique sequence averaging 41 bp. These repeated sequences are similar to the short repeated sequences found in M. jannaschii.

Nine classes of long (>500 bp) repeated sequences with ≥95% sequence identity were found (LR1-LR9; Table 1). LR-3 is a novel element with 14-bp inverted repeats and two genes, one of which has weak similarity to a transposase from Halobacterium salinarium. One copy of LR-3 interrupts AF0290, a homologue of a large M. jannaschii gene encoding a protein of unknown function. LR-4 and LR-6 encode putative transposases not identified in M. jannaschii that may represent IS elements. The remaining LR elements are not similar to known IS elements.

**Central intermediary and energy metabolism**

Sulphur oxide reduction may be the dominant respiratory process in anaerobic marine and freshwater environments, and is an important aspect of the sulphur cycle in anaerobic ecosystems. In this pathway, sulphate (SO$_4^{2-}$) is first activated to adenylylsulphate (adenosine-5’-phosphosulphate; APS), then reduced to sulphite and subsequently to sulphide$^{11}$ (Fig. 3). The most important enzyme in dissimilatory sulphate reduction, adenylylsulphate reductase, reduces the activated sulphate to sulphite, releasing AMP. In A. fulgidus, the APS reductase has a high degree of similarity and identical physiological properties to APS reductases in sulphate-reducing delta proteobacteria. A desulphoviridin-type sulphite reductase then adds six electrons to sulphite to produce sulphide. As in the Eubacteria, three sulphite-reductase genes, dsrABD, constitute an operon. The genes for adenylylsulphate reductase and sulphate adenylyltransferase reside in a separate operon. In A. fulgidus, sulphate can be replaced as an electron acceptor by both thiosulphate (S$_2$O$_3^{2-}$) and sulphite (SO$_3^{2-}$), but not by elemental sulphur.

A. fulgidus VC-16 has been shown to use lactate, pyruvate, methanol, ethanol, 1-propanol and formate as carbon and energy sources$^2$. Glucose has been described as a carbon source$^1$, but neither an uptake-transporter nor a catabolic pathway could be identified. Although it has been reported that A. fulgidus is incapable of growth on acetate$^4$, multiple genes for acetyl-CoA synthetase (which converts acetate to acetyl-CoA) were found. The organism may degrade a variety of hydrocarbons and organic acids because of the presence of 57 β-oxidation enzymes, at least one lipase, and a minimum of five types of ferredoxin-dependent oxidoreductases (Fig. 3). The predicted β-oxidation system is similar to those in Eubacteria and mitochondria, and has not previously been described in the Archaea. Escherichia coli requires both the fadD and fadL gene products to import long-chain fatty acids across the cell envelope into the cytosol.$^{15}$ A. fulgidus has 14 acyl-CoA ligases related to FadD, but as expected given that it has no outer membrane, no FadL. In E. coli, FadB has several metabolic functions, but in A. fulgidus these functions seem to be distributed among separate enzymes. For example, AF0435 encodes an orthologue of enoyl-CoA hydratase and resembles the amino-terminal domain of FadB. This gene is immediately upstream of a gene encoding an orthologue of 3-hydroxyacyl-CoA dehydrogenase that resembles the carboxy-terminal domain of FadB.

Acetyl-CoA is degraded by A. fulgidus through a C$_4$-pathway, not by the citric acid cycle or glyoxylate bypass.$^{16,17}$ This degradation is catalysed through the carbon monoxide dehydrogenase (CODH) pathway that consists of a five-subunit acetyl-CoA decarboxylase/synthase complex (ACDS) and five enzymes that are typically involved in methanogenesis.$^{18}$ In A. fulgidus, however, reverse methanogenesis occurs, resulting in CO$_2$ production. All of the enzymes and cofactors of methanogenesis from formylmethanofuran to N$_2$-methyltetrahydromethanopterin are used, but the absence of methyl-CoM reductase eliminates the possibility of methane production by conventional pathways. Production of trace amounts of methane (<0.1 μmol ml$^{-1}$) is probably a result of the reduction of N$_2$-methyltetrahydromethanopterin to methane and tetrahydromethanopterin by carbon monoxide (CO) dehydrogenase.

A. fulgidus also contains genes suggesting it has a second CO dehydrogenase system, homologous to that which enables Rhodospirillum rubrum to grow without light using CO as its sole energy source. Genes were detected for the nickel-containing CO dehydrogenase (CooS), an iron–sulphur redox protein, and a protein associated with the incorporation of nickel in CooS. These represent elements of a system that could catalyse the conversion of CO and H$_2$O to CO$_2$ and H$_2$.

In contrast to M. jannaschii, A. fulgidus contains genes representing multiple catabolic pathways. Systems include CoA-SH-dependent ferredoxin oxidoreductases specific for pyruvate, 2-ketoisovalerate, 2-ketoglutarate and indolepyruvate, as well as a 2-oxoacid with little substrate specificity.$^{19,20}$ Four genes with similarity to the tungsten-dependent oxidoreductase were also found.$^{22}$

Biochemical pathways characteristic of eubacterial metabolism, including the pentose-phosphate pathway, the Entner–Doudoroff pathway, glycolysis and gluconeogenesis, are either completely absent or only partly represented (Fig. 3). A. fulgidus does not have typical eubacterial polysaccharide biosynthesis machinery, yet it has been shown to produce a protein and carbohydrate-containing biofilm.$^{22}$ Nitrogen is obtained by importing inorganic molecules or degrading amino acids (Fig. 3); neither a glutamate dehydrogenase nor a relevant fix or nif gene is present.

