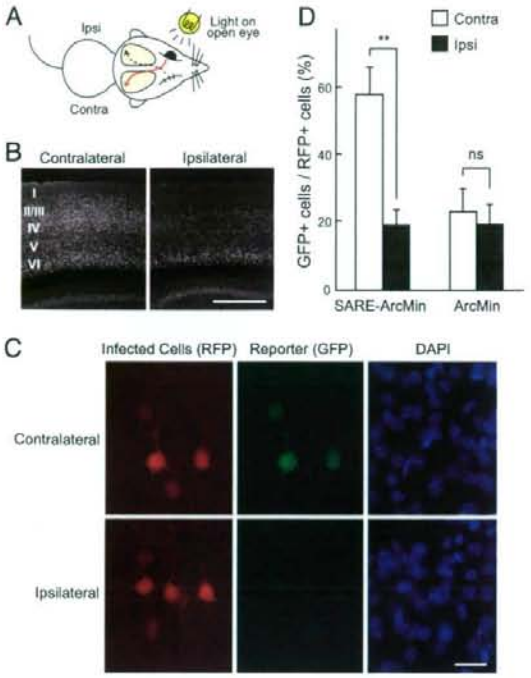

Fig. 6. Visualization of activated neurons in vivo by the SARE viral vector. (*A*) Manipulation of neuronal activity in the mouse visual cortex. The virus-infected mice were sensory-deprived on 1 eye by suture, dark reared for 2–3 days, and exposed to light on the intact open eye. (*B*) Endogenous Arc/Arg-3.1 immunohistochemistry showing unilateral activation of the visual cortex. (Scale bar, 0.5 mm.) (*C*) The SARE reporter virus-infected neurons in layers 2/3 of the visual cortex. Activity-reporter GFP signals were detected in the contralateral side, but not in the ipsilateral side. (Scale bar, 20 μm.) (*D*) The percentages of GFP-positive neurons over RFP-positive neurons in each hemisphere. $n = 6$ mice for SARE-ArcMin, $n = 5$ mice for ArcMin. **, $P < 0.01$; ns, not significant, compared to the ipsilateral side (paired t test).

contralateral, but not in the ipsilateral, visual cortex (Fig. 6*B*) (21). Consistent with this pattern, we found that the majority (58.3 ± 7.9%) of RFP-positive neurons were GFP-positive in the contralateral hemisphere that received visual inputs, whereas only a small number (19.3 ± 4.4%) of neurons were positive in the ipsilateral hemisphere (Fig. 6 *C* and *D*). The percentage of GFP- and RFP-double positive neurons in the contralateral hemisphere were comparable with that of Arc/Arg-3.1- and RFP-double positive neurons (57.8 ± 4.6%) (Fig. S6), consistent with coexpression of reporter GFP and endogenous Arc/Arg-3.1. The control ArcMin virus-infected neurons showed low GFP positivity (about 20%) and did not show imbalances between hemispheres (Fig. 6*D*). Taken together, these results indicate that SARE-ArcMin lentivirus could reliably mark activated neurons within a targeted neuronal circuit.

Discussion

Elucidation of a Novel Transcriptional Switch Mechanism That Links Synaptic Activity with Nuclear Transcription: Participation of 3 Major Activity-Regulated Transcription Factors CREB, SRF, and MEF2 in *Arc/Arg-3.1* Transcriptional Response. Our work indicates that the Arc7000 promoter essentially replicates the transcriptional profile of endogenous *Arc/Arg-3.1*, and that a relatively short DNA sequence of ≈100 bp provides a key regulatory mechanism for synapse-to-nucleus signaling. This novel *cis*-acting element contains binding sites for 3 major activity-dependent transcription factors, CREB, MEF2, and SRF, and drove as much reporter induction as Arc7000, in response to synaptic stimuli, when placed upstream of the *Arc/Arg-3.1* minimal promoter. Mutations in either one of the binding sequences resulted in a severe block of transcriptional activity. Thus, intact promoter occupancy by all 3 factors appears to be essential for proper transcriptional activation upon synaptic activity. Pharmacological analyses revealed that synaptic Ca^{2+} influx through NMDA-R was critical for its activation. On the basis of these findings, we named this main activity-sensor for *Arc/Arg-3.1* transcription "synaptic activity-responsive element (SARE)."

A functional CRE is present on many activity-regulated neuronal genes, and CREB has repeatedly been demonstrated to be a critical factor in establishment of long-term plasticity and long-term memory (1, 2, 20). Similarly, a critical role for SRE and SRF in activity-dependent gene expression has extensively been documented (3). Recent evidence also supports a role for MEF2 upstream of various activity-induced genes which directly regulate synaptic functions (28). However, the discovery of a physical clustering, corequirement and sufficiency of the 3 activity-dependent *cis*-regulatory elements located in such close proximity (≈100 bp) in SARE is particularly striking, suggesting that SARE may have a unique role as a coincidence detector for the 3 activity-dependent transcription factors. Intriguingly, we observed that mutations in the SRF-binding site not only abolished SARE activation, but also significantly augmented basal activity under silenced conditions, suggesting an additional role of SRF for transcriptional repression as previously reported in certain contexts in nonneuronal cells (29).

Although distally located SARE likely acts as a major element for activity-dependent *Arc/Arg-3.1* induction, other proximal elements may further cooperate with SARE. Our own data suggested that a proximal region possessed a 2- to 3-fold transcriptional activation ability, consistent with a previous study that showed the contribution of cAMP/MAPK pathways in the activity-dependent response mediated by this region (23). Furthermore, *Egr-1/3* may also mediate the late, protein-synthesis dependent phase of *Arc/Arg-3.1* induction via binding to this proximal region (24).

Mapping of Active Ensembles Within a Neuronal Circuit Can Be Achieved Using SARE-Based Reporters. Detailed analyses of *Arc/Arg-3.1* expression using in situ hybridization have revealed that *Arc/Arg-3.1* transcription is specifically triggered by task- or sensory input-related information processing in several brain areas, including the hippocampus, amygdala, and cerebral cortex (19–21). The shortness of SARE sequence allows designing a virus-based tool to monitor *Arc/Arg-3.1* expression for both in vitro and in vivo imaging. We here performed an experiment where we injected a SARE-reporter lentivirus into embryonic mouse cerebral ventricles in utero, at embryonic day 15 (E15). Prior work established that only the neuronal progenitor cells, present in the ventricular zone (VZ) on the day of injection, are susceptible to viral infection. VZ cells at E15 are destined to become layers 2/3 neurons in the cerebral cortex (27). More than 20 days after lentiviral infection, when the viral integration into the genome has been completed, we quantified the extent of visual activation in layers 2/3 of the primary visual cortex. The virus-injected mice were first dark reared for more than 24 h, which suppressed the reporter expression to the baseline. Following light exposure to 1 eye for 2–3 h, we found that ≈60% of the infected neurons of layers 2/3 were GFP reporter positive, which is consistent with prior functional mapping/electrophysiological reports (30), indicating that the majority of neurons indeed experienced an intense period of input activity, presumably emanating from layer 4 neurons.

Taken together, our work demonstrates that a SARE reporter, both in vitro and in vivo, can readily trace and mark an ensemble of cells that have experienced intense activity in the recent past. Further works are ongoing to design novel pharmacogenetical and

optogenetical experiments using SARE as a driver for conditional Cre recombinase such as hormone-activated CreER^{T2} (31), or for channelrhodopsins/halorhodopsins (32), with a view to manipulating specific gene expression or neuronal excitability in an activated ensemble of neurons within a circuit.

Materials and Methods

Plasmids, Lentiviruses, and Reagents. Reporter plasmid construction, lentiviral vector preparation, and reagents are described in *SI Materials and Methods*.

Neuronal Culture Preparation, Luciferase Reporter Assay, Fluorescent Protein Reporter Assay, and Luciferase Live-Cell Imaging. Dissociated neuronal culture was prepared from rat embryonic neocortex or postnatal hippocampus as described previously (10, 33). For luciferase assay, cortical neurons were transfected with a firefly luciferase reporter plasmid and an internal control Renilla luciferase plasmid by electroporation. At 10 days in vitro (div), the cells were silenced with TTX for 12–24 h, stimulated with a 4-aminopyridine/bicuculline (4AP/BIC) mixture for 4 h, and lysed. Luciferase activities were measured using the Dual Luciferase assay system (Promega).

For lentiviral reporter assay, hippocampal neurons cultured on coverslips were infected with lentiviruses at 2 div. At 15 div, neurons were silenced with TTX and then stimulated with BDNF (a gift from Dainippon Sumitomo Pharma, Osaka, Japan). For luciferase live-cell imaging, hippocampal neurons transfected with the Arc7000-ELuc-PEST plasmid were imaged in a culture medium containing d-luciferin (0.5 mM, Toyobo) at 14–18 div.

All animal experiments were carried out in accordance with regulations and guidelines of the University of Tokyo and approved by the institutional review committee of University of Tokyo Graduate School of Medicine. More information is provided in *SI Materials and Methods*.

Electrophoretic Mobility Shift Assay (EMSA) and Chromatin Immunoprecipitation (ChIP) Assay. Preparation of brain nuclear extracts is described in *SI Materials and Methods*. The nuclear extracts (2 μg for CREB and SRF and 10 μg for MEF2) were reacted with ^{32}P-radiolabeled DNA probes. Excess amounts (300-fold) of unlabeled probes were added to the reaction mixture for competition assay. Protein-DNA complexes were separated on acrylamide gels and analyzed by an image analyzer (BAS2500, Fujifilm).

Chromatin immunoprecipitation was performed using a kit (Active Motif) essentially following the manufacture's instructions. Precipitates were analyzed using a real-time PCR system (Roche) and gel electrophoresis. More information is provided in *SI Materials and Methods*.

Viral Reporter Assay in Mouse Visual Cortex. Production of lentivirus-infected mice is described in *SI Materials and Methods*. After 3–4 postnatal weeks, the infected mice were subjected to monocular deprivation, dark reared for 1–3 days, exposed to a bright environment for 2–3 h, and then perfused for immunohistochemical analysis. For quantification, infected neurons were identified by native RFP signals and TSA-enhanced GFP-immunoreactive cells were counted. More information is provided in *SI Materials and Methods*.

Statistical Analysis. Statistical analyses were performed using Prism 4.0 (Graphpad Software). Student's paired *t* test and 1-way analysis of variance with repeated measures (ANOVA) followed by Tukey's post hoc test were used for comparisons between 2 groups and 3 groups, respectively. Mann-Whitney *U* test was used for fluorescence reporter assay. All data are shown as mean ± standard error of the mean (SEM), unless otherwise stated.

ACKNOWLEDGMENTS. We thank H. Miyoshi and D. Trono for the lentiviral vector and packaging constructs, respectively; Dainippon Sumitomo Pharma (Osaka, Japan) for supply of BDNF through the courtesy of C. Nakayama and T. Ishiyama; C. Tohyama, M. Kakeyama, and W. Yoshioka (University of Tokyo) for technical help in qPCR analyses; T. Furukawa (Olympus-Japan) for assistance in building a live luciferase imaging microscope. We also thank all members of the Bito laboratory for support and discussion. We are particularly indebted to K. Saiki, Y. Kondo, and T. Kinbara for assistance. This work was supported in part by grants-in-aid from the Ministry of Education, Culture, Sports, Science and Technology (to H.O., M.O., S.T.-K., and H.B.) and from the Ministry of Health, Labor and Welfare (to H.O. and H.B.), by 21st century Center of excellence (COE) and Global COE programs (to H.B.), by a grant from National Institute of Mental Health (MH053608 to P.F.W.), and by awards from the Human Frontier Science Program Organization (career development award to H.O. and a program grant to H.B.), from the Takeda Foundation, from the Toray Science Foundation, and from the Yamada Science Foundation (to H.B.).

