Core Signaling Pathways in Human Pancreatic Cancers Revealed by Global Genomic Analyses

Siân Jones,[1*] Xiaosong Zhang,[1*] D. Williams Parsons,[1,2*] Jimmy Cheng-Ho Lin,[1*] Rebecca J. Leary,[1*] Philipp Angenendt,[1*] Parminder Mankoo,[3] Hannah Carter,[3] Hirohiko Kamiyama,[4] Antonio Jimeno,[1] Seung-Mo Hong,[4] Baojin Fu,[4] Ming-Tseh Lin,[4] Eric S. Calhoun,[1] Mihoko Kamiyama,[4] Kimberly Walter,[4] Tatiana Nikolskaya,[5] Yuri Nikolsky,[6] James Hartigan,[7] Douglas R. Smith,[7] Manuel Hidalgo,[1] Steven D. Leach,[1,8] Alison P. Klein,[1,4] Elizabeth M. Jaffee,[1,4] Michael Goggins,[1,4] Anirban Maitra,[1,4] Christine Iacobuzio-Donahue,[1,4] James R. Eshleman,[1,4] Scott E. Kern,[1,4] Ralph H. Hruban,[1,4] Rachel Karchin,[3] Nickolas Papadopoulos,[1] Giovanni Parmigiani,[1,9] Bert Vogelstein,[1]† Victor E. Velculescu,[1]† Kenneth W. Kinzler[1]†

There are currently few therapeutic options for patients with pancreatic cancer, and new insights into the pathogenesis of this lethal disease are urgently needed. Toward this end, we performed a comprehensive genetic analysis of 24 pancreatic cancers. We first determined the sequences of 23,219 transcripts, representing 20,661 protein-coding genes, in these samples. Then, we searched for homozygous deletions and amplifications in the tumor DNA by using microarrays containing probes for ~10^6 single-nucleotide polymorphisms. We found that pancreatic cancers contain an average of 63 genetic alterations, the majority of which are point mutations. These alterations defined a core set of 12 cellular signaling pathways and processes that were each genetically altered in 67 to 100% of the tumors. Analysis of these tumors' transcriptomes with next-generation sequencing-by-synthesis technologies provided independent evidence for the importance of these pathways and processes. Our data indicate that genetically altered core pathways and regulatory processes only become evident once the coding regions of the genome are analyzed in depth. Dysregulation of these core pathways and processes through mutation can explain the major features of pancreatic tumorigenesis.

Worldwide, over 213,000 patients will develop pancreatic cancer in 2008, and nearly all will die of their disease (*1*–*3*). Several genetic alterations have been identified in these lethal cancers, including those in the *CDKN2A*, *SMAD4*, and *TP53* tumor suppressor genes and in the *KRAS* oncogene (*4*–*8*). Although the discoveries of these genes have provided important insights into the natural history of the disease and have spurred efforts to develop improved diagnostic and therapeutic agents, the vast majority of human genes have not been analyzed in this cancer type.

We examined the genetic makeup of human pancreatic cancers in unprecedented detail. Because all human cancers are primarily genetic diseases, we hoped to identify additional genes and signaling pathways that could guide future research on this disease.

Sequencing strategy. The sequences of protein-coding exons from 20,735 genes were identified and used to design primers for 219,229 amplicons covering these regions (*9*). DNA from 24 advanced pancreatic adenocarcinomas (table S1) was polymerase chain reaction (PCR)–amplified with these primers and sequenced with the use of fluorescent dye terminators (*9*). The 24 tumors were passaged in vitro as cell lines or in nude mice as xenografts to remove contaminating non-neoplastic cells, facilitating detection of mutations (*10*–*12*). Exons containing variant sequences were reamplified and resequenced from the tumor DNA as well as from normal DNA from the same patient to confirm the mutation and to ensure that the mutation was somatic (i.e., that it was not present in normal cells). PCR products from 208,311 amplicons resulted in PCR products that were successfully sequenced and met stringent quality controls (table S2). These amplicons included 94.5% of the targeted sequences and yielded high-quality sequencing data for 98.5% of the target bases within these amplicons. The 208,311 successfully sequenced amplicons yielded mutational data on 23,219 transcripts representing 20,661 genes.

Somatic mutations. Among the 1562 somatic mutations detected with this strategy, 25.5% were synonymous, 62.4% were missense, 3.8% were nonsense, 5.0% were small insertions and deletions, and 3.3% were at splice sites or within the untranslated region (UTR) (Table 1 and table S3). The spectra of somatic mutations can yield insights into potential carcinogens and other environmental exposures. Table 1 lists the spectra observed in the four tumors that have been subjected to large-scale sequencing analyses of the majority of protein-encoding genes. It is evident that breast tumors have a unique somatic mutation spectrum, with a preponderance of mutations at 5′-TpC sites and a relatively small number of mutations at 5′-CpG sites. However, the spectra of colorectal (*13*, *14*), brain (*15*), and pancreatic tumors are similar, suggesting that breast epithelial cells are exposed to different levels or types of carcinogens or use distinctive repair systems (*16*, *17*). Given that cells in the colon are expected to be exposed to dietary carcinogens more than breast, brain, or pancreatic cells, one possible interpretation of these results is that dietary components are not directly responsible for causing most of the mutations found in human cancers.