The F$_{\text{O}_{\text{H}}}$$^2$H$_2$O-Q oxidoreductase complex$^{14}$ is recognized as...
the main generator of proton-motive force. However, our analysis indicates the presence of heterodisulphide reductase and several molybdoenzyme-binding oxidoreductases, with polysulphide, nitrate, dimethyl sulphoxide, and thiosulphate as potential substrates, which might contribute to energizing the cell membrane.

**Figure 3** An integrated view of metabolism and solute transport in *A. fulgidus*. Biochemical pathways for energy production, biosynthesis of organic compounds, and degradation of amino acids, aldehydes and acids are shown with the central components of A. fulgidus metabolism, sulphate, lactate and acetate-CoA highlighted. Pathways or steps for which no enzymes were identified are represented by a red arrow. A question mark is attached to pathways that could not be completely elucidated. Membrane-associated reactions that establish the proton-motive force (PMF) and generate ATP (electron transport chain and V-VATPase) are linked to cytosolic pathways for proton production. The oxalate-formate antiporters (oxT) may also contribute to the PMF by mediating electron-coupled anion exchange. Each gene product with a predicted function in ion or solute transport is illustrated. Proteins are grouped by substrate specificity with transporters for cations (green), anions (red), carbohydrates (alcohols/ acids (yellow), and amino acids/peptides/amine (blue) depicted. Ion-coupled permeases are represented by ovals (green), diamonds and circles (proWVX, glnHP, dppABCDF, potABCD, braCDFEG, hemUV, nrtBC, catABC, rbsAC, rbsAB). Genes corresponding to protein products are indicated. The central metabolic pathways are dipeptides

**Figure 3** An integrated view of metabolism and solute transport in *A. fulgidus*. Biochemical pathways for energy production, biosynthesis of organic compounds, and degradation of amino acids, aldehydes and acids are shown with the central components of A. fulgidus metabolism, sulphate, lactate and acetate-CoA highlighted. Pathways or steps for which no enzymes were identified are represented by a red arrow. A question mark is attached to pathways that could not be completely elucidated. Membrane-associated reactions that establish the proton-motive force (PMF) and generate ATP (electron transport chain and V-VATPase) are linked to cytosolic pathways for proton production. The oxalate-formate antiporters (oxT) may also contribute to the PMF by mediating electron-coupled anion exchange. Each gene product with a predicted function in ion or solute transport is illustrated. Proteins are grouped by substrate specificity with transporters for cations (green), anions (red), carbohydrates (alcohols/ acids (yellow), and amino acids/peptides/amine (blue) depicted. Ion-coupled permeases are represented by ovals (green), diamonds and circles (proWVX, glnHP, dppABCDF, potABCD, braCDFEG, hemUV, nrtBC, catABC, rbsAC, rbsAB). Genes corresponding to protein products are indicated. The central metabolic pathways are dipeptides

**Figure 3** An integrated view of metabolism and solute transport in *A. fulgidus*. Biochemical pathways for energy production, biosynthesis of organic compounds, and degradation of amino acids, aldehydes and acids are shown with the central components of A. fulgidus metabolism, sulphate, lactate and acetate-CoA highlighted. Pathways or steps for which no enzymes were identified are represented by a red arrow. A question mark is attached to pathways that could not be completely elucidated. Membrane-associated reactions that establish the proton-motive force (PMF) and generate ATP (electron transport chain and V-VATPase) are linked to cytosolic pathways for proton production. The oxalate-formate antiporters (oxT) may also contribute to the PMF by mediating electron-coupled anion exchange. Each gene product with a predicted function in ion or solute transport is illustrated. Proteins are grouped by substrate specificity with transporters for cations (green), anions (red), carbohydrates (alcohols/ acids (yellow), and amino acids/peptides/amine (blue) depicted. Ion-coupled permeases are represented by ovals (green), diamonds and circles (proWVX, glnHP, dppABCDF, potABCD, braCDFEG, hemUV, nrtBC, catABC, rbsAC, rbsAB). Genes corresponding to protein products are indicated. The central metabolic pathways are dipeptides
four-electron reduction of molecular oxygen to water, with the concurrent regeneration of NAD.

**Transports**

*A. fulgidus* may synthesize several transporters for the import of carbon-containing compounds, probably contributing to its ability to switch from autotrophic to heterotrophic growth. Both *M. jannaschii* and *A. fulgidus* have branched-chain amino-acid ABC transport systems and a transporter for the uptake of arginine and lysine. *A. fulgidus* encodes proteins for dipeptide, spermidine/putrescine, proline/glycine-betaine and glutamine uptake, as well as transporters for sugars and acids, rather like the membrane systems described in eubacterial heterotrophs. These compounds provide the necessary substrates for numerous biosynthetic and degradative pathways (Fig. 3).

Many *A. fulgidus* redox proteins are predicted to require iron. Correspondingly, iron transporters have been identified for the import of both oxidized (FeIII) and reduced (FeII) forms of iron. There are duplications in functional and regulatory genes in both systems. The uptake of FeIII may depend on haemin or a haemin-like compound because *A. fulgidus* has orthologues to the eubacterial hem transport systems, HemU and HemV. *A. fulgidus* may also use the regulatory protein Fur to modulate FeIII transport; this protein is not present in *M. jannaschii*. FeII uptake occurs through a modified Feo system containing FeoB. This is the third example of an isolated feoB gene: *M. jannaschii* and *Helicobacter pylori* also appear to lack feoA, implying that FeoA is not essential for iron transport in these organisms.