1. Kandel ER (2001) The molecular biology of memory storage: a dialogue between genes and synapses. *Science* 294:1030–1038.
2. Bourtchuladze R, *et al.* (1994) Deficient long-term memory in mice with a targeted mutation of the cAMP-responsive element-binding protein. *Cell* 79:59–68.
3. Ramanan N, *et al.* (2005) SRF mediates activity-induced gene expression and synaptic plasticity but not neuronal viability. *Nat Neurosci* 8:759–767.
4. Tokuyama W, Okuno H, Hashimoto T, Xin Li Y, Miyashita Y (2000) BDNF upregulation during declarative memory formation in monkey inferior temporal cortex. *Nat Neurosci* 3:1134–1142.
5. Morris RG (2006) Elements of a neurobiological theory of hippocampal function: the role of synaptic plasticity, synaptic tagging and schemas. *Eur J Neurosci* 23:2829–2846.
6. Pham TA, Impey S, Storm DR, Stryker MP (1999) CRE-mediated gene transcription in neocortical neuronal plasticity during the developmental critical period. *Neuron* 22:63–72.
7. Bliss TV, Collingridge GL (1993) A synaptic model of memory: long-term potentiation in the hippocampus. *Nature* 361:31–39.
8. Bito H, Deisseroth K, Tsien RW (1997) Ca^{2+}-dependent regulation in neuronal gene expression. *Curr Opin Neurobiol* 7:419–429.
9. West AE, Griffith EC, Greenberg ME (2002) Regulation of transcription factors by neuronal activity. *Nat Rev Neurosci* 3:921–931.
10. Bito H, Deisseroth K, Tsien RW (1996) CREB phosphorylation and dephosphorylation: a $Ca^{(2+)}$- and stimulus duration-dependent switch for hippocampal gene expression. *Cell* 87:1203–1214.
11. Robertson LM, *et al.* (1995) Regulation of c-fos expression in transgenic mice requires multiple interdependent transcription control elements. *Neuron* 14:241–252.
12. Flavell SW, Greenberg ME (2008) Signaling mechanisms linking neuronal activity to gene expression and plasticity of the nervous system. *Annu Rev Neurosci* 31:563–590.
13. Link W, *et al.* (1995) Somatodendritic expression of an immediate early gene is regulated by synaptic activity. *Proc Natl Acad Sci USA* 92:5734–5738.
14. Lyford GL, *et al.* (1995) Arc, a growth factor and activity-regulated gene, encodes a novel cytoskeleton-associated protein that is enriched in neuronal dendrites. *Neuron* 14:433–445.
15. Plath N, *et al.* (2006) Arc/Arg3.1 is essential for the consolidation of synaptic plasticity and memories. *Neuron* 52:437–444.
16. Shepherd JD, *et al.* (2006) Arc/Arg3.1 mediates homeostatic synaptic scaling of AMPA receptors. *Neuron* 52:475–484.
17. Chowdhury S, *et al.* (2006) Arc/Arg3.1 interacts with the endocytic machinery to regulate AMPA receptor trafficking. *Neuron* 52:445–459.
18. Guzowski JF, *et al.* (2000) Inhibition of activity-dependent arc protein expression in the rat hippocampus impairs the maintenance of long-term potentiation and the consolidation of long-term memory. *J Neurosci* 20:3993–4001.
19. Guzowski JF, McNaughton BL, Barnes CA, Worley PF (1999) Environment-specific expression of the immediate-early gene Arc in hippocampal neuronal ensembles. *Nat Neurosci* 2:1120–1124.
20. Han JH, *et al.* (2007) Neuronal competition and selection during memory formation. *Science* 316:457–460.
21. Tagawa Y, Kanold PO, Majdan M, Shatz CJ (2005) Multiple periods of functional ocular dominance plasticity in mouse visual cortex. *Nat Neurosci* 8:380–388.
22. Ramirez-Amaya V, *et al.* (2005) Spatial exploration-induced Arc mRNA and protein expression: evidence for selective, network-specific reactivation. *J Neurosci* 25:1761–1768.
23. Waltereit R, *et al.* (2001) Arg3.1/Arc mRNA induction by Ca^{2+} and cAMP requires protein kinase A and mitogen-activated protein kinase/extracellular regulated kinase activation. *J Neurosci* 21:5484–5493.
24. Li L, Carter J, Gao X, Whitehead J, Tourtellotte WG (2005) The neuroplasticity-associated arc gene is a direct transcriptional target of early growth response (Egr) transcription factors. *Mol Cell Biol* 25:10286–10300.
25. Wallace CS, Lyford GL, Worley PF, Steward O (1998) Differential intracellular sorting of immediate early gene mRNAs depends on signals in the mRNA sequence. *J Neurosci* 18:26–35.
26. Thomas GM, Huganir RL (2004) MAPK cascade signalling and synaptic plasticity. *Nat Rev Neurosci* 5:173–183.
27. Hashimoto M, Mikoshiba K (2004) Neuronal birthdate-specific gene transfer with adenoviral vectors. *J Neurosci* 24:286–296.
28. Flavell SW, *et al.* (2006) Activity-dependent regulation of MEF2 transcription factors suppresses excitatory synapse number. *Science* 311:1008–1012.
29. Davis FJ, Gupta M, Camoretti-Mercado B, Schwartz RJ, Gupta MP (2003) Calcium/calmodulin-dependent protein kinase activates serum response factor transcription activity by its dissociation from histone deacetylase, HDAC4. Implications in cardiac muscle gene regulation during hypertrophy. *J Biol Chem* 278:20047–20058.
30. Bear MF, Kleinschmidt A, Gu QA, Singer W (1990) Disruption of experience-dependent synaptic modifications in striate cortex by infusion of an NMDA receptor antagonist. *J Neurosci* 10:909–925.
31. Feil R, Wagner J, Metzger D, Chambon P (1997) Regulation of Cre recombinase activity by mutated estrogen receptor ligand-binding domains. *Biochem Biophys Res Commun* 237:752–757.
32. Zhang F, Aravanis AM, Adamantidis A, de Lecea L, Deisseroth K (2007) Circuit-breakers: optical technologies for probing neural signals and systems. *Nat Rev Neurosci* 8:577–581.
33. Takemoto-Kimura S, *et al.* (2007) Regulation of dendritogenesis via a lipid-raft-associated Ca^{2+}/calmodulin-dependent protein kinase CLICK-III/CaMKIgamma. *Neuron* 54:755–770.

Supporting Information

Kawashima *et al.* 10.1073/pnas.0806518106

SI Materials and Methods

Plasmid Construction. Genomic fragments containing 5′ UTR and the regulatory region of the *Arc/Arg-3.1* gene were amplified by PCR from a 129/Sv strain mouse bacterial artificial chromosome (BAC, clone24533, Incyte Genomics), and subcloned into the XhoI/BglII sites of a firefly luciferase reporter plasmid pGL4.11[luc2P] (Promega). Detailed genomic positions of each genomic fragment were as follows: −1000 (≈ −996 to +198), −2000 (−1996 to +198), −3000 (−2996 to +198), −4000 (−3996 to +198), −5000 (−4996 to +198), and −7000 (−7065 to +198), where +1 denotes the transcription initiation site. The resulting reporter plasmid harboring the −1000 fragment was designated as pGL4.11-Arc1000-luc2P, and a similar nomenclature was used with other reporter plasmids. The pGL4.11-Arc7000-del no. 1-luc2P and pGL4.11-Arc7000-del no. 2-luc2P plasmids were created by excising fragments between an SwaI site (−6668) and a XhoI site (−7065), and a Tth111I site (−6894) and the XhoI site, of the pGL4.11-Arc7000-luc2P, respectively. To directly examine the enhancer activity of SARE (Figs. 3 and 4), two distinct TATA-containing minimum promoter plasmids, pGL4.11-ArcMin-luc2P and pGL4.11-minCMV-luc2P, were constructed by inserting the ArcMin fragment, which contains a short upstream sequence and 5′ UTR (−222 to +198) of the *Arc* gene, and the minCMV fragment (1), which corresponds to the minimum sequence (−51 to +6) of the CMV IE promoter, into the BglII/HindIII sites of pGL4.11[luc2P], respectively. The SARE fragment (−6793 to −6690) was further subcloned into the XhoI/BglII sites of pGL4.11-ArcMin-luc2P and pGL4.11-minCMV-luc2P, thus creating pGL4.11-SARE-ArcMin-luc2P and pGL4.11-SARE-minCMV-luc2P, respectively. For mutation analyses of SARE (Fig. 4), point mutations were introduced into pGL4.11-Arc7000-luc2P and pGL4.11-SARE-minCMV-luc2P by PCR-based mutagenesis using the following primers: box *A*, 5′-ctgagccatgacgcagcaaggcgcttacccaagcaggaaggctctgtgcgctg-3′; box *B*, 5′-gaggctgagaatagctgagccatgttcagcaaggagcttccccacgcag-3′; box *C*, 5′-ccggcaccataaaaggagagaggcggagcatactgagccatgacg-cagcaag-3; and box *D*, 5′-gcagcctgcctgcttccggcacaatacaagtagagaggctgagaatagctga-3′; where bold and italic characters denote mutated sites.

To create d2EGFP reporter plasmids, a destabilized EGFP (d2EGFP) cDNA fragment was excised from pTAL-d2EGFP (Clontech) and replaced with the luc2P fragment of pGL4.11[luc2P]. The resulting plasmid was designated as pGL4.11-d2EGFP, and Arc4000, Arc7000, ArcMin, and SARE-ArcMin fragments described above were further subcloned into the XhoI/BglII sites of the pGL4.11-d2EGFP. Arc7000-ELuc-PEST (Emerald Luc-PEST) plasmid was created by inserting the Arc7000 fragment into the XhoI/BglII sites of pELuc(PEST)-test (Toyobo).

Genomic positions noted above were numbered according to the NCBI mouse genome database (Build 37.1). All plasmid constructs were verified by sequencing.

Reagents. Tetrodotoxin (TTX) (Wako), 4-aminopyridine (4-AP) (Sigma), water-soluble bicuculline (BIC) (Tocris), strychnine (Sigma), glycine (Nacalai Tesque), water-soluble 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX) (Tocris), and D-(-)-2-amino-5-phosphonopentanoic acid (D-AP5) (Tocris) were dissolved in water, and KN-92 (Merck), KN-93 (Merck), U0124 (Merck), and U0126 (Merck) were dissolved in dimethyl sulfoxide (DMSO) (Sigma) to make 500-1000 x stock solutions. All other chemicals were purchased from Nacalai, unless otherwise stated.

Sequence Homology Analysis and Transcription Factor Binding Site Prediction. Genomic sequences surrounding the *Arc/Arg-3.1* gene of various species were obtained from the NCBI genome database (Gene ID: mouse, 11838; rat, 54323; cow, 519403; macaque, 702690; human, 23237; opossum, 12632237; platypus, 149410163). The cavy sequence was obtained from genomic sequence AAKN02023196. For pairwise comparison of the mouse and human *Arc/Arg-3.1* promoters, genomic regions 10 kb upstream of the transcription initiation site were aligned using the mVISTA/AVID alignment program (2, 3). Local sequence comparison around SARE of multiple species was done using the AlignX (ClustalW) program in the VectorNTI Suite 9 software (Invitrogen). The dendrogram was drawn with the Njprot program. Putative transcription factor binding sites were identified with the help of the MatInspector program (Genomatix Software).

Immunocytochemistry and Western Blotting. For immunostaining of endogenous Arc, cortical neurons were stimulated as for the luciferase assay. After stimulation, cells were fixed and immunostained essentially as described previously (4). The primary antibodies used were a mouse anti-luciferase mAb (C-21, Santa Cruz) (1:200) and a rabbit anti-Arc pAb (OP-1, 1:2000) that was raised against bacterially expressed recombinant GST-fused full-length Arc (Okuno *et al.*, unpublished data). The secondary antibodies were AlexaFluor594- or AlexaFlour488-conjugated anti-rabbit and anti-mouse antibodies (1:1000) (Invitrogen).

Western blotting was performed essentially as described previously (4). Cells were lysed in a sample buffer and separated on a standard discontinuous acrylamide gel (10%) or a gradient acrylamide gel (4−12%) (Invitrogen). The primary antibodies used were a mouse anti-β-tubulin mAb (TUB2.1) (Sigma) and the rabbit anti-Arc pAb. The secondary antibodies were HRP-conjugated anti-mouse and anti-rabbit antibodies (GE Healthcare). Chemiluminescence was detected using ECL-Plus reagent (GE Healthcare) and a LAS4000mini image analyzer (Fujifilm).

Luciferase Reporter Assay. Cortical neurons were transfected with a firefly luciferase reporter plasmid and an internal control Renilla luciferase plasmid pGL4.74 [hRluc/TK] (Promega) at a ratio of 4:1 (0.60 μg and 0.15 μg for each well) by electroporation using a Nucleofector (Amaxa) and plated at a density of 1.25×10^6 cells in a 12-well dish. At 10 days in vitro (div), the cells were pretreated with a medium containing 2 μM TTX. After 12–24 h (i.e., 11 div), cells were stimulated with a medium containing 100 μM 4-aminopyridine (4AP), 30 μM bicuculline (BIC), 100 μM glycine, and 1 μM strychnine. For analyses of signaling pathways (Fig. S5), kinase inhibitors and channel blockers were added to the medium 30 min before the stimulation. Cells were lysed 4 h after stimulation, and luciferase activities of both firefly and renilla were measured using the Dual Luciferase assay system (Promega) and a Fluoroskan Ascent FL luminometer (Thermo Fisher Scientific), according to the manufacturers' protocols. The firefly luciferase activities were normalized with the intra-sample renilla luciferase activities, corrected for transfected plasmid copy numbers, and further normalized to an arbitrary standard value (Arc1000 in TTX) for appropriate comparison of different reporter plasmids. Assays using different batches of neuronal culture (i.e., primary culture prepared

from different animals on different days) were regarded as independent experiments.