Of the 20,661 genes analyzed by sequencing, 1327 had at least one mutation, and 148 had two or more mutations among the 24 cancers surveyed (table S3). In addition to the frequency of mutations, the type of mutation can provide information useful for evaluating its potential role in disease (*18*). Nonsense mutations, out-of-frame insertions or deletions, and splice-site changes generally lead to inactivation of the protein products. To evaluate missense mutations, we developed an algorithm that uses machine learning of 58 predictive features based on the physical-chemical properties of amino acids involved in the substitutions and their evolutionary conservation at equivalent positions of conserved proteins (*9*). Of the 924 missense mutations that could be scored with this algorithm, 160 (17.3%) were predicted to contribute to tumorigenesis when assessed by this method (table S3).

We also generated structural models of 404 of the missense mutations identified in this study [links to structural models available at (*19*)]. In each case, the model was based on x-ray crystallography or nuclear magnetic resonance spectroscopy of the normal protein or a closely related homolog. This analysis showed that 55 of the 404 mutations were located near a domain interface or ligand-binding site and were likely to affect function (examples in Fig. 1).

The average number of somatic mutations in pancreatic cancers (48; Table 2) is considerably

[1]Sol Goldman Pancreatic Cancer Research Center, Ludwig Center and Howard Hughes Medical Institute at the Johns Hopkins Kimmel Cancer Center, Baltimore, MD 21231, USA. [2]Department of Pediatrics, Section of Hematology-Oncology, Baylor College of Medicine, Houston, TX 77030, USA. [3]Department of Biomedical Engineering, Institute of Computational Medicine, Johns Hopkins Medical Institutions, Baltimore, MD 21218, USA. [4]Department of Pathology, Johns Hopkins Medical Institutions, Baltimore, MD 21231, USA. [5]Vavilov Institute for General Genetics, Moscow 8333, 117809, Russia. [6]GeneGo, Incorporated, St. Joseph, MI 49085, USA. [7]Agencourt Bioscience Corporation, Beverly, MA 01915, USA. [8]Department of Surgery, Johns Hopkins Medical Institutions, Baltimore, MD 21231, USA. [9]Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: bertvog@gmail.com (B.V.); velculescu@jhmi.edu (V.E.V.); kinzlke@jhmi.edu (K.W.K.)

less than that in breast cancer (101) or colorectal cancers (77) ($P < 0.001$), even though fewer genes were sequenced in the latter two tumor types (*14*). One plausible explanation for this lower rate is that the cells that initiate pancreatic tumorigenesis have gone through fewer divisions than colorectal or breast cancer cells. It has been previously shown that the majority of mutations observed in colorectal cancers are likely to have occurred in the normal stem cells that gave rise to the initiating neoplastic cell (*12*). Our data are thus consistent with observations showing that normal pancreatic epithelial cells divide infrequently (*20*, *21*).

We further evaluated 39 genes that were mutated in more than one of the 24 discovery screen cancers in a prevalence screen consisting of 90 pancreatic cancers. In this screen, we detected 255 nonsilent somatic mutations among 23 genes (table S4). The nonsilent mutation rate of the genes in the prevalence screen (excluding *KRAS*, *TP53*, *CDK2NA*, and *SMAD4*) was higher than that in the discovery screen (3.6 versus 1.47 nonsilent mutations per Mbase, $P < 0.001$). The fraction of nonsilent mutations observed in these 19 genes was also higher than that observed in the genes assessed in the discovery screen ($P = 0.052$). These data are consistent with the hypothesis that a greater fraction of the genes tested in the prevalence screen were positively selected during tumorigenesis.

Deletions. By using oligonucleotide arrays containing probes for 1,069,688 SNPs and robust algorithms for identifying deletion events from SNP array data (*22*), we identified 198 separate homozygous deletions among the 24 pancreatic cancers (table S5). The average size of these deletions was 335,000 bp. Additionally, we observed many regions that had undergone single-copy losses, often manifest as losses of heterozygosity, including losses of whole chromosomes or whole chromosome arms. We did not pursue these changes because it is difficult to reliably identify target genes from such large regions unless the residual copy of the gene on the nondeleted chromosome is mutated. Such target genes would have already been called to our attention by the results of the discovery sequencing screen and would have been scored as homozygous changes (table S3).

According to the allelic two-hit hypothesis, the presence of a homozygous deletion indicates that a tumor suppressor gene exists within the deleted region (*23*). To determine the most likely target within these deletions, we used the results from our analysis of point mutations as well as expression analyses (see below) and previously published studies. For a gene to be considered the target, a portion of its coding region had to be affected by the homozygous deletion, and the gene (i) had to harbor a nonsilent sequence alteration in a different tumor, (ii) had to be a well-documented tumor suppressor gene, or (iii) had to have corroborating expression data. The presumptive target genes for the homozygous deletions that met these criteria are listed in table S5. This list includes the classic tumor suppressor genes *CDKN2A* (p16), *SMAD4*, and *TP53*, as well as genes that had not previously been implicated in pancreatic cancer development.

When an exon of a gene is truly deleted in a tumor, no sequencing information should be obtainable, providing confirmation of the deletion. Without exception, the homozygous deletions found through the SNP arrays were consistent with the sequencing data (*9*). Furthermore, there was only one homozygous deletion revealed by sequencing that was not evident in the microarray hybridizations (a four-exon deletion of *SMAD4* in tumor Pa21C).

The number of deletions in a tumor was more variable than the number of somatic mutations, averaging 8.3 and ranging between 2 and 20 per tumor (Fig. 2). However, each homozygous deletion completely abrogated the function of the target gene as well as all other genes within the deleted region, whereas only a fraction of the somatic mutations were predicted to alter the gene's function. In a typical pancreatic cancer, ~10 genes (including targets and nearby genes) are eradicated by homozygous deletion, providing fertile grounds for therapeutic strategies that target such losses (*24*, *25*).