A complex suite of proteins regulates ionic homeostasis. Ten distinct transporters facilitate the flux of the physiological ions K+, Na+, Mg2+, Fe2+, Fe3+, NO3, SO4 and inorganic phosphate (P). Most of these transporters have homologues in *M. jannaschii* and are therefore likely to be critical for nutrient acquisition during autotrophic growth. *A. fulgidus* has additional ion transporters for the elimination of toxic compounds including copper, cyanate and arsenite. As in *M. jannaschii*, the *A. fulgidus* genome contains two paralogous operons of cobalamin biosynthesis-cobalt transporters, *ctbM*QO.

**Sensory functions and regulation of gene expression**

Consistent with its extensive energy-producing metabolism and versatile system for carbon utilization, *A. fulgidus* has complex sensory and regulatory networks. These networks contain over 55 proteins with presumed regulatory functions, including members of the ArsR, AsnC and Sir2 families, as well as several iron-dependent repressor proteins. There are at least 15 signal-transducing histidine kinases, but only nine response regulators; this difference suggests there is a high degree of cross-talk between kinases and regulators. Only four response regulators appear to be in operons with histidine kinases, including those in the methyl-directed chemotaxis system (Che), which lies adjacent to the flagellar biosynthesis operon. Although rich in regulatory proteins, *A. fulgidus* apparently lacks regulators for response to amino-acid and carbon starvation as well as to DNA damage. Finally, *A. fulgidus* contains a homologue of the mammalian mitochondrial benzodiazepine receptor, which functions as a sensor in signal-transduction pathways. These receptors have been previously identified only in Proteobacteria and Cyanobacteria.

**Replication, repair and cell division**

*A. fulgidus* possesses two family B DNA polymerases, both related to the catalytic subunit of the eukaryal delta polymerase, as previously observed in the Sulfolobales. It also has a homologue of the proofreading ε subunit of *E. coli* Pol III, not previously observed in the Archaea. The DNA repair system is more extensive than that found in *M. jannaschii*, including a homologue of the eukaryal Rad25, a 3-methyladenine DNA glycosylase, and exodeoxyribonuclease III. As well as reverse gyrase, topoisomerase I (ref. 9), and topoisomerase VI (ref. 27), the genes for the first archaean DNA gyrase were identified.

*A. fulgidus* lacks a recognizable type II restriction-modification system, but contains one type I system. In contrast, two type II and three type I systems were identified in *M. jannaschii*. No homologue of the *M. jannaschii* thermochrome was identified.

The cell-division machinery is similar to that of *M. jannaschii*, with orthologues of eubacterial fts and euukaryal cdc genes. However, several cdc genes found in *M. jannaschii*, including homologues of cdc23, cdc27, cdc47 and cdc54, appear to be absent in *A. fulgidus*.

**Transcription and translation**

*A. fulgidus* and *M. jannaschii* have transcriptional and translational systems distinct from their eubacterial and euukaryal counterparts. In both, the RNA polymerase contains the large universal subunits and five smaller subunits found in both Archaea and euukaryotes. Transcription initiation is a simplified version of the euukaryotic mechanism. However, *A. fulgidus* alone has a homologue of euukaryotic TBP-interacting protein 49 not seen in *M. jannaschii*, but apparently present in *Sulfolobus solfataricus*.

Translation in *A. fulgidus* parallels *M. jannaschii* with a few exceptions. The organism has only one tRNA operon with an Ala-tRNA gene in the spacer and lacks a contiguous 5S tRNA gene. Genes for 46 tRNAs were identified, five of which contain introns in the anticodon region that are presumably removed by the intron excision enzyme EndA. The gene for selenocysteine tRNA (SelC) was not found, nor were the genes for SelA, SelB and SelD. With the exception of Asp-tRNA2C and Val-tRNA3C, tRNA genes are not linked in the *A. fulgidus* genome. The RNA component of the tRNA maturation enzyme RNase P is present. Both *A. fulgidus* and *M. jannaschii* appear to possess an enzyme that inserts the tRNA-modified nucleoside archaeosine, but only *A. fulgidus* has the related enzyme that inserts the modified base queine.

Both *A. fulgidus* and *M. jannaschii* lack glutamine synthetase and asparagine synthetase; the relevant tRNAs are presumably aminoacylated with glutamic and aspartic acids, respectively. An enzymatic in situ transamidation then converts the amino acid to its amide form, as seen in other Archaea and in Gram-positive Eubacteria. Indeed, genes for the three subunits of the Glu-tRNA amidotransferase (gitABC) have been identified in *A. fulgidus*. The Lys aminoacyl-tRNA synthetase in both organisms is a class I-type, not a class II-type. *A. fulgidus* possesses a normal tRNA synthetase for both Cys and Ser, unlike *M. jannaschii* in which the former was not identifiable and the latter was unusual.

*M. jannaschii* has a single gene belonging to the TCP-1 chaperonin family, whereas *A. fulgidus* has two that encode subunits α and β of the thermosome. Phylogenetic analysis of the archaeal TCP-1 family indicates that these *A. fulgidus* genes arose by a recent species-specific gene duplication, as is the case for the two subunits of the Thermoplasma acidophilum thermosome and the *Sulfolobus shibatae* rosettosome. As in *M. jannaschii*, no dnaK gene was identified.

**Biosynthesis of essential components**

Like most autotrophic microorganisms, *A. fulgidus* is able to synthesize many essential compounds, including amino acids, cofactors, carriers, purines and pyrimidines. Many of these biosynthetic pathways show a high degree of conservation between *A. fulgidus* and *M. jannaschii*. These two Archaea are similar in their biosynthetic pathways for siroheme, cobalamin, molybdopterin, riboflavin, thiamin and nicotinate, the role category with greatest conservation between these two organisms being amino-acid bio-synthesis. Of 78 *A. fulgidus* genes assigned to amino-acid biosynthetic pathways, at least 73 (94%) have homologues in *M. jannaschii*. For both archaean species, amino-acid biosynthetic pathways resemble those of *Bacillus subtilis* more closely than
those of E. coli. For example, in A. fulgidus and M. jannaschii, tryptophan biosynthesis is accomplished by seven enzymes, TrpA, B, C, D, E, F, G as in B. subtilis, rather than by five enzymes, TrpA, B, C, D, E (including the bifunctional TrpC and TrpD) as found in E. coli.