Fluorescent Protein Reporter Assay. Hippocampal neurons cultured on coverslips were cotransfected with a d2EGFP reporter plasmid and a marker plasmid pTagRFP-C (Evrogen) at a ratio of 5:1 using Lipofectamine 2000 (Invitrogen) at 7 div. Neurons were pretreated with a medium containing TTX at 15 div, and stimulated with a medium containing 4AP/BIC/glycine/strychnine at 16 div as for the luciferase assay. After 4 h, cells were fixed and mounted on slides. The average GFP intensity in the soma of all RFP-positive neurons was quantified.

To evaluate the reporter activity of SARE-ArcMin lentivirus, hippocampal neurons (2 div) cultured on coverslips were infected with SARE-ArcMin lentivirus at a MOI of about 1. Neurons were pretreated with 2 μM TTX at 15 div and stimulated at 16 div with 50 ng/ml BDNF (generously provided by Dainippon Sumitomo Pharma through the courtesy of Dr. Chikao Nakayama). After 5–6 h, cells were fixed and mounted on slides.

Nuclear Extract Preparation from Stimulated Cerebral Cortex. Eight-week-old male Wistar rats were given electrical shocks (100 V, 5 msec, 100 Hz for 5 sec) through ear clips connected to an electrical stimulator (SIU-102, Warner Instruments) controlled by a pulse generator (Master-8, A.M.P.I.) and killed after 30–40 min. The brains were rapidly removed, and the neocortex was dissected. Nuclear extracts were isolated essentially as described previously (5), except that protease and phosphatase inhibitor cocktails, Complete and PhosStop (Roche Diagnostics), were added in extraction buffers. The resulting protein extracts were desalted using Amicon Ultra-15 filter devices (Millipore) in a buffer [20 mM Hepes (pH 7.6), 150 mM KCl, 1 mM EDTA, 1.5 mM $MgCl_2$, 1 mM DTT, 10% glycerol] and stored at −80 °C. Protein concentrations were measured by the Quick Start Bradford Protein Assay (Bio-Rad).

Electrophoretic Mobility Shift Assay. For protein-DNA binding reaction, 2 μg (for CREB and SRF) or 10 μg (for MEF2) of nuclear extracts were reacted with a ^{32}P-radiolabeled DNA probe (see below) in a 20-μL reaction mixture [25 mM Hepes (pH 7.6), 60 mM KCl, 1 mM EDTA, 5 mM $MgCl_2$, 1 μg poly(dI-dC), 0.1 μg poly L-lysine, 10% glycerol]. For competition assay, the mixture was reacted in the presence of excess amounts (300-fold) of unlabeled probes. For supershift assay, anti-CREB mAb (48H2, Cell Signaling), anti-MEF2D mAb (clone 9, BD Biosciences), anti-SRF pAb (H-300x, Santa Cruz), and control anti-Myc mAb (9B11, Cell Signaling) were added to the reaction mixture. Essentially the same results were obtained using different anti-CREB (E306, Epitomics), anti-MEF2 (H-300x, Santa Cruz), and anti-SRF (ab53130, Abcam) antibodies. The mixtures were incubated for 15 min at room temperature, and the resulting protein-DNA complexes were separated by electrophoresis on a 5% nondenaturing acrylamide gel in 0.25 × TBE buffer. The gels were dried, exposed overnight to a BAS-SR2040 imaging plate (Fujifilm), and analyzed by the BAS2500 image analyzer (Fujifilm).

The DNA probes used were the followings:
box *B*, 5′-ACATACTTGCTGCGTCATGGCTCAGAGATG-3′;
mutB, 5′-ACATACTTGCTG*AAA*CATGGCTCAGAGATG-3′;
box *C*, 5′-ATTGGCTCAGCTATTCTCAGCCTCTCTACG-3′;
mutC, 5′-ATTGGCTCAG*A*TATGCTCCGCCTCTCTACG-3′;
box *D*, 5′-AATCCTCTCTCCTTTTATGGTGCCGTCGGC-3′;
mutD, 5′-AATCCTCTCT*A*CTTGTAT*T*GTGCCGTCGGC-3′;
where italic characters denote mutated nucleotides.

The analysis of MEF2-box*C* complex required relatively large amounts of nuclear extracts, which inevitably brought higher nonspecific signals. To reduce these nonspecific signals, mutated competitors were added to the reaction in MEF2-antibody supershift assays.

Chromatin Immunoprecipitation (ChIP) Assay. Cortical neurons were cultured on a 10-cm dish at a density of 1.25×10^7 cells/dish. At 13–15 div, the neurons were treated with 1 μM TTX overnight and stimulated with the 4AP/BIC mixture for 30 min. The chromatin immunoprecipitation assay was performed using the ChIP-IT Express kit (Active Motif) according to the manufacturer's instructions with some modifications. Briefly, the neurons were fixed with 1% PFA in a culture medium for 10 min at room temperature, homogenized, and centrifuged. The pellets were resuspended in a shearing buffer and sheared using a cup-horn sonicator (XL2020/431A, Misonix Inc.) with ice-cold water perfusion. The sonication-sheared lysates were cleared by centrifugation and used for chromatin immunoprecipitation. The quality of shearing was monitored by DNA electrophoresis for each shearing batch. We only used the sheared chromatin lysates in which the length of DNA fragments were ranging from 200 to 1000 bp with an average of ≈500 bp. The sheared chromatin lysates (50–60 μg proteins) were reacted with a specific antibody (1–2 μg) overnight at 4 °C and immunoprecipitated with protein G magnetic beads. The precipitated DNA-chromatin complex was washed, eluted, incubated for 6 h at 65 °C in a reverse cross-link buffer, and digested with proteinase K for 1 h at 37 °C. The resultant DNA was purified using a purification column. Antibodies used for immunoprecipitation were the following: rabbit polyclonal anti-CREB and control IgG (ChIPAb+ CREB kit, Millipore), rabbit polyclonal anti-SRF (sc-335X, Santa Cruz), mouse monoclonal anti-MEF2D (BD Biosciences), and control IgG (ChIP-IT control kit, Active Motif).

The precipitated DNA fractions were analyzed using a real-time PCR system (Light Cycler 1.5, Roche) with SYBR Green I kits (TaKaRa and Roche). The primers used were: SARE primers set 1 (−6778 to −6649, for CREB and MEF2), 5′-ggctggctctgggaggtattta-3′ and 5′-cccccagagatgagagttcaga-3′; SARE primers set 2 (−6815 to −6649, for SRF), 5′-ctccttttatggtgccggaag-3′ and 5′-cccccagagatgagagttcaga-3′; negative control primers (+2904 to +3149, in *Arc/Arg-3.1* 3′-UTR locus), 5′-agaaccttgcaggagcctta-3′ and 5′-atggaggaacctcaacatgg-3′; somatostatin primers (−142 to +37), 5′- tgcgaggctaatggtgcg-taaa-3′ and 5′-cttgagcgcggtgggtcagt-3′; Nur77 primers (−547 to −364), 5′-gtccgggactgactgggaaa-3′ and 5′-gtcccggggtccgaaataac-3′; β-actin primers (+14 to +232), 5′-cacccgcgagtacaaccttctt-3′ and 5′-caccctaggcggaaagttaagc-3′. The PCR products were visualized by electrophoresis in 4% agarose gels (NuSieve 3:1, Lonza). The identity of each product was confirmed by restriction enzyme digestion.

Reanalysis of Recently Published SRF ChIP-Seq Raw Data Confirmed the Potential of SRF Binding to the SARE Sequence (6).

Live Imaging. For luciferase live-cell imaging, hippocampal neurons plated on glass-bottom dishes were cotransfected with the Arc7000-ELuc-PEST plasmid and a marker plasmid pEGFP-C1 (Clontech) at 7 div as described before. At 14–18 div, neurons were incubated in a culture medium containing d-luciferin (0.5 mM, Toyobo) at 37 °C in a stage CO_2 chamber (Tokai Hit). Transfected neurons were identified by GFP signals, and luminescent images were acquired with 5-min exposure every 60 min. After collecting baseline images, neurons were stimulated by field-electrical pulses (3 ms, 50 mA, 100 pulses at 100 Hz, 10 times with 30-sec intervals) from 2 parallel platinum electrodes using an electrical stimulator (SIU-102, Warner Instruments) controlled by a pulse generator (Master-8, A.M.P.I.) (7). The integrated luminescence in the soma (L) of individual neurons was normalized to the basal luminescence (L_0) before the stimulation.

Lentivirus Vector Construction and Viral Production. A lentiviral SARE-ArcMin reporter plasmid (CS-SARE-ArcMin-d2EGFP)

was created by introducing 2 independent expression cassettes, SARE-ArcMin-d2EGFP (GFP, activity reporter) and PGK-TurboFP635 (RFP, infection marker), into a lentiviral plasmid CS-CA-MCS (8). These 2 cassettes were placed in the opposite directions, and an intermediate insulator sequence was placed to prevent interactions between the cassettes (9). To insert multiple fragments, the chicken actin-promoter fragment of the lentiviral plasmid CS-CA-MCS (generously provided by Dr. Hiroyuki Miyoshi, RIKEN-BRC, Japan) was excised by PstI/XhoI digestion and replaced with a synthetic oligonucleotide 5′-tgaattcttcgaactgcagggcgcgccgctagcaccggtctcgaggttaacggatccg-3′, which contains (EcoRI-BstBI-PstI-AscI-NheI-AgeI-XhoI-HpaI-BamHI) unique restriction enzyme sites. A synthetic oligonucleotide containing the enhancer blocking FII sequence from the chicken β-globin insulator (10) 5′-cgcgccgccccgggcg-gctcgctgctgccccctagcgggggagg-gacgtaattacatccctgggggctttgggggggg-3′ was inserted into the AscI/NheI sites. The mouse phosphoglycerate kinase (PGK) promoter and TurboFP635 cDNA (Evrogen) were amplified by PCR and subcloned into the XhoI/HpaI sites and the HpaI/BamHI sites, respectively. The destabilized EGFP (d2EGFP) cDNA with SV40 polyA sequence were amplified from pTAL-d2EGFP (Clontech) and subcloned into the BstBI/EcoRI sites. The SARE-ArcMin fragment was subcloned into the AscI/BstBI sites, thus creating CS-SARE-ArcMin-d2EGFP. A control lentiviral plasmid, CS-ArcMin-d2EGFP, was created by the same process as described above, except that the ArcMin fragment was inserted instead of the SARE-ArcMin.

To generate self-inactivating lentiviruses, CS-SARE-ArcMin-d2EGFP or CS-ArcMin-d2EGFP were cotransfected with ps-PAX2 (Addgene plasmid 12260) and pMD2.G (Addgene plasmid 12259) (courtesy of Dr. Didier Trono, EPFL, Switzerland) into HEK293T cells using the calcium phosphate method. Culture supernatants were collected 48–72 h after transfection, filtered with 0.45 μm cellulose acetate membrane filters (Asahi Techno Glass), and ultracentrifuged at 80,000 $\times$ g for 2 h. The resulting pellets were dissolved in PBS to obtain concentrated virus solutions. Virus titer was determined using HEK293T cells by counting RFP-positive cells.

Live Imaging of Lentivirus-Infected Neurons. Hippocampal neurons cultured on glass-bottom dishes were infected at 7–8 div with the SARE-ArcMin reporter or the ArcMin control lentiviruses at a nominal MOI of 2–12. After 1–2 weeks, neurons were incubated in a stage CO_2 chamber (Tokai Hit) and fluorescent images were acquired every 20 min. Neurons were stimulated by field electrical pulses (3 ms, 100 mA, 100 pulses at 100 Hz, 9 times with 1-min intervals). The GFP and RFP fluorescence in the soma of individual neurons was monitored before and after the stimulation. GFP fluorescence was divided by RFP fluorescence (F) and further normalized to the basal fluorescence (F_0) before the stimulation.

Viral Reporter Assay in Mouse Visual Cortex. Injection of lentiviruses into mouse embryos in the uterus was performed similarly to the procedure previously reported for in vivo electroporation (11, 12). Virus solutions (1–2 μL) were injected into the lateral ventricle of E15 embryos in the uterus. The pups were born naturally and grew without any apparent abnormality.

After 3–4 postnatal weeks, the pups were subjected to monocular deprivation by suture and glue. The pups were subsequently dark reared for 1–3 days, and exposed to a bright environment for 2–3 h. The pups were then perfused with 4% paraformaldehyde/phosphate buffer. The brains were removed, postfixed, cryoprotected, and frozen in M1 Embedding Matrix (Shandon, Thermo Scientific). Coronal sections (30 μm) of the visual cortex were cut using a cryostat. The primary visual cortex was defined according to a brain atlas (13).

Immunohistochemistry was performed as previously described (5, 14, 15), except that the TSA system was used for detection. The sections were treated with a rabbit anti-GFP pAb (A-11122, 1:3000) (Invitrogen). After washing, the sections were treated with an HRP-conjugated anti-rabbit goat pAb (1:500, Jackson Immunoresearch), and the signals were developed using a TSA Plus FITC System (PerkinElmer) according to the manufacturer's protocols. A subset of sections were immunostained with a rat anti-GFP mAb (GF090R, Nacalai) and an anti-rat AlexaFlour488-conjugated antibody.