Amplifications. With the use of algorithms similar to those described above for deletions (*22*), we identified 144 focal high-copy amplifications in the 24 tumors (table S6). We also identified a variety of low-copy-number gains of entire chromosomes, chromosomal arms, or other large genomic regions that were not pursued because of the difficulty in reliably identifying candidate cancer genes from such large chromosomal regions. To determine the most likely target of the focal amplifications, we again used the results from our mutational data, expression analyses, and previously published data. The presumptive target genes for each of the amplifications that met predefined criteria (*9*) are listed in table S6. There were fewer amplifications than homozygous deletions or point mutations in most pancreatic tumors (Fig. 2).

Passenger mutation rates. The primary goal of cancer genome studies is the identification of genes that are likely to play a causal role in the neoplastic process (potential drivers). One can categorize the best candidate cancer genes (*CAN* genes) on the basis of their mutation frequencies and types. This categorization requires an estimate of the passenger mutation rate (*13*, *14*, *26*). For each of the genes containing somatic mutations, passenger probabilities were determined by using estimated minimal and maximal passenger mutation rates after taking into account the size of the gene, its nucleotide composition, and other relevant factors [Table 1 and (*9*)]. To analyze the probability that a given gene would be involved in an amplification or deletion, we made the

Table 1. Summary of somatic mutations in four tumor types. Pancreas data have their basis in 24 tumors analyzed in the current study; brain data have their basis in 21 nonhypermutable tumors analyzed in (*15*); and colorectal and breast data have their basis in 11 breast and 11 colorectal tumors analyzed in (*14*). Nonsilent numbers in parentheses refer to percentage of total nonsynonymous mutations, and substitutions numbers in parentheses refer to percentage of total substitutions. The total number of substitutions includes synonymous as well as nonsilent mutations identified in the indicated study.

	Pancreas	Brain	Colorectal	Breast
Number of mutated genes	1007	685	769	1026
Number of nonsilent mutations	1163	748	849	1112
Missense	974 (83.7)	622 (83.2)	722 (85)	909 (81.7)
Nonsense	60 (5.2)	43 (5.7)	48 (5.7)	64 (5.8)
Insertion	4 (0.3)	3 (0.4)	4 (0.5)	5 (0.4)
Deletion	43 (3.7)	46 (6.1)	27 (3.2)	78 (7.0)
Duplication	31 (2.7)	7 (0.9)	18 (2.1)	3 (0.3)
Splice site or UTR	51 (4.4)	27 (3.6)	30 (3.5)	53 (4.8)
Total number of substitutions	1484	937	893	1157
	Substitutions at C:G base pairs			
C:G to T:A	798 (53.8)	601 (64.1)	534 (59.8)	422 (36.5)
C:G to G:C	142 (9.6)	67 (7.2)	61 (6.8)	325 (28.1)
C:G to A:T	246 (16.6)	114 (12.1)	130 (14.6)	175 (15.1)
	Substitutions at T:A base pairs			
T:A to C:G	142 (9.6)	87 (9.3)	69 (7.7)	102 (8.8)
T:A to G:C	79 (5.3)	24 (2.6)	59 (6.6)	57 (4.9)
T:A to A:T	77 (5.2)	44 (4.7)	40 (4.5)	76 (6.6)
	Substitutions at specific dinucleotides			
5′-CpG-3′	563 (37.9)	404 (43.1)	427 (47.8)	195 (16.9)
5′-TpC-3′	218 (14.7)	102 (10.9)	99 (11.1)	395 (34.1)

conservative assumption that the overall frequency of all observed amplifications and deletions represented the passenger mutation rate (*9*).

Passenger probabilities for all genes in which at least two genetic alterations were identified in the discovery screen are listed in table S7. This list includes all genes previously known to play an important role in pancreatic cancer through mutation or copy number change, providing experimental confirmation of our general approach. Importantly, the *CAN* genes listed in table S7 included numerous other genes of potential biological interest, many of which had not previously been identified as playing an important role in this tumor type. Examples include the transcriptional activator *MLL3*; cadherin homologs *CDH10*, *PCDH15*, and *PCDH18*; the α-catenin *CTNNA2*; the dipeptidyl-peptidase *DPP6*; the angiogenesis inhibitor *BAI3*; the heterotrimeric guanine nucleotide–binding protein (G-protein)–coupled receptor *GPR133*; the guanylate cyclase *GUCY1A2*; the protein kinase *PRKCG*; and *Q9H5F0*, a gene of unknown function. These genes were generally mutated at much lower frequencies than those previously identified to be mutated in pancreatic cancers (table S7). This is compatible with the idea that conventional strategies were able to identify frequently mutated genes but not the bulk of the genes that are genetically altered in pancreatic cancers.