No biotin biosynthetic genes were identified, yet biotin can be detected in A. fulgidus cell extracts\(^4,\) and several genes encode a biotin-binding consensus sequence. Similarly, A. fulgidus lacks the genes for pyridoxine biosynthesis although pyridoxine can be found in cell extracts (albeit at lower levels than seen in E. coli and several Archaea\(^6\)). No gene encoding ferrochelatase, the terminal enzyme in haem biosynthesis, has been identified, although A. fulgidus is known to use cytochromes\(^4,\)\(^5,\)\(^6\). These cofactors may be obtained by mechanisms that we have not recognized. Although all of the enzymes required for pyrimidine biosynthesis appear to be present, three enzymes in the purine pathway (GAR transformylase, AICAR formyltransferase and the ATPase subunit of AIR carboxylase) have stereochemistry\(^3\) for which there are multiple biosynthetic lipids containing a glycerophosphate backbone with a 2,3-...
4. Stetter, K. O. *Archaeoglobus fulgidus* gen. nov., sp. nov: a new taxon of extremely thermophilic archaea. *Syst. Appl. Microbiol.* 10, 172–173 (1988).

5. Stetter, K. O. et al. Hyperthermophilic archaea are thriving in deep North Sea and Alaskan oil reservoirs. *Nature* 363, 743–745 (1993).

6. Vorholt, J., Kunow, J., Stetter, K. O. & Thauer, R. K. Enzymes and cofactors of the carbon monoxide dehydrogenase pathway for autotrophic CO2 fixation in *Archaeoglobus lutidihydrogenicus* and the lack of carbon monoxide dehydrogenase in the heterotrophic *A. profundum*. *Arch. Microbiol.* 163, 112–118 (1995).

7. Weese, C. R. & Fox, G. E. Phylogenic structure of the prokaryotic domain: The primary kingdoms. *Proc. Natl Acad. Sci. USA* 74, 5088–5090 (1977).

8. Weese, C. R., Kandler, O. & Whitt, W. L. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl Acad. Sci. USA* 87, 4576–4579 (1990).

9. Bull, C. J. et al. Complete genome sequence of the methanogenic archaon *Methanococcus jannaschii*. *Science* 273, 1058–1073 (1996).

10. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410 (1990).

11. Riley, M. Functions of gene products of *Escherichia coli*. *Microbiol. Rev.* 57, 862–952 (1993).

12. Cooling, F. B. III, Maloney, C. L., Nagel, E., Tabinowski, J. & Odorn, J. M. Inhibition of sulfate respiration by L-3-dehydroxyanthraquinone and other anthraquinone derivatives. *Appl. Environ. Microbiol.* 62, 3099–3104 (1996).

13. Thauer, R. R. & Kunow, J. in *Sulfur Reducing Bacteria* (ed. Barton, L. L.) 33–48 (Plenum, New York, 1995).

14. Speich, D. et al. Adenylylsulfate reductase from the sulfite-reducing archaon *Archaeoglobus fulgidus*: cloning and characterization of the genes and comparison of the enzyme with other iron-sulfur flavoproteins. *Microbiology* 140, 1273–1284 (1994).

15. Clark, D. P. & Cronan, J. E. Jr in *Escherichia coli* (ed. Neidhardt, F. C.) 343–357 (ASM Press, Washington DC, 1996).

16. Moller-zirkhan, D. & Thauer, R. K. Anaerobic lactate oxidation to 3 CO2 by *Acetobacterium woodii*. *Arch. Microbiol.* 166, 216–220 (1997).

17. Schauder, R., Eikmanns, B., Thauer, R. K. & Widdel, F. Acetate oxidation to CO2 in *Acetobacterium woodii*. *Arch. Microbiol.* 163, 112–118 (1995).

18. Dai, Y.-R. in *Anaerobic Lactic Acid Fermentation* (ed. Sato, K.) 133–144 (Springer, Tokyo, 1997).

19. Zhang, Q., Iwasaki, T., Wakagi, T. & Oshima, T. 2-oxoacid:ferredoxin oxidoreductase from the thermoacidophilic archaeon, *Thermococcus litoralis*. *Eur. J. Biochem.* 232, 397–403 (1995).

20. Marsh, T. L., Reich, C. I., Whitlock, R. B. & Olsen, G. J. Transcription factor IID in the Archaea: the minimal gene complement of *Haloferax volcanii*. *Proc. Natl Acad. Sci. USA* 118, 11819–11824 (1991).

21. Huber, R. et al. Isolation of a hyperthermophilic archaeon predicted by in situ RNA analysis. *Nature* 376, 57–58 (1995).

22. Kleczkowska, M. et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae Rd*. *Science* 284, 331–338 (1999).

23. Getz, V. et al. Acetyl-CoA decarboxylase/synthase complex from *Archaeoglobus fulgidus*: purification, characterization, and properties. *Arch. Microbiol.* (submitted).

24. Hedges, C. P., Kandler, O. & Thauer, R. K. Acetate oxidation to CO2 in archaea: a novel pathway not involving reactions of the citric-acid cycle. *Arch. Microbiol.* 165, 216–220 (1996).

25. Huber, R. et al. A novel archaean transcription factor related to TATA-box-binding protein and transcription factor II A of the eukaryotic RNA polymerase. *Nature* 376, 57–58 (1995).