Infected neurons in the primary visual cortex were identified by native RFP signals. GFP immunoreactivity-positive cells were counted, and the ratio of the number of RPF/GFP double-positive cells to the RFP-positive cells was calculated for each hemisphere of individual brains. The binocular areas were excluded from the regions of interest. Endogenous Arc/Arg-3.1 induction in the visual cortex was confirmed by staining sections with a rabbit anti-Arc pAb (OP-2, 1:2000, Okuno *et al.*, unpublished data).

Image Acquisition. The fluorescence protein reporter assay used a 40× objective (NA 1.3, Oil) with an EM-CCD camera (iXon, Andor Technology) attached to an inverted microscope (IX81, Olympus, Tokyo, Japan). Luciferase live-cell imaging and fluorescence live-cell imaging used a 40x objective (NA 0.95, Air) with the Andor CCD camera. Fixed fluorescent images of lentivirus-infected cultured neurons and mouse visual cortex were obtained using a 20× objective (NA 0.65, Air) with a color CCD camera (DP-70, Olympus) attached to an upright microscope (BX-51, Olympus) or a 20× objective (NA 0.5) with a confocal laser microscopy system (LSM 510META-V3.2, Carl Zeiss) attached to an inverted microscope (Axiovert 200M). Collected images were analyzed using Metamorph software (Molecular Devices) or ImageJ software (National Institutes of Health).

1. Gossen M, *et al.* (1995) Transcriptional activation by tetracyclines in mammalian cells. *Science* 268:1766–1769.
2. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I (2004) VISTA: computational tools for comparative genomics. *Nucleic Acids Res* 32:W273–279.
3. Bray N, Dubchak I, Pachter L (2003) AVID: a global alignment program. *Genome Res* 13:97–102.
4. Takemoto-Kimura S, *et al.* (2007) Regulation of dendritogenesis via a lipid-raft-associated Ca^{2+}/calmodulin-dependent protein kinase CLICK-III/CaMKIgamma. *Neuron* 54:755–770.
5. Okuno H, Saffen DW, Miyashita Y (1995) Subdivision-specific expression of ZIF268 in the hippocampal formation of the macaque monkey. *Neuroscience* 66:829–845.
6. Valouev A, *et al.* (2008) Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. *Nat Methods* 5:829–834.
7. Furuyashiki T, Arakawa Y, Takemoto-Kimura S, Bito H, Narumiya S (2002) Multiple spatiotemporal modes of actin reorganization by NMDA receptors and voltage-gated Ca^{2+} channels. *Proc Natl Acad Sci USA* 99:14458–14463.
8. Miyoshi H, Blomer U, Takahashi M, Gage FH, Verma IM (1998) Development of a self-inactivating lentivirus vector. *J Virol* 72:8150–8157.
9. Vogel R, Amar L, Thi AD, Saillour P, Mallet J (2004) A single lentivirus vector mediates doxycycline-regulated expression of transgenes in the brain. *Hum Gene Ther* 15:157–165.
10. Bell AC, West AG, Felsenfeld G (1999) The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* 98:387–396.
11. Saito T, Nakatsuji N (2001) Efficient gene transfer into the embryonic mouse brain using in vivo electroporation. *Dev Biol* 240:237–246.
12. Tabata H, Nakajima K (2001) Efficient in utero gene transfer system to the developing mouse brain using electroporation: visualization of neuronal migration in the developing cortex. *Neuroscience* 103:865–872.
13. Franklin K, Paxinos G (1997) *The Mouse Brain in Stereotaxic Coordinates* (Academic, New York).
14. Okuno H, Kanou S, Tokuyama W, Li YX, Miyashita Y (1997) Layer-specific differential regulation of transcription factors Zif268 and Jun-D in visual cortex V1 and V2 of macaque monkeys. *Neuroscience* 81:653–666.
15. Okuno H, Miyashita Y (1996) Expression of the transcription factor Zif268 in the temporal cortex of monkeys during visual paired associate learning. *Eur J Neurosci* 8:2118–2128.

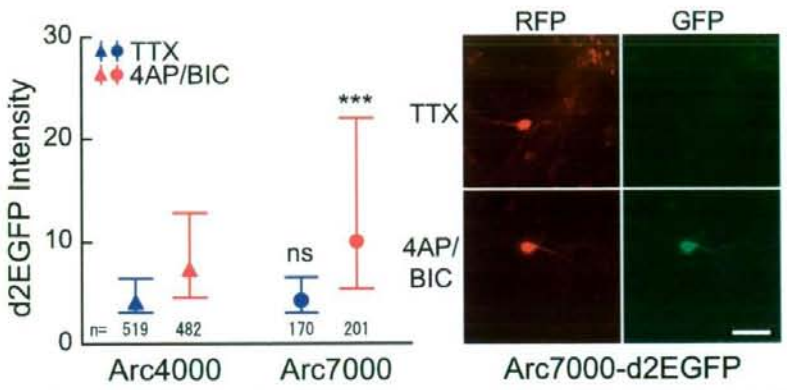

Fig. S1. Activation of Arc7000 in individual hippocampal neurons. Neurons were cotransfected with reporter (d2EGFP) and transfection marker (tagRFP) plasmids, stimulated with 4AP/BIC, and average GFP fluorescence in the soma was quantified. The triangles and circles represent median values, and bars represent quadrant points. Representative images of Arc7000 transfected neurons are shown on the *Right*. GFP reporter signals were more elevated after stimulation downstream of Arc7000 than of Arc4000. The difference between Arc7000 and Arc4000 appeared smaller than that obtained by the luciferase assays, presumably owing to higher background autofluorescence. Numbers in the *Left* panel indicate the number of neurons analyzed. ***, $P < 0.001$; ns, not significant (Mann-Whitney test) compared with the 4AP/BIC or the TTX value of Arc4000. (Scale bar, 50 μm.)

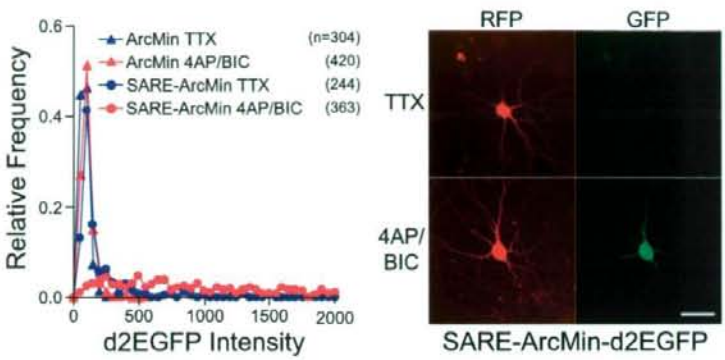

Fig. S2. Activation of SARE in individual hippocampal neurons. Neurons were cotransfected with reporter (d2EGFP) and transfection marker (tagRFP) plasmids, stimulated with 4AP/BIC and average GFP fluorescence in the soma was quantified. The frequency histogram of reporter GFP signals in individual neurons is shown. The distribution of activated GFP signals from SARE-ArcMin was dramatically shifted to the right compared to that from ArcMin alone. Representative images of SARE-ArcMin transfected neurons are shown on the *Right*. Numbers in the *Left* panel indicate the number of neurons analyzed. (Scale bar, 50 μm.)

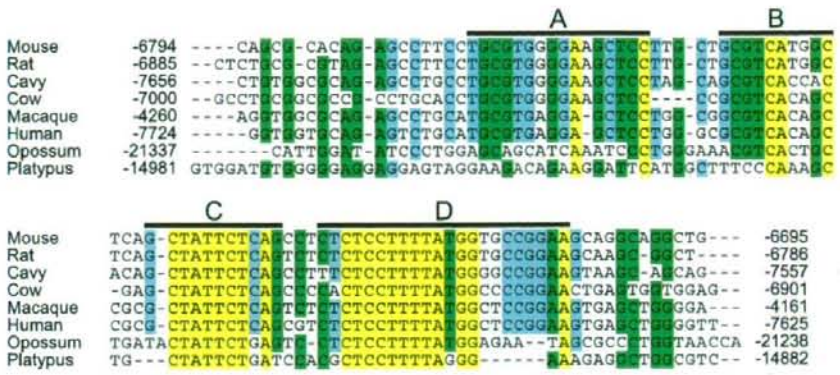

Fig. S3. Divergence of *Arc/Arg-3.1* SARE sequence across various mammalian species.

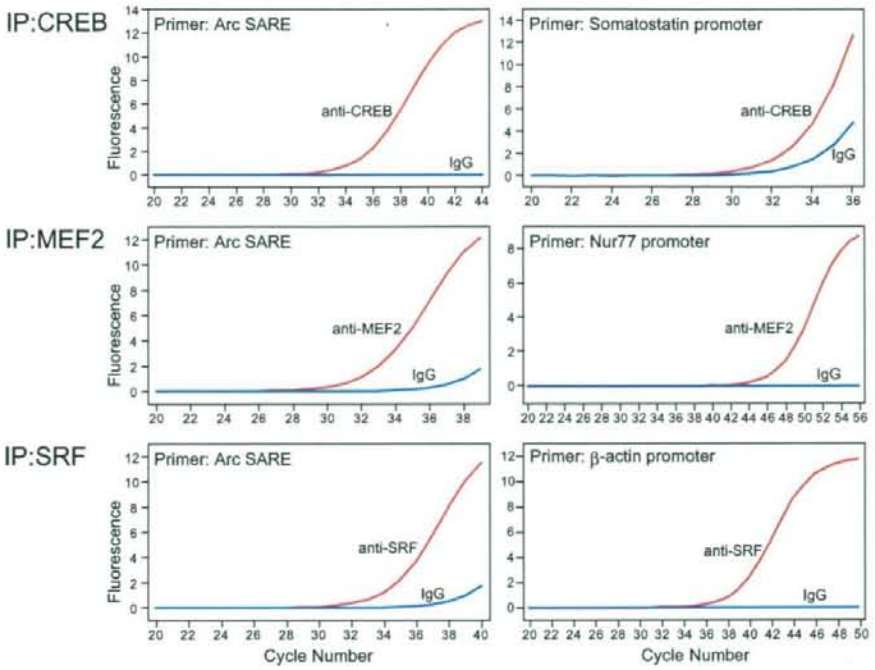

Fig. S4. Representative real-time PCR data of ChIP assays. SYBR Green I fluorescence was plotted against PCR cycle numbers. The chromatin lysates were immunoprecipitated (IP) with antibodies against 3 transcription factors. Genomic fragments in the precipitates were analyzed using specific primers indicated in each panel. The onset of fluorescence increment in the precipitates (red) always preceded that in the control IgG precipitates (blue), indicating the specific binding of CREB, MEF2, and SRF to target genomic sequences. The specificity of the PCR was confirmed by melting curve analyses.

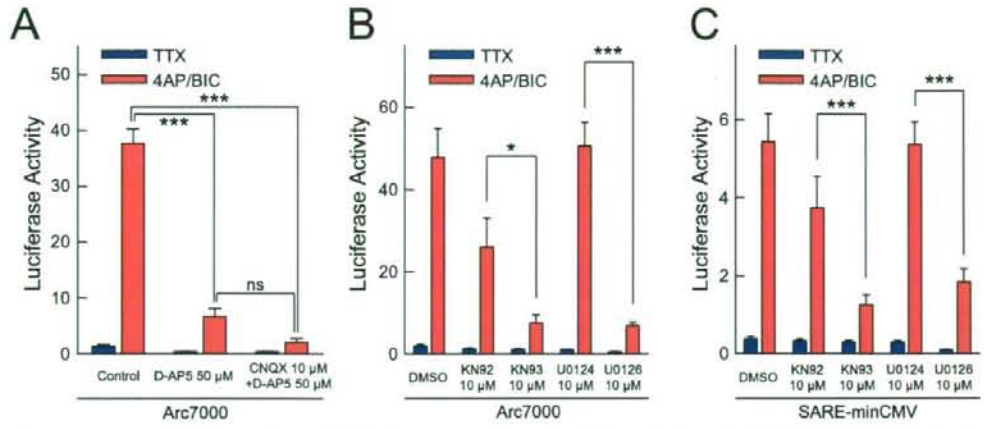

Fig. S5. Pharmacological analysis of the Arc7000- and SARE-activating signaling cascades. (*A*) The activation of Arc7000 is suppressed by ionotropic glutamate receptor antagonists. Activation of Arc7000 was almost completely suppressed when both AMPA-R and NMDA-R were blocked by CNQX and AP5. Activation of Arc7000 was also diminished by 80% when only NMDA-R was blocked. ***, $P < 0.001$; ns, not significant (1-way ANOVA with Tukey's test). (*B* and *C*) The CaMK and MAPK pathways are involved in the activation of Arc7000 via SARE. Neurons were treated with a CaMK inhibitor, KN-93, and a MEK inhibitor, U0126, shortly before and during the 4AP/BIC stimulation and subjected to the luciferase assay. The activation levels (4AP/BIC) of the Arc7000 and the SARE-minCMV activities were significantly diminished by both KN-93 and U0126 compared with their inactive analogues, KN-92 and U0124, respectively. *, $P < 0.05$; ***, $P < 0.001$ (1-way ANOVA with Tukey's test).