Candidate pathways and processes promoting pancreatic tumorigenesis. Because most cellular pathways and processes involve multiple

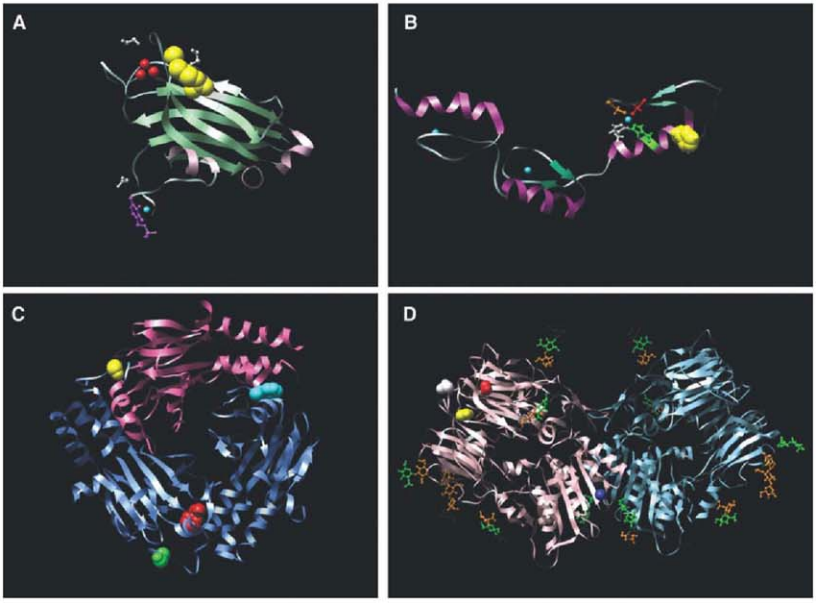

Fig. 1. Examples of structural models of mutations. (**A**). The x-ray crystal structure of the C2 domain of protein kinase C γ (PKCG) [Protein Data Bank identification number (PDBID) 2UZP]. R252 (*41*) is shown as yellow space-fills; Ca^{2+} ions are shown as red spheres. The ligands 1,2-ethanediol and pyridoxal-5-phosphate are shown in white and purple ball-and-stick representations, respectively. The R252→H252 (R252H) mutation could reduce the membrane binding of the C2 domain of PRKCG and thereby affect function. (**B**) The nuclear magnetic resonance solution structure of the three tandem repeats of zf-C2H2 domains from human Kruppel-like factor 5 (KLF5) (PDBID 2EBT). H389 is shown as yellow space-fills; Zn^{2+} ions are shown as cyan spheres. The residues comprising the C2H2 group that coordinate the nearby Zn^{2+} ion are shown as ball-and-stick representations, H393 and H397 are shown in green and white, whereas C380 and C375 are shown in orange and red. The mutation at position 389 (H389N) may disrupt the structure of the zinc finger or nearby zinc coordination site. (**C**) The x-ray crystal structure of the heterotrimer of SMAD3 (two subunits shown as blue ribbons) and SMAD4 (one subunit shown as pink ribbons) (PDBID 1U7F). The residues corresponding to two of the mutant positions (F260S and S422F, shown as red and yellow space-fills, respectively, in chain A) are located at interfaces and could perturb Smad3-Smad3 or Smad3-Smad4 interactions. In chain B, F260 is shown as cyan space-fills and S422 as green space-fills. (**D**) The x-ray crystal structure of the extracellular domain of human DPP6 as a homodimer (PDBID 1XFD). Two of the mutated residues found in this study, T409I (shown as red space-fills) and D475N (shown in yellow space-fills) are in spatial proximity and are close to one of the glycosylation sites, N471 (shown as white space-fills). These mutations fall in the β-propeller domain of the protein (residues 142 to 322 and 351 to 581) thought to be involved in protein-protein interactions. The A778T mutation (shown as blue space-fills) falls in the */β hydrolase domain (residues 127 to 142 and 581 to 849) and is close to the homodimer region of the protein and could perturb the homodimer association. Carbohydrates with glycosylation sites are shown in stick representation. Images created with UCSF Chimera version 1.2422 for Linux (*42*).

proteins functioning in a concerted manner, it is possible that mutations in different genes result in similar tumorigenic effects. Because nearly all of the protein-coding genes in the human genome were evaluated in the current study, the data provided a unique opportunity to investigate groups of genes operating though specific signaling pathways and processes. Sets of genes involved in signaling pathways or cellular processes were defined through three well-annotated GeneGo MetaCore databases: gene ontology (GO), canonical gene pathway maps (MA), and genes participating in defined cellular processes and networks (GG) (*27*). We developed a statistical approach that provided a combined probability that a gene set contained driver alterations, taking into account all types of genetic alterations evaluated in this study (*22*). For each gene set, we considered whether the component genes were more likely to be affected by a genetic alteration than would be predicted by the passenger mutation rate.

These analyses identified 69 gene sets that were genetically altered in the majority of the 24 cancers examined (table S8). Thirty-one of these sets could be further grouped into 12 core signaling pathways and processes that were each altered in 67 to 100% of the 24 cancers analyzed and had clear functional relevance to neoplasia based on annotations in the databases described above (Table 2). The core pathways included those in which a single, frequently altered gene predominated, such as in KRAS signaling and in the regulation of the G1/S cell cycle transition; pathways in which a few altered genes predominated, such as in TGF-β signaling; and pathways in which many different genes were altered, such as in integrin signaling, regulation of invasion, homophilic cell adhesion, and small guanine triphosphatase (GTPase)–dependent signaling.

Analysis of gene expression. Gene expression patterns can inform the analysis of pathways because they can reflect epigenetic alterations not detectable by sequencing or copy number analyses. They can also point to downstream effects on gene expression resulting from the altered signaling pathways and processes described above. To analyze the transcriptome of pancreatic cancers, we performed SAGE [serial analysis of gene expression (*28*)] on RNA from the same 24 cancers used for mutation analysis. When combined with massively parallel sequencing by synthesis, SAGE provides a highly quantitative and sensitive measure of gene expression. The approach described above is similar to that used in recent RNA-Seq studies (*29*–*32*), but SAGE has the advantage that the quantification does not depend on the length of the transcript.