26. Edgell, D. R., Klenk, H.-P. & Doolittle, W. F. Gene duplications in evolution of archaeal family B DNA polymerases. *J. Bacteriol.* 179, 2632–2640 (1997).

27. Berge et al. Characterization of the TATA-box-binding protein from thermococcal strains. *Nature* 386, 414–417 (1997).

28. Shih, L. T., Reich, C. I., Whitlock, R. B. & Olsen, G. J. Transcription factor IID in the Archaea: sequences in the *Thermococcus celer* genome would encode a product closely related to the TATA-box-binding protein of eukaryotes. *Proc. Natl Acad. Sci. USA* 91, 4180–4184 (1994).

29. Sass, F. P., Gong, G., DeDecker, B. S. & Sigler, P. B. The 2.1-A crystal structure of an archaeal preinitiation complex: TATA-box-binding protein/transcription factor II B of *H. volcanii*. *Proc. Natl Acad. Sci. USA* 94, 6042–6047 (1997).

30. Woese, C. R., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl Acad. Sci. USA* 87, 4576–4579 (1990).

31. Koga, H. K. et al. The 60 kDa heat shock proteins in the hyperthermophlic archaeon *Haloferax volcanii*. *J. Mol. Biol.* 253, 712–723 (1995).

32. Noll, K. M. B. & Barber, T. S. Vitamin contents of archaebacteria. *J. Bacteriol.* 170, 4313–4321 (1988).

33. Kagawa, H. K. *et al.* The 60 kDa heat shock proteins in the hyperthermophlic archaeon *Haloferax volcanii*. *J. Mol. Biol.* 253, 712–723 (1995).

34. Sonnhammer, E. L., Eddy, S. R. & Durbin, R. Pfam: A comprehensive database of protein families. *Nucleic Acids Res.* 27, 29–37 (1999).

35. Huber, R. et al. Isolation of a hyperthermophilic archaeon predicted by in situ RNA analysis. *Nature* 376, 57–58 (1995).

36. Krauss, A. et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae Rd*. *Science* 284, 331–338 (1999).

37. Fraher, R. D. et al. The minimal gene complement of *Mycoplasm a genitalium*. *Genome Res.* 10, 397–403 (2000).

38. Sonnhammer, E. L., Eddy, S. R. & Durbin, R. Pfam: A comprehensive database of protein families based on seed alignments. *Proteins* 28, 405–420 (1997).

39. Papp, M. & von Heijne, G. TopPred II: a new tool for predicting membrane protein structure predictions. *Comput. Appl. Biosci.* 16, 485–486 (2000).

Acknowledgements. We thank M. Heaney, J. Scott and R. Shirley for software and database support; V. Sapiro, V. Vincent, J. Mehan and P. Mass for computer system support; B. Cameron and D. J. Doyle for editorial assistance: and K. O. Stetter for providing A. fulgidus WC–16. This work was supported by the US Department of Energy.

Correspondence and requests for materials should be addressed to J.C.V. (e-mail: gaf@tigr.org). The annotated genome sequence and the gene family alignments are available on the World-Wide Web at http://www.tigr.org/db/mbd/afdb/afdb.html. The sequence has been deposited in Genbank with accession number AEH000782.
### Table 2. List of A. fulgidus genes with putative identification. Gene numbers correspond to those in Fig. 2. Percentages