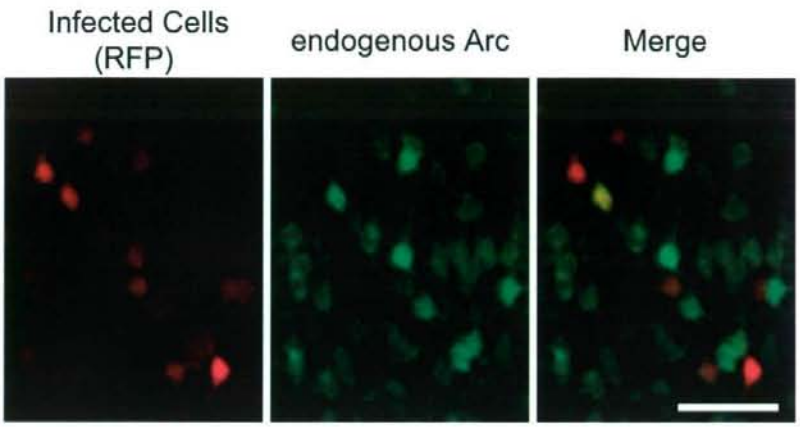

Fig. S6. Expression of endogenous *Arc/Arg-3.1* in virus-infected neurons in the visual cortex. A majority of reporter-virus-infected neurons expressed increased levels of endogenous *Arc/Arg-3.1* in the activated visual cortex. (Scale bar, 50 μm.)

A Cryptic Algal Group Unveiled: A Plastid Biosynthesis Pathway in the Oyster Parasite *Perkinsus marinus*

Motomichi Matsuzaki,[1] *Haruko Kuroiwa,† Tsuneyoshi Kuroiwa,† Kiyoshi Kita,‡ and Hisayoshi Nozaki**

*Department of Biological Sciences, Graduate School of Science, University of Tokyo, Tokyo, Japan; †Department of Life Science, College of Science, and Research Information Center for Extremophile, Rikkyo University, Tokyo, Japan; and ‡Department of Biomedical Chemistry, Graduate School of Medicine, University of Tokyo, Tokyo, Japan

Plastids are widespread in plant and algal lineages. They are also exploited by some nonphotosynthetic protists, including malarial parasites, to support their diverse modes of life. However, cryptic plastids may exist in other nonphotosynthetic protists, which could be important in studies on the diversity and evolution of plastids. The parasite *Perkinsus marinus*, which causes mass mortality in oyster farms, is a nonphotosynthetic protist that is phylogenetically related to plastid-bearing dinoflagellates and apicomplexans. In this study, we searched for *P. marinus* methylerythritol phosphate (MEP) pathway genes, responsible for de novo isoprenoid synthesis in plastids, and determined the full-length gene sequences for 6 of 7 of these genes. Phylogenetic analyses revealed that each *P. marinus* gene clusters with orthologs from plastid-bearing eukaryotes, which have MEP pathway genes with essentially the same mosaic pattern of evolutionary origin. A new analytical method called sliding-window iteration of TargetP was developed to examine the distribution of targeting preferences. This analysis revealed that the sequenced genes encode bipartite targeting peptides that are characteristic of proteins targeted to secondary plastids originating from endosymbiosis of eukaryotic algae. These results support our idea that *Perkinsus* is a cryptic algal group containing nonphotosynthetic secondary plastids. In fact, immunofluorescent microscopy indicated that 1 of the MEP pathway enzymes, 1-deoxy-D-xylulose 5-phosphate reductoisomerase, was localized to small compartments near mitochondrion, which are possibly plastids. This tiny organelle seems to contain very low quantities of DNA or may even lack DNA entirely. The MEP pathway genes are a useful tool for investigating plastid evolution in both of the photosynthetic and nonphotosynthetic eukaryotes and led us to propose the hypothesis that ancestral "chromalveolates" harbored plastids before a secondary endosymbiotic event.

Introduction

"Plastids" are intrinsic organelles in plants and algae, but gaps remain in our knowledge regarding their diversity and distribution. Plastids originally arose from endosymbiotic cyanobacteria and are now involved in processes including photosynthesis and other biochemical processes in plant cells. Among the protists, several lines of algae or plastid-bearing eukaryotes (PBEs), such as giant kelp and diverse bloom-forming algae are known. Furthermore, intracellular parasites of the phylum Apicomplexa, including the malarial parasite, have been recently shown to harbor nonphotosynthetic but essential plastids (Wilson 2005). In a very recent environmental sequencing study, a distinct group of PBEs, the picobiliphytes, were discovered (Not et al. 2007). Thus, it seems very likely that unknown PBEs still exist.

On the basis of their deduced evolutionary history, plastids can be divided into 2 classes: primary plastids with 2 bounding membranes, which are direct descendants of endosymbiotic cyanobacteria, and secondary plastids with more bounding membranes, which originate from past engulfed eukaryotic algae (Bhattacharya et al. 2004). Because secondary plastids remain "outside" with regard to membrane topology, proteins targeted to these compartments must be transported via a secretory pathway; they must contain an N-terminal bipartite targeting peptide, composed of a signal peptide (SP) to lead the polypeptide to the endoplasmic reticulum (ER), and a subsequent transit peptide (TP) to deliver the mature protein into the plastid lumen (van Dooren et al. 2001). Secondary PBEs are scattered over the protist phylogeny; however, it has been proposed that members of the so-called chromalveolate group, which consists of diatoms and other stramenopiles, haptophytes, cryptophytes, dinoflagellates, and apicomplexans, ancestrally contain secondary plastids of a single red algal origin (Cavalier-Smith 1999). Several lines of "evidence" support the "chromalveolate" hypothesis (Fast et al. 2001; Yoon et al. 2002), but critics note that the hypothesis assumes too many independent losses of plastids (Bodył 2005). Thus, a better understanding of plastid distribution in the basal chromalveolates is desirable to address these criticisms.

Perkinsus spp. are marine unicellular protists with a worldwide distribution that attack a wide range of mollusks, including clams and abalones, causing mass mortality (Villalba et al. 2004). *Perkinsus marinus* is the most notorious species of the genus because it parasitizes the eastern oyster *Crassostrea virginica* and has heavily impacted oyster fisheries and hence coastal water quality in the United States (Villalba et al. 2004). Molecular phylogenetic data have shown that this species is a basal chromalveolate derived from the ancestral dinoflagellates just after the split from apicomplexans (Cavalier-Smith and Chao 2004; Leander and Keeling 2004; Adl et al. 2005); thus, examining *P. marinus* for the existence of ancestral secondary plastids is of interest. Although electron microscopy (EM) observations have revealed no signs of plastids (Perkins and Menzel 1967; Perkins 1976, 1996; Sunila et al. 2001), 2 quite recent studies have suggested that *Perkinsus* spp. contain secondary plastids; *P. marinus* was shown to possess genes for a plant-type ferredoxin system that possibly encode plastid-targeting signals (Stelter et al. 2007) and an EM observation indicated a tiny organelle bounded by 4 membranes in *Perkinsus olseni* (=*Perkinsus*

[1] Present address: Department of Biomedical Chemistry, Graduate School of Medicine, University of Tokyo, Tokyo, Japan.

Key words: secondary endosymbiosis, protein sorting signal, chromalveolates, methylerythritol phosphate pathway, *Perkinsus marinus*.

E-mail: mzaki@m.u-tokyo.ac.jp

Mol. Biol. Evol. 25(6):1167–1179. 2008
doi:10.1093/molbev/msn064
Advance Access publication March 20, 2008

atlanticus) (Teles-Grilo et al. 2007). However, gene sequence or morphology in itself cannot prove the existence of vestigial plastids; at the least, localization data should be presented in order to confirm the presence of plastids.

Genome sequencing has revealed that biosynthesis of isoprenoid precursors is a key metabolic role of both photosynthetic (Matsuzaki et al. 2004; Derelle et al. 2006) and nonphotosynthetic (Gardner et al. 2002, 2005) plastids. Isoprenoids are a diverse and versatile group of compounds including sterols, carotenoids and other terpenes, and the side chains of quinones and chlorophylls. Isoprenoids are all derived from isopentenyl diphosphate and its isomer, both of which are synthesized by the methylerythritol phosphate (MEP) pathway in plastids but the mevalonate (MVA) pathway in the cytosol of many eukaryotes, including higher plants and animals (Rodríguez-Concepción 2004). It has been suggested that in eukaryotes the MEP pathway only exists in PBEs, with a discussion of the evolutionary origins of the genes based on the orthologs of higher plants and several bacteria, although only 5 out of 7 MEP genes were known at the time (Lange et al. 2000). Many eukaryotic genomes have been sequenced since then, and to date the MEP pathway genes have been found to be specific to PBEs and would seem to be necessary for most given that they lack the MVA pathway genes; thus, it seems that the MEP pathway is a specific feature of plastids, photosynthetic, or otherwise. Therefore, the 7 MEP pathway genes (shown in fig. 1) would be good molecular markers for the study of plastid distribution and evolutionary history; however, to the best of our knowledge, no prior studies have involved the widespread sampling of PBEs.

A preliminary database of the *P. marinus* genome at The Institute for Genomic Research (TIGR) contains partial sequences of the MEP pathway genes (the existence of ispC, ispG, and ispH has already been discussed by Stelter et al. [2007]). In this study, we attempted to clone full-length MEP pathway genes of *P. marinus* to elucidate the evolution of secondary plastids by phylogenetic analyses. A new analytical method for investigating protein sorting signals was also developed to compare signals between genes. Finally, immunofluorescent microscopy was performed to demonstrate the existence of plastids in *P. marinus*.

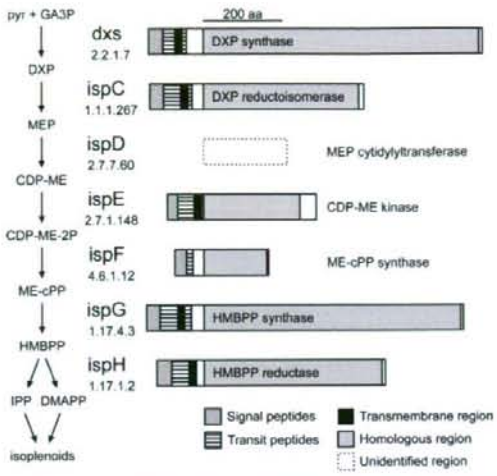

FIG. 1.—The MEP pathway and responsible genes in *Perkinsus marinus*. The flowchart on the left shows the compounds and reactions involved in the MEP pathway, with the name of the gene responsible and an Enzyme Committee number listed for each reaction. Compounds are abbreviated as follows: pyr, pyruvate; GA3P, glyceraldehyde 3-phosphate; DXP, 1-deoxy-D-xylulose 5-phosphate; MEP, 2-*C*-methyl-D-erythritol 4-phosphate; CDP-ME, 4-diphosphocytidyl methylerythritol; CDP-ME-2P, CDP-ME 2-phosphate; ME-cPP, methylerythritol 2,4-cyclodiphosphate; HMBPP, 1-hydroxy-2-methyl-2-butenyl 4-diphosphate; IPP, isopentenyl diphosphate; and DMAPP, dimethylallyl diphosphate. On the right are schematic representations of the inferred structures of MEP pathway enzymes in *P. marinus*. N-terminal extensions that were putatively composed of SP (diagonal hatch) and TP (horizontal stripe) with a transmembrane region (filled) were located next to the regions that were homologous to each enzyme's bacterial counterpart (shaded). The gene for the third step, an *ispD* ortholog, is unidentified and indicated by a box with dotted edges. The scale is indicated at the top of the figure.

Materials and Methods

Culture Conditions

Perkinsus marinus strain CRTW-3HE was obtained from the American Type Culture Collection (number 50439; ATCC, Manassas, VA) and maintained at 26 °C in ATCC Medium 1886. Discontinued products for the medium components were substituted as follows during the course of the study: Instant Ocean Sea Salt (Aquarium Systems, Mentor, OH) for artificial seawater (S1649; Sigma, St Louis, MO) and Lipid Mixture (1000×; L5146; Sigma) for Lipid Concentrate (100×; 21900-014; Gibco, Grand Island, NY).