As a control for the current study, we microdissected histologically normal pancreatic duct epithelial cells because these cells are the presumed precursors of pancreatic cancers. As an additional control, we used human papillomavirus (HPV)–immortalized pancreatic duct epithelial (HPDE) cells, which have been shown to have many properties in common with normal duct epithelial cells (*33*, *34*). SAGE libraries were prepared from these cells as well as the 24 pancreatic cancers; an average of 5,737,000 tags was obtained from each library, and an average of 2,268,000 tags per library matched the sequence of known transcripts (table S9).

The expression analysis was first used to help identify target genes from amplified and homozygously deleted regions. Although a small fraction of these regions contained a known tumor suppressor gene or oncogene, many contained more than one gene that had not previously been implicated in cancer. In tables S5 and S6, a presumptive target gene was identified within these regions through the use of the mutational and transcriptional data. For example, we assumed that a gene could not have been the target of an amplification event if that gene was not wholly contained within the amplicon and expressed in the tumor containing the amplification. Similarly, expression data can be used to help gauge the importance of genes containing missense mutations. A missense mutation in a gene that is not expressed in the tumor containing it is more likely to be a passenger than a mutation in a gene that is expressed (table S3).

Second, we determined whether the genes in the core signaling pathways and processes described above were differentially expressed. If the pathways and processes containing genetic alterations were indeed responsible for tumorigenesis, one might expect that many of the genes within these pathways would be aberrantly expressed. To test this hypothesis, we examined the expression of the gene sets constituting the 12 core signaling pathways and processes (Table 2 and table S8). The 31 gene sets constituting these pathways were more highly enriched for differentially expressed genes than the remaining 3041 gene sets ($P < 0.001$). These expression data thus independently support the contribution of these signaling pathways and processes to pancreatic tumorigenesis.

Lastly, we attempted to identify individual genes rather than pathways that were differentially expressed in the cancers. The data in table S9 represent the largest compendium of digital expression data derived for any tumor type to

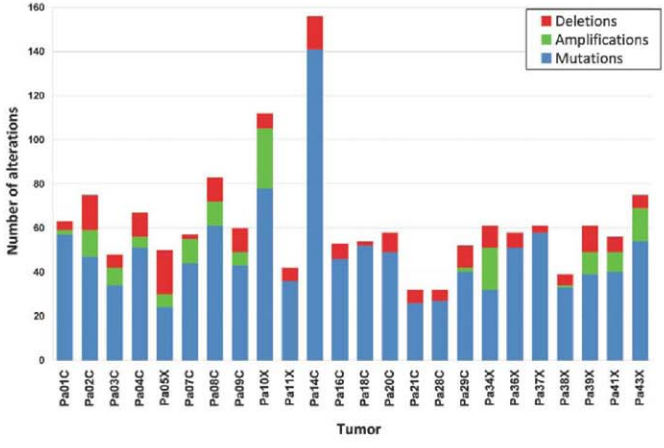

Fig. 2. Number of genetic alterations detected through sequencing and copy number analyses in each of the 24 cancers.

date. There was a remarkably high number (541) of genes that were at least 10-fold overexpressed in >90% of the 24 cancers (compared to normal pancreatic duct cells or HPDE cells). To determine whether these genes were also overexpressed in the primary tumors from which the cell lines were made, we performed SAGE on five such primary tumors. These results confirmed these 541 genes' overexpression in situ: The genes were, on average, expressed at 75-fold higher levels in the cell lines and at 88-fold higher levels in the primary tumors compared with their expression in normal duct epithelial cells. Notably, 54 of the overexpressed genes encoded proteins that are predicted to be secreted or expressed on the cell surface (table S9). These overexpressed genes provide leads for a variety of diagnostic and therapeutic approaches.

Implications for Pancreatic Tumorigenesis

The extensive genetic studies described above suggest that the key to understanding pancreatic cancers lies in an appreciation of a core set of pathways and processes. We identified 12 partially overlapping processes that are genetically altered in the great majority of pancreatic cancers (Fig. 3A). However, the pathway components that are altered in any individual tumor vary widely (Fig. 3, B and C). For example, the two tumors depicted in Fig. 3, B and C, each contain mutations of a gene involved in the TGF-β pathway (one *SMAD4*, the other *BMPR2*). Similarly, these two tumors both contain mutations of genes involved in most of the other 11 core processes and pathways, but the specific genes altered in each tumor are largely different. Although we cannot be certain that every identified mutation plays a functional role in the pathway or process in which it is implicated, it is clear from both the current and the previously published genetic data, as well as from past functional studies, that many of them are likely to affect these pathway(s).

This perspective is likely to apply to most, if not all, epithelial tumors. It is consistent with the idea that genetic alterations can be classified as mountains (high-frequency mutations) or hills (low-frequency mutations), with the hills predominating in terms of the total number of alterations involved (*14*). The heterogeneity among pathway components and the varied nature of mutations within individual genes can explain tumor heterogeneity, a fundamental facet of all solid tumors (*35*).

From an intellectual viewpoint, the pathway perspective helps bring order and rudimentary understanding to a very complex disease (*36*–*38*). Although the importance of signaling pathways in understanding neoplasia has been recognized (*39*, *40*), genomewide genetic analyses such as that described here can identify the precise genetic alterations that may be responsible for pathway dysregulation in each patient's tumor. Because most genes are mutated in only a small fraction of tumors, it is only through analysis of functional gene groups that an appreciation for the true importance of these genes' mutations in neoplasia can be reached. For example, from Table 2 it

Table 2. Core signaling pathways and processes genetically altered in most pancreatic cancers. A complete listing of the gene sets defining these signaling pathways and processes and the statistical significance of each gene set are provided in table S8.