| Gene     | Description                                                                 | Percentage |
|----------|-----------------------------------------------------------------------------|------------|
| AF1365   | hcaA, bacteriochlorophyll synthase, 63 kDa subunit                           | 51.2%      |
| AF0140   | hcaB, bacteriochlorophyll synthase, 63 kDa subunit                           | 44.7%      |
| AF1385   | hcaC, bacteriochlorophyll synthase, 63 kDa subunit                           | 38.2%      |
| AF1386   | hcaD, bacteriochlorophyll synthase, 63 kDa subunit                           | 37.8%      |
| AF1366   | hcaE, bacteriochlorophyll synthase, 63 kDa subunit                           | 35.1%      |
| AF1387   | hcaF, bacteriochlorophyll synthase, 63 kDa subunit                           | 30.8%      |
| AF1388   | hcaG, bacteriochlorophyll synthase, 63 kDa subunit                           | 30.4%      |
| AF1389   | hcaH, bacteriochlorophyll synthase, 63 kDa subunit                           | 28.3%      |
| AF1390   | hcaI, bacteriochlorophyll synthase, 63 kDa subunit                           | 25.0%      |
| AF1391   | hcaJ, bacteriochlorophyll synthase, 63 kDa subunit                           | 23.8%      |
| AF1392   | hcaK, bacteriochlorophyll synthase, 63 kDa subunit                           | 22.6%      |
| AF1393   | hcaL, bacteriochlorophyll synthase, 63 kDa subunit                           | 22.2%      |
| AF1394   | hcaM, bacteriochlorophyll synthase, 63 kDa subunit                           | 21.9%      |
| AF1395   | hcaN, bacteriochlorophyll synthase, 63 kDa subunit                           | 21.6%      |
| AF1396   | hcaO, bacteriochlorophyll synthase, 63 kDa subunit                           | 20.2%      |
| AF1397   | hcaP, bacteriochlorophyll synthase, 63 kDa subunit                           | 19.7%      |
| AF1398   | hcaQ, bacteriochlorophyll synthase, 63 kDa subunit                           | 19.3%      |
| AF1399   | hcaR, bacteriochlorophyll synthase, 63 kDa subunit                           | 18.9%      |
| AF1400   | hcaS, bacteriochlorophyll synthase, 63 kDa subunit                           | 18.5%      |
| AF1401   | hcaT, bacteriochlorophyll synthase, 63 kDa subunit                           | 18.1%      |
| AF1402   | hcaU, bacteriochlorophyll synthase, 63 kDa subunit                           | 17.7%      |
| AF1403   | hcaV, bacteriochlorophyll synthase, 63 kDa subunit                           | 17.3%      |
| AF1404   | hcaW, bacteriochlorophyll synthase, 63 kDa subunit                           | 16.9%      |
| AF1405   | hcaX, bacteriochlorophyll synthase, 63 kDa subunit                           | 16.5%      |
| AF1406   | hcaY, bacteriochlorophyll synthase, 63 kDa subunit                           | 16.1%      |
| AF1407   | hcaZ, bacteriochlorophyll synthase, 63 kDa subunit                           | 15.7%      |
| AF1408   | hcaAA, bacteriochlorophyll synthase, 63 kDa subunit                          | 15.3%      |
| AF1409   | hcaAB, bacteriochlorophyll synthase, 63 kDa subunit                          | 14.9%      |
| AF1410   | hcaAC, bacteriochlorophyll synthase, 63 kDa subunit                          | 14.5%      |
| AF1411   | hcaAD, bacteriochlorophyll synthase, 63 kDa subunit                          | 14.1%      |
| AF1412   | hcaAE, bacteriochlorophyll synthase, 63 kDa subunit                          | 13.7%      |
| AF1413   | hcaAF, bacteriochlorophyll synthase, 63 kDa subunit                          | 13.3%      |
| AF1414   | hcaAG, bacteriochlorophyll synthase, 63 kDa subunit                          | 12.9%      |
| AF1415   | hcaAH, bacteriochlorophyll synthase, 63 kDa subunit                          | 12.5%      |
| AF1416   | hcaAI, bacteriochlorophyll synthase, 63 kDa subunit                          | 12.1%      |
| AF1417   | hcaAJ, bacteriochlorophyll synthase, 63 kDa subunit                          | 11.7%      |
| AF1418   | hcaAK, bacteriochlorophyll synthase, 63 kDa subunit                          | 11.3%      |
| AF1419   | hcaAL, bacteriochlorophyll synthase, 63 kDa subunit                          | 10.9%      |
| AF1420   | hcaAM, bacteriochlorophyll synthase, 63 kDa subunit                          | 10.5%      |
| AF1421   | hcaAN, bacteriochlorophyll synthase, 63 kDa subunit                          | 10.1%      |
| AF1422   | hcaAO, bacteriochlorophyll synthase, 63 kDa subunit                          | 9.7%       |
| AF1423   | hcaAP, bacteriochlorophyll synthase, 63 kDa subunit                          | 9.3%       |
| AF1424   | hcaAQ, bacteriochlorophyll synthase, 63 kDa subunit                          | 8.9%       |
| AF1425   | hcaAR, bacteriochlorophyll synthase, 63 kDa subunit                          | 8.5%       |
| AF1426   | hcaAS, bacteriochlorophyll synthase, 63 kDa subunit                          | 8.1%       |
| AF1427   | hcaAT, bacteriochlorophyll synthase, 63 kDa subunit                          | 7.7%       |
| AF1428   | hcaAU, bacteriochlorophyll synthase, 63 kDa subunit                          | 7.3%       |
| AF1429   | hcaAV, bacteriochlorophyll synthase, 63 kDa subunit                          | 6.9%       |
| AF1430   | hcaAW, bacteriochlorophyll synthase, 63 kDa subunit                          | 6.5%       |
| AF1431   | hcaAX, bacteriochlorophyll synthase, 63 kDa subunit                          | 6.1%       |
| AF1432   | hcaAY, bacteriochlorophyll synthase, 63 kDa subunit                          | 5.7%       |
| AF1433   | hcaAZ, bacteriochlorophyll synthase, 63 kDa subunit                          | 5.3%       |
| AF1434   | hcaAA, bacteriochlorophyll synthase, 63 kDa subunit                          | 4.9%       |
| AF1435   | hcaAB, bacteriochlorophyll synthase, 63 kDa subunit                          | 4.5%       |
| AF1436   | hcaAC, bacteriochlorophyll synthase, 63 kDa subunit                          | 4.1%       |
| AF1437   | hcaAD, bacteriochlorophyll synthase, 63 kDa subunit                          | 3.7%       |
| AF1438   | hcaAE, bacteriochlorophyll synthase, 63 kDa subunit                          | 3.3%       |
| AF1439   | hcaAF, bacteriochlorophyll synthase, 63 kDa subunit                          | 2.9%       |
| AF1440   | hcaAG, bacteriochlorophyll synthase, 63 kDa subunit                          | 2.5%       |
| AF1441   | hcaAH, bacteriochlorophyll synthase, 63 kDa subunit                          | 2.1%       |
| AF1442   | hcaAI, bacteriochlorophyll synthase, 63 kDa subunit                          | 1.7%       |
| AF1443   | hcaAJ, bacteriochlorophyll synthase, 63 kDa subunit                          | 1.3%       |
| AF1444   | hcaAK, bacteriochlorophyll synthase, 63 kDa subunit                          | 0.9%       |
REGULATORY FUNCTIONS