Sequencing MEP Pathway Genes

Total RNA was extracted from cell pellets using TRIzol Reagent (Invitrogen, Carlsbad, CA), and the mRNA was enriched using the PolyATtract mRNA Isolation System III (Promega, Madison, WI). Complementary DNA was synthesized using the CapFishing Full-length cDNA Premix Kit (Seegene, Seoul, Korea) with random hexamers or the oligo dT adapter as a primer. Reverse transcriptase–polymerase chain reaction and rapid amplification of cDNA ends were performed using PrimeSTAR HS DNA polymerase (Takara Bio, Tokyo, Japan) and the primers listed in supplementary table S1 (Supplementary Material online). Polymerase chain reaction products were purified using Wizard SV Gel and PCR Clean-Up System (Promega) and then cloned with the ZeroBlunt TOPO PCR Cloning Kit for Sequencing (Invitrogen). Inserts were sequenced using vector-specific or gene-specific (shown in supplementary table S1, Supplementary Material online) primers. Full-length sequences have been deposited in DDBJ/EMBL/GenBank under the accession numbers AB284361–AB284366 (shown in supplementary table S3, Supplementary Material online). All experiments described here were performed according to the manufacturers' instructions. The amino acid sequences were inferred from the most upstream ATG with the universal codon table.

Phylogenetic Analysis

All possible MEP pathway orthologs for bacteria for which complete genome sequences were available as of 1

November 2006 were retrieved from the National Center for Biotechnology Information Refseq database. Very large multiple sequence alignments were constructed using MUSCLE 3.6 (Edgar 2004) and then 50 bacteria were chosen (supplementary table S2, Supplementary Material online) based on the alignments to retain at least a single species for each bacterial phylum and to eliminate species with organism-specific indels when possible. Amino acid sequences for 8 eukaryotes were deduced based on the sequenced genome databases (supplementary table S3, Supplementary Material online). Data matrices were constructed as follows: structure-based alignments were created for bacterial sequences using the Expresso (Armougom et al. 2006) Web service (http://www.tcoffee.org/), then eukaryotic sequences were added using ClustalX 1.83 (Thompson et al. 1997), and the results were manually refined and sites chosen for analysis. Although the Expresso Web service was used, ispG and ispH alignments were constructed by conventional methods because no suitable structural data were available for these genes. First, maximum likelihood (ML) trees were inferred with a WAG substitution matrix (Whelan and Goldman 2001) using Phyml 2.4.4 (Guindon and Gascuel 2003) with bootstrap values (100 replications), and distant problematic bacterial sequences (underlined in supplementary table S2, Supplementary Material online) were eliminated. Final ML trees were then inferred within additional operational taxonomic units (OTUs) for dinoflagellates and haptophytes, which have been obtained and synthesized from expressed sequence tags (EST) (source organisms and the original accession numbers are shown in the supplementary table S3, Supplementary Material online). Apicomplexan orthologs were excluded because they contain a vast number of changed or inserted residues, making it difficult to robustly infer phylogeny. In fact, trees drawn to include them resulted in placing them at deeper branches, disturbing tree topology, or reducing statistical support to greater or lesser degrees (data not shown).

Targeting Presequence Analysis

The N-terminal extensions of the translated amino acid sequences for each MEP pathway gene were examined using SignalP-HMM (Nielsen and Krogh 1998; Bendtsen et al. 2004) for SP and SOSUI (Hirokawa et al. 1998) for the transmembrane region. The distribution of protein sorting signals was examined using newly developed sliding-window iteration of TargetP (SWIT) analysis. Amino acid sequences were inferred from the first in-frame start codon, and subcellular localization was predicted by submitting the first 130 residues to the TargetP server (http://www.cbs.dtu.dk/services/TargetP/) (Nielsen et al. 1997; Emanuelsson et al. 2000); this was then repeated successively after eliminating the first residue from the N-terminus of the previously analyzed amino acid sequence (i.e., sliding a 130-residue window with a 1-residue step). Iterations were performed using the newly developed Ruby script, facilitated by the BioRuby library (http://www.bioruby.org/) (Goto et al. 2003). TargetP yields scores for SP, mitochondrial TP (mTP), chroloplast TP (cTP), and an other category for each iteration. The 4 scores were plotted with window positions on the y and x axes. Two superoxide dismutase genes, PmSOD1 and PmSOD2 (AY095212 and AY095213, respectively), were used for comparison (Wright et al. 2002).

Antibody

Rabbits were immunized with 2 synthesized peptides, TATVEDALKHPNWS and YTLAYPQRLHHDGS, which were partial fragments of the deduced ispC peptide sequence. A fraction of IgG (270 μg/ml) from the sera was obtained by affinity purification using the peptide cocktail to capture specific antibody. Recombinant ispC protein (with an N-terminal 6×His tag) lacking the predicted bipartite targeting peptide was expressed in *Escherichia coli* BL21-AI (Invitrogen) and was purified using HisTrap FF crude (GE Healthcare, Little Chalfont, Buckinghamshire, UK). We discovered that the cytosolic fraction of *P. marinus* contained a protein that nonspecifically reacted with normal rabbit IgG (Santa Cruz Biotechnology, Santa Cruz, CA); therefore, all antibodies were first adsorbed with proteins from the cytosolic fraction of *P. marinus* that were obtained by precipitation with 50–70% saturated ammonium sulfate.

Centrifugal Fractionation

Cells were collected by centrifugation (200 × g for 5 min at ambient temperature), washed with phosphate-buffered saline (PBS), resuspended in 50 mM *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid buffer (pH 7.4) supplemented with protease inhibitor cocktail (Complete; F. Hofmann-La Roche, Basel, Switzerland), and disrupted by sonication on ice. The homogenate was subjected to differential centrifugal fractionation at 4 °C and divided into 4 fractions: 200 × g sediment (5 min), 2,000 × g sediment (10 min), 20,000 × g sediment (15 min), and the supernatant. The 20,000 × g sediment was incubated for 30 min at 4 °C in buffer containing 0.5% Triton X-100 and was then centrifuged at 20,000 × g for 15 min. All fractions were subjected to western blotting performed according to the following protocol: proteins separated by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (10% gel) were transferred to PVDF membrane (Immobilon-P; Millipore, Billerica, MA), membranes were blocked for 1 h in blocking buffer (3% bovine serum albumin and 0.05% Tween 20 in PBS), and primary antibody (1:1000 dilution of the affinity-purified antibody) was added for 1 h. Next, alkaline phosphatase-conjugated Goat Anti-Rabbit IgG (Fc) (Promega S3731) in blocking buffer (1:7500) was added for 1 h and then the blots were developed by addition of BCIP/NBT Color Development Substrate (Promega S3771).

Immunofluorescent Microscopy

Cells were incubated for 30 min in culture medium containing 200 nM CMXRos (MitoTracker Red; Invitrogen M7512), washed for 15 min, and then fixed for 30 min at

−20 °C with 4% (w/v) paraformaldehyde in 85% (v/v) methanol. Fixed cells were washed with PBS and then mounted on Matsunami adhesive silane-coated slides (Matsunami Glass Industries, Kishiwada, Japan) and air dried. The slides were blocked for 1 h with BlockAce (Dainippon Sumitomo Pharma, Osaka, Japan) containing 4 μg/ml normal goat IgG (Santa Cruz Biotechnology), labeled for 1 hr with a 1:1000 dilution of affinity-purified antibody, and then incubated for 1 h with 1:400 dilution of Alexa Fluor 488 Goat Anti-Rabbit IgG (H + L; Invitrogen A11008) in Can Get Signal immunostain Solution A (Toyobo, Osaka, Japan). After counterstaining with 1 μg/ml 4′,6-diamidino-2-phenylindole (DAPI), the slides were observed under an epifluorescence microscope (Olympus BX60) equipped with cooled digital color camera (Olympus DP70). For negative control experiments, the affinity-purified antibody was omitted or substituted with the same concentration of normal rabbit IgG.

Results

MEP Pathway Genes

We first searched for MEP pathway genes using a similarity search service at the *P. marinus* genome database of TIGR (http://www.tigr.org/tdb/e2k1/pmg/) using red alga, green alga, and apicomplexan homologs as queries for each gene (identifiers are shown in supplementary table S3, Supplementary Material online). Irrespective of species used as queries, all homologs resulted in the same significant hits: single contigs for *dxs*, *ispC*, *ispE*, and *ispF* and 2 contigs for *ispG* and *ispH*. For the latter genes, the hits were located at the termini of contigs and were assumed to be single genes astride contig gaps. Thus, all these genes were thought to exist only in single copy. In contrast, MVA pathway genes were not found when we searched for them in a similar fashion.

Six out of 7 MEP pathway genes (fig. 1) were then sequenced in their entirety using gene-specific primers designed based on the contig sequences. No homolog of the gene responsible for the third step, *ispD*, was found in the database, and an attempt to amplify the gene using degenerate primers failed. Untranslated regions were short, 19–45 and 15–97 nt for the 5′ and 3′ ends, respectively.

Evolutionary Origin and Relationships

To investigate the evolutionary origin of the sequenced MEP pathway genes, we performed phylogenetic analyses using PBEs and a wide range of bacteria. The dxs ortholog of *P. marinus* clusters with those of PBEs, and they form a clade with the alpha-proteobacterial counterparts with 100% bootstrap support, which is independent of cyanobacterial orthologs (fig. 2). The ispC orthologs of *P. marinus* and PBEs form a clade with the cyanobacterial orthologs with 100% bootstrap support (fig. 3). For the ispE tree, eukaryotic orthologs, including that of *P. marinus* and chlamydial counterparts, form a weakly supported (76%) clade (fig. 4). The ispF ortholog of *P. marinus* clusters with those of PBE with weak bootstrap support (58%), but the clade has failed to show a relationship with specific

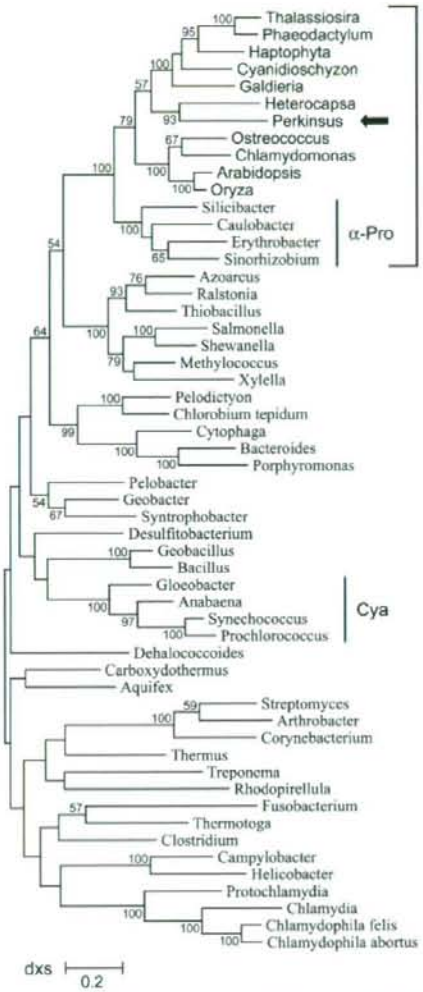

FIG. 2.—Unrooted ML tree for dxs (log likelihood = −35,861.100656) constructed using Phyml with WAG substitution matrix, based on a matrix comprising 55 OTUs and 497 sites. OTU names in sans serif indicate they are eukaryotic orthologs, and the bold arrow denotes the *Perkinsus marinus* ortholog. Haptophyta OTU is a synthetic sequence composed of *Emiliania*, *Isochrysis*, and *Pavlova* ESTs (supplementary table S3, Supplementary Material online) and lacks 23% of sites. Heterocapsa OTU is also a synthetic sequence derived from ESTs and lacks 6% of sites. Cyanobacterial (Cya) and alpha-proteobacterial (α-Pro) orthologs are indicated. Numbers adjacent to the nodes indicate bootstrap support (100 replicates), and values below 50% have been omitted. The scale bar indicates the number of substitutions per site.

bacteria (fig. 5). This is probably because of the shorter data matrix (104 sites), corroborated by poor phylogenetic resolution of the overall tree. For the ispG tree, the *P. marinus* ortholog and eukaryotic orthologs other than those from red algae form a clade with the chlamydial orthologs with 100% bootstrap support, whereas red algal orthologs form another clade with the cyanobacterial orthologs (fig. 6). The

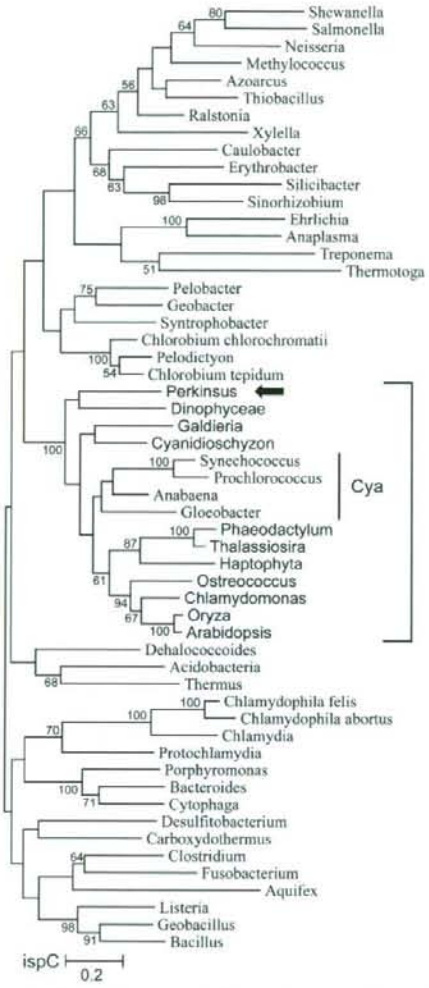

FIG. 3.—Unrooted ML tree for ispC (log likelihood = −25,588.615193) constructed using Phyml with WAG substitution matrix, based on a matrix comprising 55 OTUs and 337 sites. Haptophyta and Dinophyceae OTUs are synthetic sequences composed of *Emiliania* and *Isochrysis* ESTs, and *Alexandrium* and *Crypthecodinium* ESTs (supplementary table S3, Supplementary Material online), and lack 13% and 31% of sites, respectively. Cyanobacterial (Cya) orthologs are indicated. See also the legend of figure 2.