Regulatory process or pathway	Number of genetically altered genes detected	Fraction of tumors with genetic alteration of at least one of the genes	Representative altered genes
Apoptosis	9	100%	*CASP10, VCP, CAD, HIP1*
DNA damage control	9	83%	*ERCC4, ERCC6, EP300, RANBP2, TP53*
Regulation of G_1/S phase transition	19	100%	*CDKN2A, FBXW7, CHD1, APC2*
Hedgehog signaling	19	100%	*TBX5, SOX3, LRP2, GLI1, GLI3, BOC, BMPR2, CREBBP*
Homophilic cell adhesion	30	79%	*CDH1, CDH10, CDH2, CDH7, FAT, PCDH15, PCDH17, PCDH18, PCDH9, PCDHB16, PCDHB2, PCDHGA1, PCDHGA11, PCDHGC4*
Integrin signaling	24	67%	*ITGA4, ITGA9, ITGA11, LAMA1, LAMA4, LAMA5, FN1, ILK*
c-Jun N-terminal kinase signaling	9	96%	*MAP4K3, TNF, ATF2, NFATC3*
KRAS signaling	5	100%	*KRAS, MAP2K4, RASGRP3*
Regulation of invasion	46	92%	*ADAM11, ADAM12, ADAM19, ADAM5220, ADAMTS15, DPP6, MEP1A, PCSK6, APG4A, PRSS23*
Small GTPase–dependent signaling (other than KRAS)	33	79%	*AGHGEF7, ARHGEF9, CDC42BPA, DEPDC2, PLCB3, PLCB4, RP1, PLXNB1, PRKCG*
TGF-β signaling	37	100%	*TGFBR2, BMPR2, SMAD4, SMAD3*
Wnt/Notch signaling	29	100%	*MYC, PPP2R3A, WNT9A, MAP2,TSC2, GATA6, TCF4*

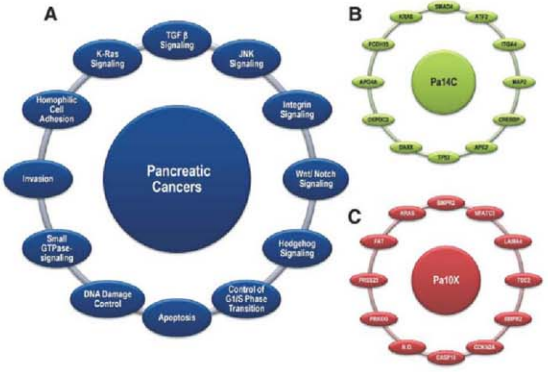

Fig. 3. Signaling pathways and processes. **(A)** The 12 pathways and processes whose component genes were genetically altered in most pancreatic cancers. (**B** and **C**) Two pancreatic cancers (Pa14C and Pa10X) and the specific genes that are mutated in them. The positions around the circles in (B) and (C) correspond to the pathways and processes in (A). Several pathway components overlapped, as illustrated by the BMPR2 mutation that presumably disrupted both the SMAD4 and Hedgehog signaling pathways in Pa10X. Additionally, not all 12 processes and pathways were altered in every pancreatic cancer, as exemplified by the fact that no mutations known to affect DNA damage control were observed in Pa10X. N.O. indicates not observed.

is evident that all pancreatic cancers studied had alterations in genes in the Wnt/Notch and Hedgehog signaling pathways, a finding that could not have been appreciated in the absence of global analyses.

In addition to yielding insights into tumor pathogenesis, such studies provide the data required for personalized cancer medicine. Unlike certain forms of leukemia, in which tumorigenesis appears to be driven by a single, targetable oncogene, pancreatic cancers result from genetic alterations of a large number of genes that function through a relatively small number of pathways and processes. Our studies suggest that the best hope for therapeutic development may lie in the discovery of agents that target the physiologic effects of the altered pathways and processes rather than their individual gene components. Thus, rather than seeking agents that target specific mutated genes, agents that broadly target downstream mediators or key nodal points may be preferable. Pathways that could be targeted include those causing metabolic disturbances, neoangiogenesis, misexpression of cell surface proteins, alterations of the cell cycle, cytoskeletal abnormalities, and an impaired ability to repair genomic damage (Table 2 and table S8).