| Gene          | Description                                      | Identity |
|---------------|--------------------------------------------------|----------|
| AF1392        | NDhex methyl-dehydrogenase, fumarate reductase   | 53.2%    |
| AF0204        | Ribose-5-phosphate 3-kinase                      | 46.9%    |
| AF0205        | Signal transduction histidine kinase             | 46.9%    |
| AF0431        | tyrosine kinase                                   | 32.4%    |
| AF1700        | Signal transduction histidine kinase             | 32.4%    |
| AF0402        | 2-oxoglutarate dehydrogenase                     | 30.2%    |
| AF1407        | Phenylpyruvate kinase                             | 44.4%    |
| AF1649        | Phenylpyruvate kinase                             | 59.3%    |
| AF1639        | Serine-1-phosphate phosphatase                   | 46.2%    |
| AF1646        | Protein phosphatase                               | 37.9%    |
| AF1259        | Signal transduction histidine kinase             | 35.8%    |
| AF1407        | Signal transduction histidine kinase             | 35.8%    |
| AF0811        | Signal transduction histidine kinase             | 35.8%    |
| AF1377        | Signal transduction histidine kinase             | 35.8%    |
| AF0034        | Signal transduction histidine kinase             | 35.8%    |
| AF1407        | Signal transduction histidine kinase             | 35.8%    |
| AF1259        | Signal transduction histidine kinase             | 35.8%    |
| AF1407        | Signal transduction histidine kinase             | 35.8%    |
| AF0811        | Signal transduction histidine kinase             | 35.8%    |
| AF1377        | Signal transduction histidine kinase             | 35.8%    |
| AF0034        | Signal transduction histidine kinase             | 35.8%    |
| AF1407        | Signal transduction histidine kinase             | 35.8%    |
| AF0811        | Signal transduction histidine kinase             | 35.8%    |
| AF1377        | Signal transduction histidine kinase             | 35.8%    |
| AF0034        | Signal transduction histidine kinase             | 35.8%    |
| AF1407        | Signal transduction histidine kinase             | 35.8%    |
| AF0811        | Signal transduction histidine kinase             | 35.8%    |
| AF1377        | Signal transduction histidine kinase             | 35.8%    |
| AF0034        | Signal transduction histidine kinase             | 35.8%    |
| AF1407        | Signal transduction histidine kinase             | 35.8%    |
| AF0811        | Signal transduction histidine kinase             | 35.8%    |
| AF1377        | Signal transduction histidine kinase             | 35.8%    |
| AF0034        | Signal transduction histidine kinase             | 35.8%    |
| AF1407        | Signal transduction histidine kinase             | 35.8%    |
| AF0811        | Signal transduction histidine kinase             | 35.8%    |
| AF1377        | Signal transduction histidine kinase             | 35.8%    |
| AF0034        | Signal transduction histidine kinase             | 35.8%    |
| AF1407        | Signal transduction histidine kinase             | 35.8%    |
| AF0811        | Signal transduction histidine kinase             | 35.8%    |
| AF1377        | Signal transduction histidine kinase             | 35.8%    |
| AF0034        | Signal transduction histidine kinase             | 35.8%    |
| AF1407        | Signal transduction histidine kinase             | 35.8%    |
| AF0811        | Signal transduction histidine kinase             | 35.8%    |
| AF1377        | Signal transduction histidine kinase             | 35.8%    |
| AF0034        | Signal transduction histidine kinase             | 35.8%    |
| AF1407        | Signal transduction histidine kinase             | 35.8%    |
| AF0811        | Signal transduction histidine kinase             | 35.8%    |
| AF1377        | Signal transduction histidine kinase             | 35.8%    |
| AF0034        | Signal transduction histidine kinase             | 35.8%    |
| AF1407        | Signal transduction histidine kinase             | 35.8%    |
| AF0811        | Signal transduction histidine kinase             | 35.8%    |
| AF1377        | Signal transduction histidine kinase             | 35.8%    |
| AF0034        | Signal transduction histidine kinase             | 35.8%    |
| AF1407        | Signal transduction histidine kinase             | 35.8%    |
| AF0811        | Signal transduction histidine kinase             | 35.8%    |
| AF1377        | Signal transduction histidine kinase             | 35.8%    |
| AF0034        | Signal transduction histidine kinase             | 35.8%    |
| AF1407        | Signal transduction histidine kinase             | 35.8%    |
| AF0811        | Signal transduction histidine kinase             | 35.8%    |
| AF1377        | Signal transduction histidine kinase             | 35.8%    |
| AF0034        | Signal transduction histidine kinase             | 35.8%    |
| AF1407        | Signal transduction histidine kinase             | 35.8%    |
| AF0811        | Signal transduction histidine kinase             | 35.8%    |
| AF1377        | Signal transduction histidine kinase             | 35.8%    |
| AF0034        | Signal transduction histidine kinase             | 35.8%    |
| AF1407        | Signal transduction histidine kinase             | 35.8%    |
| AF0811        | Signal transduction histidine kinase             | 35.8%    |
| AF1377        | Signal transduction histidine kinase             | 35.8%    |
| AF0034        | Signal transduction histidine kinase             | 35.8%    |
| AF1407        | Signal transduction histidine kinase             | 35.8%    |
| AF0811        | Signal transduction histidine kinase             | 35.8%    |
| AF1377        | Signal transduction histidine kinase             | 35.8%    |
| AF0034        | Signal transduction histidine kinase             | 35.8%    |
| AF1407        | Signal transduction histidine kinase             | 35.8%    |
| AF0811        | Signal transduction histidine kinase             | 35.8%    |
| AF1377        | Signal transduction histidine kinase             | 35.8%    |
| AF0034        | Signal transduction histidine kinase             | 35.8%    |
| AF1407        | Signal transduction histidine kinase             | 35.8%    |
| AF0811        | Signal transduction histidine kinase             | 35.8%    |
| AF1377        | Signal transduction histidine kinase             | 35.8%    |
| AF0034        | Signal transduction histidine kinase             | 35.8%    |
| AF1407        | Signal transduction histidine kinase             | 35.8%    |
| AF0811        | Signal transduction histidine kinase             | 35.8%    |
| AF1377        | Signal transduction histidine kinase             | 35.8%    |
| AF0034        | Signal transduction histidine kinase             | 35.8%    |
| AF1407        | Signal transduction histidine kinase             | 35.8%    |
| AF0811        | Signal transduction histidine kinase             | 35.8%    |
| AF1377        | Signal transduction histidine kinase             | 35.8%    |
| AF0034        | Signal transduction histidine kinase             | 35.8%    |
| AF1407        | Signal transduction histidine kinase             | 35.8%    |
| AF0811        | Signal transduction histidine kinase             | 35.8%    |
| AF1377        | Signal transduction histidine kinase             | 35.8%    |
| AF0034        | Signal transduction histidine kinase             | 35.8%    |
| AF1407        | Signal transduction histidine kinase             | 35.8%    |
| AF0811        | Signal transduction histidine kinase             | 35.8%    |
| AF1377        | Signal transduction histidine kinase             | 35.8%    |
| AF0034        | Signal transduction histidine kinase             | 35.8%    |
| AF1407        | Signal transduction histidine kinase             | 35.8%    |
| AF0811        | Signal transduction histidine kinase             | 35.8%    |
| AF1377        | Signal transduction histidine kinase             | 35.8%    |
| AF0034        | Signal transduction histidine kinase             | 35.8%    |
| AF1407        | Signal transduction histidine kinase             | 35.8%    |
| AF0811        | Signal transduction histidine kinase             | 35.8%    |
| AF1377        | Signal transduction histidine kinase             | 35.8%    |
| AF0034        | Signal transduction histidine kinase             | 35.8%    |
| AF1407        | Signal transduction histidine kinase             | 35.8%    |
| AF0811        | Signal transduction histidine kinase             | 35.8%    |
| AF1377        | Signal transduction histidine kinase             | 35.8%    |
| AF0034        | Signal transduction histidine kinase             | 35.8%    |
| AF1407        | Signal transduction histidine kinase             | 35.8%    |
| Protein Name | Description | Identity |
|-------------|-------------|----------|
| 38.0% | dipeptide ABC transporter, permease protein (dppC) | 38.0% |
| 38.0% | dipeptide ABC transporter, permease protein (dppB) | 38.0% |
| 38.0% | dipeptide ABC transporter, permease protein (dppA) | 38.0% |
| 38.0% | dipeptide ABC transporter, ATP-binding component (dppD) | 38.0% |
| 38.0% | dipeptide ABC transporter, ATP-binding component (dppE) | 38.0% |
| 38.0% | dipeptide ABC transporter, ATP-binding component (dppF) | 38.0% |
| 38.0% | dipeptide ABC transporter, ATP-binding component (dppG) | 38.0% |
| 38.0% | dipeptide ABC transporter, ATP-binding component (dppH) | 38.0% |