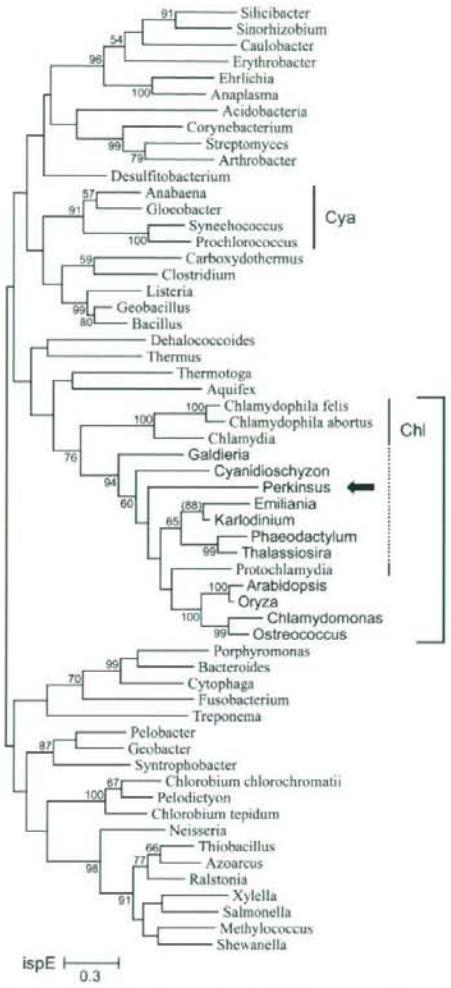

FIG. 4.—Unrooted ML tree for ispE (log likelihood = −16,236.630250) constructed using Phyml with WAG substitution matrix, based on a matrix comprising 58 OTUs and 175 sites. Emiliania and Karlodinium OTUs are partial sequences derived from ESTs (supplementary table S3, Supplementary Material online), lack 50% and 54% of sites, respectively, and do not overwrap; therefore, it could be an artifact that they form a clade although with a high bootstrap value (88, as indicated in parentheses). Cyanobacterial (Cya) and chlamydial (Chl) orthologs are indicated. See also the legend of figure 2.

ispH orthologs of *P. marinus* and PBEs form a clade with the cyanobacterial orthologs with 100% bootstrap support (fig. 7). All these results indicate that each of *P. marinus* orthologs clusters with the corresponding PBE orthologs; thus, the *P. marinus* orthologs are unlikely to have been independently transferred from bacteria.

Some of the *P. marinus* orthologs showed weak phylogenetic affinity to the dinoflagellate orthologs. The dxs ortholog formed a highly supported clade with the dinoflagellate *Heterocapsa* (fig. 2). The ispC ortholog formed a clade with a synthetic dinoflagellate OTU with <50% bootstrap support (fig. 3). Additionally, when a short sequence (105 sites) from the dinoflagellate *Amphidinium* ortholog was included in the ispG analysis, it was sister to the *Perkinsus* ortholog but with weak bootstrap support (54%). However, bootstrap supports for the adjacent nodes became weak, possibly due to the short dinoflagellate sequence (data not shown). The statistical weakness seems attributed to missing data derived from using partial ESTs of the dinoflagellate orthologs. In fact, the dxs tree, which has

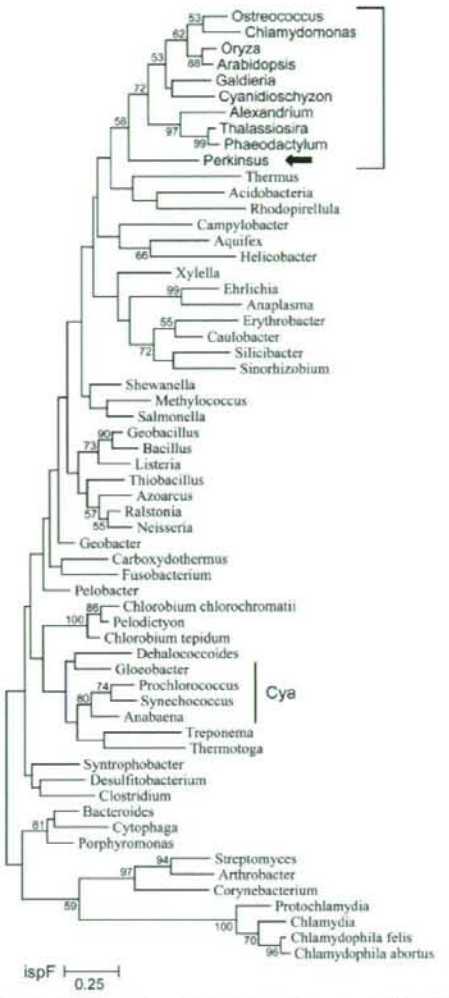

FIG. 5.—Unrooted ML tree for ispF (log likelihood = −7,405.683963) constructed using Phyml with WAG substitution matrix, based on a matrix comprising 60 OTUs and 104 sites. Alexandrium OTU is a partial sequence derived from ESTs (supplementary table S3, Supplementary Material online) and lacks 11% sites. Cyanobacterial (Cya) orthologs are indicated. See also the legend of figure 2.

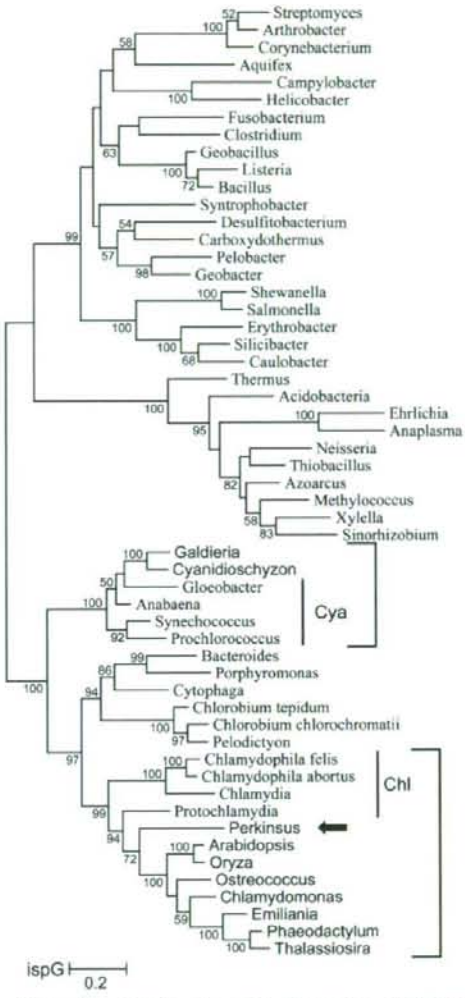

FIG. 6.—Unrooted ML tree for ispG (log likelihood = −21,369.78023) constructed using Phyml with WAG substitution matrix, based on a matrix comprising 55 OTUs and 344 sites. Emiliania OTU is a synthetic sequence derived from ESTs (supplementary table S3, Supplementary Material online). Amphidinium OTU was excluded because it was derived from a short EST and lacked 70% of sites; however, it was sister to the *Perkinsus* ortholog with weak (54%) bootstrap support when it was included in the matrix. Cyanobacterial (Cya) and chlamydial (Chl) orthologs are indicated. See also the legend of figure 2.

the strongest support, has the lowest amount of missing data (6% of sites) in the dinoflagellate ortholog. Therefore, the strength of this phylogenetic model may improve if full-length sequences become available. Reliable relationships for ispE and ispF orthologs of *P. marinus* were not resolved, and this was probably due to long-branch attraction or the shorter data matrix (figs. 4 and 5). The ispE tree (fig. 4) shows a highly supported close relationship between orthologs of the haptophyte *Emiliania* and the dinoflagellate *Karlodinium*. Although *Karlodinium* has a tertiary plastid derived from haptophytes, the close relationship should not be interpreted as a result of endosymbiotic gene transfer, but rather as an artifactual grouping of stray OTUs, because these 2 OTUs are derived from partial ESTs and have no overwrap on the sequence alignment. To summarize, in total, the MEP pathway genes of *P. marinus* have weak phylogenetic affinity to those of dinoflagellates, and this likely shows that the MEP pathway is descended from the common ancestor of *P. marinus* and dinoflagellates.

The present phylogenetic trees indicate that the MEP pathway genes of PBEs and *P. marinus* have essentially the

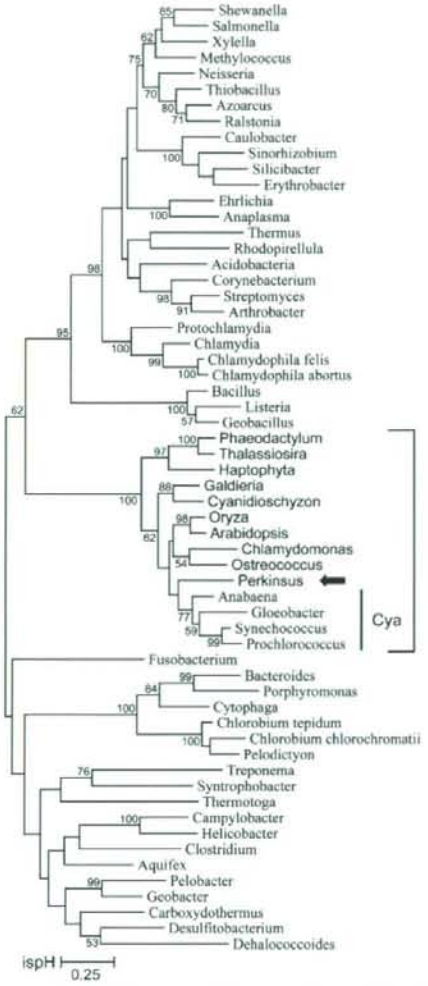

FIG. 7.—Unrooted ML tree for ispH (log likelihood = −18,623.672548) constructed using Phyml with WAG substitution matrix, based on a matrix comprising 60 OTUs and 230 sites. Haptophyta OTU is a synthetic sequence composed of *Emiliania* and *Pavlova* ESTs (supplementary table S3, Supplementary Material online) and lacks 10% of sites. Cyanobacterial (Cya) orthologs are indicated. See also the legend of figure 2.

same pattern of mosaic origins in Cyanobacteria, Proteobacteria, and Chlamydia (table 1). In PBEs, dxs (fig. 2) and ispE (fig. 4) had a sister relationship with the proteobacterial and chlamydial orthologs, respectively, whereas ispC and ispH seemed to be derived from Cyanobacteria (figs. 3 and 7). A chlamydial origin was also observed for ispD orthologs (supplementary fig. S1, Supplementary Material online). The evolutionary origin of ispF remains unclear, probably owing to insufficient informational sites being used (fig. 5). However, only red algae had cyanobacteria-like ispG orthologs, and *P. marinus* and other PBEs

Table 1
Shared Mosaic Origins of MEP Pathway Genes

	dxs	*ispC*	*ispD*	*ispE*	*ispF*	*ispG*	*ispH*
Green plants	pro	cya	chl	chl?	#	chl	cya
Red algae	pro	cya	chl	chl?	#	cya	cya
Diatoms	pro	cya	chl	chl?	#	chl	cya
Haptophytes	pro	cya	chl	chl?	—	chl	cya
Dinoflagellates	pro	cya	chl	chl?	#	chl?	—
Perkinsus	pro	cya	—	chl?	#	chl	cya

NOTE.—chl, chlamydia; cya, cyanobacteria; pro, α-proteobacteria; —, not found; ?, supported with weak bootstrap values; and #, monophyletic but showing no clear relationships.

harbored genes that were closely related to their chlamydial counterparts (fig. 6). This difference is not a phylogenetic artifact because there were long and well-aligned insertion sequences that were shared with PBEs other than red algae and Chlamydia (supplementary fig. S2, Supplementary Material online). The present phylogenetic analyses demonstrated that *P. marinus* and all PBEs analyzed here, with a curious exception of red algae, obtained their MEP pathway genes from an identical source.