References and Notes

1. D. M. Parkin, F. I. Bray, S. S. Devesa, *Eur. J. Cancer* **37** (suppl.), S4 (2001).
2. A. Jemal *et al.*, *CA Cancer J. Clin.* **58**, 71 (2008).
3. J. B. Koorstra, S. R. Hustinx, G. J. Offerhaus, A. Maitra, *Pancreatology* **8**, 110 (2008).
4. E. Efthimiou, T. Crnogorac-Jurcevic, N. R. Lemoine, *Pancreatology* **1**, 571 (2001).
5. M. Mimeault, R. E. Brand, A. A. Sasson, S. K. Batra, *Pancreas* **31**, 301 (2005).
6. D. A. Tuveson, S. R. Hingorani, *Cold Spring Harb. Symp. Quant. Biol.* **70**, 65 (2005).
7. E. M. Jaffee, R. H. Hruban, M. Canto, S. E. Kern, *Cancer Cell* **2**, 25 (2002).
8. A. Maitra, R. H. Hruban, *Annu. Rev. Pathol.* **3**, 157 (2008).
9. Materials and methods are available as supporting material on *Science* Online.
10. J. M. Winter, J. R. Brody, S. E. Kern, *Cancer Biol. Ther.* **5**, 360 (2006).
11. B. Rubio-Viqueira *et al.*, *Clin. Cancer Res.* **12**, 4652 (2006).
12. S. Jones *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 4283 (2008).
13. T. Sjöblom *et al.*, *Science* **314**, 268 (2006).
14. L. D. Wood *et al.*, *Science* **318**, 1108 (2007).
15. D. W. Parsons *et al.*, *Science* **321**, 1807 (2008).
16. A. Hartmann, H. Blaszyk, J. S. Kovach, S. S. Sommer, *Trends Genet.* **13**, 27 (1997).
17. S. P. Hussain, C. C. Harris, *Mutat. Res.* **428**, 23 (1999).
18. P. C. Ng, S. Henikoff, *Nucleic Acids Res.* **31**, 3812 (2003).
19. R. Karchin, Structural models of mutants identified in pancreatic cancers, http://karchinlab.org/Mutants/CAN-genes/pancreatic/Pancreatic_cancer.html (2008).
20. W. M. Klein, R. H. Hruban, A. J. Klein-Szanto, R. E. Wilentz, *Mod. Pathol.* **15**, 441 (2002).
21. H.-P. Elsasser, G. Adler, H. F. Kern, in *The Pancreas*, V. L. W. Go *et al.*, Eds. (Raven, New York, 1993), pp. 75–86.
22. R. J. Leary *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
23. A. G. Knudson, *Am. J. Med. Genet.* **111**, 96 (2002).
24. S. R. Hustinx *et al.*, *Mod. Pathol.* **18**, 959 (2005).
25. A. Varshavsky, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 14935 (2007).
26. C. Greenman, R. Wooster, P. A. Futreal, M. R. Stratton, D. F. Easton, *Genetics* **173**, 2187 (2006).
27. S. Ekins, Y. Nikolsky, A. Bugrim, E. Kirillov, T. Nikolskaya, *Methods Mol. Biol.* **356**, 319 (2007).
28. V. E. Velculescu, L. Zhang, B. Vogelstein, K. W. Kinzler, *Science* **270**, 484 (1995).
29. M. Sultan *et al.*, *Science* **321**, 956 (2008).
30. R. Lister *et al.*, *Cell* **133**, 523 (2008).
31. A. Mortazavi, B. A. Williams, K. McCue, L. Schaeffer, B. Wold, *Nat. Methods* **5**, 621 (2008).
32. R. Morin *et al.*, *Biotechniques* **45**, 81 (2008).
33. T. Furukawa *et al.*, *Am. J. Pathol.* **148**, 1763 (1996).
34. H. Ouyang *et al.*, *Am. J. Pathol.* **157**, 1623 (2000).
35. A. H. Owens, D. S. Coffey, S. B. Baylin, *Tumor Cell Heterogeneity* (Academic Press, New York, 1982).
36. J. Lin *et al.*, *Genome Res.* **17**, 1304 (2007).
37. T. W. Chittenden *et al.*, *Genomics* **91**, 508 (2008).
38. E. J. Edelman, J. Guinney, J. Chi, P. Febbo, S. Mukherjee, *PLoS Comput. Biol.* **4**, e28 (2008).
39. D. Hanahan, R. A. Weinberg, *Cell* **100**, 57 (2000).
40. B. Vogelstein, K. W. Kinzler, *Nat. Med.* **10**, 789 (2004).
41. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
42. E. F. Pettersen *et al.*, *J. Comput. Chem.* **25**, 1605 (2004).

43. Contact information for the authors who directed the major components of this project is as follows: J. R. Eshleman, S. E. Kern (clinical and sample coordination), jeshlem@jhmi.edu, sk@mail.jhmi.edu; R. H. Hruban (pathological review), rhruban@jhmi.edu; R. Karchin (bioinformatic analysis), karchin@jhu.edu; N. Papadopoulos (gene expression analysis), npapadop1@jhmi.edu; G. Parmigiani (statistical analysis), gp@jhu.edu; B. Vogelstein, V. E. Velculescu, K. W. Kinzler (sequencing and copy number analysis), bertvog@gmail.com, velculescu@jhmi.edu, kinzler@jhmi.edu.

We thank J. Ptak, N. Silliman, L. Dobbyn, M. Whalen, M. Borges, G. Cusatis, M. Griffith, C. Henderson, C. Karikari, G. Mo, M. Mullendore, E. Palmisano, M. Raben, S. Solt, and D. Trusty for expert technical assistance; T. Sjöblom for help with database management for mutational analyses; Y. Ding for help with bioinformatic analyses; and J. Cameron, G. Feldmann, S. Salaria, L. Hua, and C. Yeo for helpful discussions. This project was performed under the auspices of the Goldman Pancreatic Cancer Genome Initiative and was funded by the Sol Goldman Charitable Trust; the Lillian Goldman Charitable Trust; the Lustgarten Foundation for Pancreatic Cancer Research; the Virginia and D. K. Ludwig Fund for Cancer Research; the Susan G. Komen Foundation; the Michael Rolfe Pancreatic Cancer Foundation; the Joseph C. Monastra Foundation for Pancreatic Cancer; the family and friends of George Rubis; the Viragh Family Foundation; the Broad Foundation; the Emerald Foundation; NIH grants CA62924, CA43460, CA57345, and CA121113; and Beckman Coulter Corporation. Under separate licensing agreements between the Johns Hopkins University and Genzyme, Beckman Coulter, and Exact Sciences Corporations, K.W.K., B.V., and V.E.V. are entitled to a share of royalties received by the university on sales of products related to research described in this paper. These authors and the university own Genzyme and Exact Sciences stock, which is subject to certain restrictions under university policy. The terms of these arrangements are managed by the Johns Hopkins University in accordance with its conflict-of-interest policies.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1164368/DC1
Materials and Methods
Fig. S1
Tables S1 to S9
References

7 August 2008; accepted 27 August 2008
Published online 4 September 2008;
10.1126/science.1164368
Include this information when citing this paper.