**Other Categories**

- **Transport proteins**:
  - ATP-dependent dipeptide ABC transporter, permease protein (dppC) 38.0%
  - ATP-dependent dipeptide ABC transporter, permease protein (dppB) 38.0%
  - ATP-dependent dipeptide ABC transporter, permease protein (dppA) 38.0%
  - ATP-dependent dipeptide ABC transporter, ATP-binding component (dppD) 38.0%
  - ATP-dependent dipeptide ABC transporter, ATP-binding component (dppE) 38.0%
  - ATP-dependent dipeptide ABC transporter, ATP-binding component (dppF) 38.0%
  - ATP-dependent dipeptide ABC transporter, ATP-binding component (dppG) 38.0%
  - ATP-dependent dipeptide ABC transporter, ATP-binding component (dppH) 38.0%

- **Carbohydrates, organic alcohols, and acids**:
  - L-glutamine synthetase (glnA) 38.0%
  - L-glutamate synthetase (gltA) 38.0%
  - L-glutamate synthetase (gltB) 38.0%
  - L-glutamate synthetase (gltC) 38.0%
  - L-glutamate synthetase (gltD) 38.0%

- **Transposon-related functions**:
  - IS100 insertion sequence (IS100) 40.0%
  - IS100 insertion sequence (IS100) 40.0%
  - IS100 insertion sequence (IS100) 40.0%
  - IS100 insertion sequence (IS100) 40.0%
  - ISA1083-1, putative transposase 40.0%
  - ISA1083-1, putative transposase 40.0%
  - ISA1083-1, putative transposase 40.0%
  - ISA1083-1, putative transposase 40.0%
  - ISA1083-1, putative transposase 40.0%

- **Drug and antibiotic resistance**:
  - Carbenicillin resistance membrane protein (cra) 40.0%
  - Carbenicillin resistance membrane protein (cra) 40.0%
  - Carbenicillin resistance membrane protein (cra) 40.0%
  - Carbenicillin resistance membrane protein (cra) 40.0%
  - Carbenicillin resistance membrane protein (cra) 40.0%

- **Regulatory functions**:
  - RNA polymerase subunit sigma 38.0%
  - RNA polymerase subunit sigma 38.0%
  - RNA polymerase subunit sigma 38.0%
  - RNA polymerase subunit sigma 38.0%
  - RNA polymerase subunit sigma 38.0%

- **Other**:
  - ATP-binding protein (dppA) 38.0%
  - ATP-binding protein (dppB) 38.0%
  - ATP-binding protein (dppC) 38.0%
  - ATP-binding protein (dppD) 38.0%
  - ATP-binding protein (dppE) 38.0%
  - ATP-binding protein (dppF) 38.0%
  - ATP-binding protein (dppG) 38.0%
  - ATP-binding protein (dppH) 38.0%

Nature © Macmillan Publishers Ltd 1997