Targeting Presequence and Subcellular Localization

The inferred amino acid sequences of the 6 MEP pathway genes in *P. marinus* have obvious N-terminal extensions relative to their bacterial homologs (fig. 1). N-terminal extensions often function as targeting presequences to deliver the peptide into certain subcellular compartments (e.g., ER, mitochondria, and plastids). First, SignalP predicted that all genes would have SPs (fig. 1; supplementary fig. S3, Supplementary Material online), although the probability of cleavage was not very high. Each predicted SP indicated an abnormally long (10–20 residues) hydrophilic n-region, but h- and c-regions were normal (supplementary fig. S3, Supplementary Material online). Because elimination of a few residues from the N-termini of the n-regions greatly improved the cleavage probability (data not shown), we consider that SPs with minor modification exist and they are likely cleaved off. The cleavable SP indicates that the MEP pathway enzymes are trafficked via the secretory pathway. However, TargetP predicted that all MEP pathway orthologs (except dxs) would be targeted to mitochondria. A lesson learned from the SignalP result is that simple predictions using full-length sequences may yield dubious results if the organism under question is distantly related to organisms after which the predictor has been modeled. Therefore, the need exists to monitor trends and the robustness of results obtained by TargetP analysis to better understand whether the mitochondrial targeting was significant and to further investigate the characteristics of the N-terminal extension.

We developed a new method of exploiting TargetP in order to robustly examine the distribution of targeting preferences. The method iteratively invoked TargetP with a sliding 130-residue window from the N-terminus; thus, we named the method SWIT. Figure 8 shows the SWIT results for the MEP pathway genes and 2 *P. marinus* superoxide dismutases, PmSOD1 and 2 (Wright et al. 2002). For

example, for dxs, a prediction with the first 130 residues yielded scores of 0.002, 0.062, 0.165, and 0.064 for cTP, mTP, SP, and others (*y* intercepts for green, blue, yellow, and gray lines), respectively; this is equivalent to the simple TargetP prediction using the full-length sequence. As the window slides downstream, the SP score increases up to 0.393 and then decreases to near zero. Although the first result (*y* intercept) alone, or the simple TargetP prediction, seems rather unreliable, we recognize that the N-terminal region likely has SP preference (yellow arrow). As the window slides along, a significant prominence for cTP appears (green arrow) in the vicinity of the predicted SP cleavage site (yellow triangle). The MEP pathway genes shared this characteristic SWIT trend, that is, a peak in the SP score at the N-terminal region and a subsequent increase in the cTP score. The trends seem to suggest that these enzymes are localized to secondary plastids because an N-terminal bipartite (SP + cTP) presequence is a characteristic feature of proteins of secondary plastids (van Dooren et al. 2001). Thus, we hypothetically considered residues between the N-terminus and the predicted SP cleavage site and between the SP cleavage site and the most downstream predicted site for cTP cleavage (green triangles) as putative SP and cTP, respectively.

One issue is that all aforementioned genes, except dxs, have considerable mTP scores at the *y* intercepts (fig. 8), which correspond to the simple TargetP results (see above). Both PmSOD1 and 2 also have high mTP scores at their N-termini, similar to the 5 MEP pathway genes (fig. 8); however, PmSOD2 localizes to an unknown, nonmitochondrial compartment, whereas PmSOD1 localizes to the mitochondria (Schott and Vasta 2003). According to the SWIT result, the mTP score for PmSOD1 is maintained along the N-terminal extension, and this corresponds to its mitochondrial localization; however, the score for PmSOD2 rapidly declines and SP and subsequent cTP preferences appear alternatively. Given that SP and cTP preferences do not appear for PmSOD1, they can be used to discriminate localization to a compartment other than the mitochondria; additionally, a high mTP score at the N-terminus has no biological significance when there are significant SP and cTP preferences. The MEP pathway genes share the SP + cTP feature with PmSOD2; thus, the MEP pathway genes appear to have protein sorting signals that are characteristic of proteins targeted to secondary plastids and to be localized to a compartment other than the mitochondria.

The predicted TP regions suggested by SWIT analysis were further examined in comparison to the TPs of PBEs. The predicted regions were overall hydrophilic and with a net positive charge, as in PBEs, resulting from depletion of acidic residues (5.3% compared with 12.9% in the homologous region) and accumulation of hydroxylated residues (especially Ser; 16.2% compared with 6.4%) and positive-charged residues (especially Arg; 7.1% compared with 5.2%). All but ispF were shown to possess Phe residues in a hydrophobic context in proximity to the SP cleavage site, and a single transmembrane helix followed by an Arg-rich region within the TP region (fig. 1; supplementary fig. S3, Supplementary Material online). This feature resembles a similar feature in the class I TP of peridinine dinoflagellates, and no apicomplexans reported so far harbor

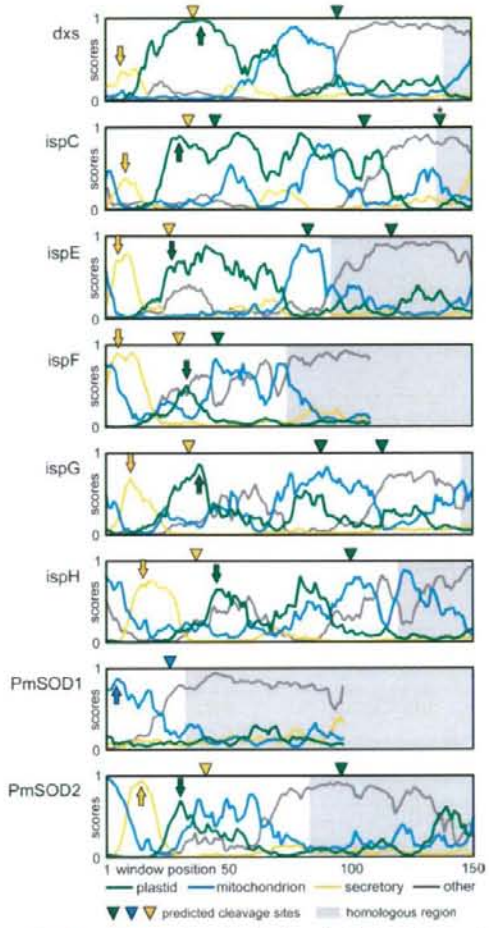

FIG. 8.—SWIT results showing the distribution of targeting preferences for the N-terminal presequences. For each gene, the neural network scores (*y* axis) of the TargetP prediction are plotted for the N-terminal 150 amino acids (*x* axis) in the following colors: green, plastid TP; light blue, mitochondrial TP; yellow, SP for secretory pathway; and gray, other location. Thus, the *y* intercepts correspond to the simple TargetP predictions. Triangles above each plot indicate the predicted cleavage sites: green, sites for plastid TP that appeared in the SWIT analysis; light blue, sites for mitochondrial TP (only for PmSOD1); and yellow, sites for SP as predicted by SignalP-HMM. The homologous region is shown in gray on each plot.

the transmembrane helix (Patron et al. 2005). No hydrophobic regions were predicted for ispF or PmSOD2, which makes them similar to the class II TP on the other hand (Patron et al. 2005). Collectively, the predicted TP regions resemble those of the peridinine dinoflagellates with respect to physicochemical features and class duality.

To examine whether the predicted bipartite targeting sequence is functional, a polyclonal antibody recognizing the second step enzyme coded by *ispC* was prepared. The affinity-purified antipeptide antibody successfully detected recombinant ispC (fig. 9*A*, lane 1), and it detected

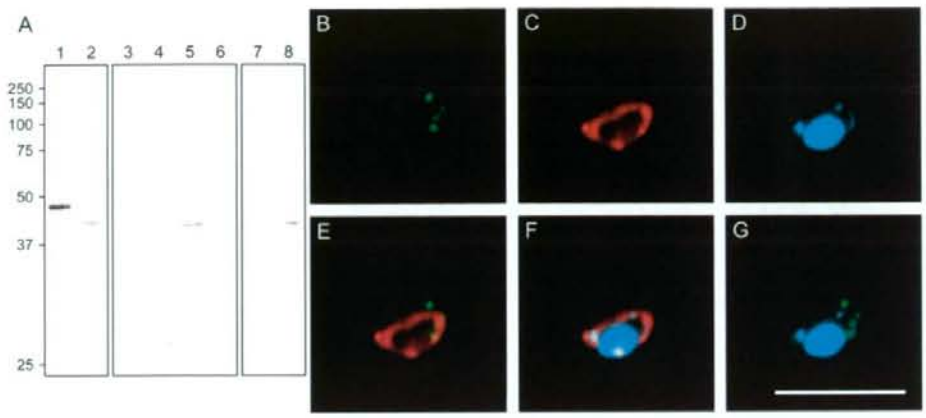

FIG. 9.—Immunological detection of the second step enzyme ispC. (*A*) Fractionation experiments for ispC. The size of ispC can be seen in lanes 1 and 2; a His-tagged recombinant ispC protein (1) acts as a positive control, and *Perkinsus marinus* cell lysate (2) contains native ispC. Lanes 3–6 show centrifugal fractionation of homogenized cell lysate; equivalent amounts of the 200 × g sediment (3), 2,000 × g sediment (4), 20,000 × g sediment (5), and 20,000 × g supernatant (6) were applied. Lanes 7 and 8 show the effect of Triton X-100 treatment on the 20,000 × g sediment (5); equivalent amounts of the 20,000 × g sediment (7) and 20,000 × g supernatant (8) were applied. Immunofluorescent microscopy indicating localization of ispC (*B–G*). Alexa Fluor 488 signals labeling ispC (*B*), CMXRos signals indicating mitochondria (*C*), and DAPI signals indicating DNA (*D*) were obtained from the same visual field. An overlaid image of ispC and mitochondria (*E*) indicates that the ispC protein localizes to punctate compartments near mitochondria. Those of DNA and mitochondria (*F*), and DNA and ispC (*G*) indicate that the mitochondria definitely contain DNA but the putative plastids do not contain detectable amounts of DNA. Bar indicates 10 μm.

a protein of approximately 43 kDa in *P. marinus* cell lysates (lane 2). Although smaller in size than predicted molecular weight of ispC (51 kDa), the size corresponds to an estimated molecular weight of the mature protein after cleavage of the predicted bipartite targeting peptides (asterisk in fig. 8). Detection at 43 kDa suggests that ispC is transported using targeting peptides that are subsequently cleaved. SOSUI predicted that the mature ispC protein was soluble. However, centrifugal fractionation showed the ispC protein specifically in the 20,000 × g sediment (lane 5), indicating that the protein was not cytosolic. The protein was easily solubilized by a mild detergent (lane 8), suggesting that it may be associated with membranes or membrane-bounded organelles. Finally, immunofluorescent microscopy revealed dotted signals for ispC (fig. 9*B*), and these signals were not associated with mitochondria (fig. 9*C*) or DNA (fig. 9*D*). Cells typically had multiple fluorescent spots that were frequently located near mitochondria (fig. 9*E*). Staining for extranuclear DNA overlapped with mitochondrial fluorescence (fig. 9*F*) but not with ispC fluorescence (fig. 9*G*).

Discussion

Plastids in *P. marinus*

The presence of 6 MEP pathway genes (fig. 1), together with the lack of any MVA pathway genes, suggests that the MEP pathway is responsible for de novo isoprenoid synthesis in *P. marinus*. A little possibility cannot be ruled out that the MEP pathway is not functional because we have been unable to find an *ispD* ortholog, but it does not impact on our logic for proving the existence of plastids in *P. marinus* (see below). Of course, activity of the MEP pathway still requires characterization in order to further discuss plastid function; however, this is beyond the scope of the present study.

The *P. marinus* genes reported here are very likely relevant to plastids because the MEP pathway is specific to PBE, all PBEs have obtained the genes from essentially the same source (table 1), and all MEP pathway genes of *P. marinus* consistently group with them (figs. 2–7). Furthermore, the fact that all these genes have the bipartite targeting sequence (figs. 1 and 8), plus the fact that the predicted targeting sequences are cleaved (fig. 9*A*, lane 2), supports our idea that these enzymes traffic using machinery homologous to those of secondary PBEs. Finally, a subcellular fractionation experiment indicated that ispC is not cytosolic and is associated with membranes or organelles (fig. 9*A*, lanes 5 and 8), and immunofluorescent microscopy showed punctate localization of ispC near mitochondria (fig. 9*E*). Generally said, if a group of proteins accumulates at a subcellular compartment by means of machinery reasonably homologous to that of plastids, that compartment should be identified as a plastid. Although recent studies have shown the evidence for *P. marinus* plastid (Stelter et al. 2007; Teles-Grilo et al. 2007), no direct evidence that completely fulfills this criterion has been presented. To our knowledge, we are the first to use proper evidence in order to demonstrate that *P. marinus* harbors secondary plastids.

We revealed the existence of *P. marinus* plastids by detecting ispC using immunofluorescent microscopy. However, our SWIT analysis showed that ispF and PmSOD2 had similar but slightly different bipartite presequences (figs. 1 and 8). Although PmSOD2 has also shown punctate localization (Schott and Vasta 2003), it remains to be elucidated whether ispC and PmSOD2 are colocalized. Another point to be examined is that the ispC spots do not