An Integrated Genomic Analysis of Human Glioblastoma Multiforme

D. Williams Parsons,[1,2]* Siân Jones,[1]* Xiaosong Zhang,[1]* Jimmy Cheng-Ho Lin,[1]* Rebecca J. Leary,[1]* Philipp Angenendt,[1]* Parminder Mankoo,[3] Hannah Carter,[3] I-Mei Siu,[4] Gary L. Gallia,[4] Alessandro Olivi,[4] Roger McLendon,[5] B. Ahmed Rasheed,[5] Stephen Keir,[5] Tatiana Nikolskaya,[6] Yuri Nikolsky,[7] Dana A. Busam,[8] Hanna Tekleab,[8] Luis A. Diaz Jr.,[1] James Hartigan,[9] Doug R. Smith,[9] Robert L. Strausberg,[8] Suely Kazue Nagahashi Marie,[10] Sueli Mieko Oba Shinjo,[10] Hai Yan,[5] Gregory J. Riggins,[4] Darell D. Bigner,[5] Rachel Karchin,[3] Nick Papadopoulos,[1] Giovanni Parmigiani,[1] Bert Vogelstein,[1]† Victor E. Velculescu,[1]† Kenneth W. Kinzler[1]†

Glioblastoma multiforme (GBM) is the most common and lethal type of brain cancer. To identify the genetic alterations in GBMs, we sequenced 20,661 protein coding genes, determined the presence of amplifications and deletions using high-density oligonucleotide arrays, and performed gene expression analyses using next-generation sequencing technologies in 22 human tumor samples. This comprehensive analysis led to the discovery of a variety of genes that were not known to be altered in GBMs. Most notably, we found recurrent mutations in the active site of isocitrate dehydrogenase 1 (*IDH1*) in 12% of GBM patients. Mutations in *IDH1* occurred in a large fraction of young patients and in most patients with secondary GBMs and were associated with an increase in overall survival. These studies demonstrate the value of unbiased genomic analyses in the characterization of human brain cancer and identify a potentially useful genetic alteration for the classification and targeted therapy of GBMs.

Malignant gliomas are the most frequent and lethal cancers originating in the central nervous system. The most biologically aggressive subtype is glioblastoma multiforme (GBM) [World Health Organization (WHO) grade IV astrocytoma], a tumor associated with a dismal prognosis (*1*). The current standard of care for GBM patients—surgical resection followed by adjuvant radiation therapy and chemotherapy with the oral alkylating agent temozolomide—produces a median survival of only 15 months (*2*). Historically, GBMs have been categorized into two groups ("primary" and "secondary") on the basis of clinical presentation (*3*). Secondary GBMs are defined as cancers that have clinical, radiologic, or histopathologic evidence of malignant progression from a preexisting lower-grade tumor, whereas primary GBMs have no such history and present at diagnosis as advanced cancers (*4*). Clinical differences have been reported between the two groups, with secondary GBMs occurring less frequently (~5% of GBMs) and predominantly in younger patients (median age ~45 years versus ~60 years for primary GBM) (*5*, *6*). The histopathologic findings of primary and secondary GBMs are indistinguishable, and the prognosis does not appear to be different after adjustment for age (*5*, *6*).

Substantial research effort has focused on the identification of genetic alterations in GBMs that might help define subclasses of GBM patients with differing prognoses and/or response to specific therapies (*7*). Distinctions between the genetic lesions found in primary and secondary GBMs have been made, with *TP53* mutations occurring more commonly in secondary GBMs and *EGFR* amplifications and *PTEN* mutations occurring more frequently in primary GBMs (*6*, *8*, *9*); however, none of these alterations is sufficiently specific to distinguish between primary and secondary GBMs. This issue is further confounded by the possibility that a fraction of GBMs designated as primary tumors may follow a sequence of genetic events similar to that of secondary lesions but not come to clinical attention until malignant progression to a GBM has occurred.

The comprehensive elucidation of genetic alterations in GBMs could provide novel targets that might be used for diagnostic, prognostic, or therapeutic purposes as well as to identify subgroups of patients that preferentially respond to particular targeted therapies. The determination of the human genome sequence and improvements in sequencing and bioinformatic technologies have recently permitted genome-wide sequence analyses in human cancers. We have previously studied the genomes of 11 breast and 11 colorectal cancers by determining the sequence of the more than 18,000 Consensus Coding Sequence (CCDS) and Reference Sequence (RefSeq) genes (*10*, *11*). Here, we have analyzed 20,661 protein coding genes in 22 human GBM samples. To complement these sequencing data, we have also performed a genome-wide analysis of focal copy number alterations, including amplifications and homozygous deletions, using high-density oligonucleotide microarrays on the same GBM tumors. Finally, we have examined the expression profiles of these same samples using serial analysis of gene expression (SAGE) and next-generation sequencing technologies.

Sequencing strategy. We extended our previous sequencing strategy for identification of somatic mutations to include 23,219 transcripts from 20,661 genes (fig. S1). These included 2783 additional genes from the Ensembl databases that were not present in the CCDS or RefSeq databases analyzed in the previous studies (*10*, *11*). In addition, we redesigned polymerase chain reaction (PCR) primers for regions of the genome that (i) were difficult to PCR amplify in previous studies or (ii) were found to share substantial identity with other human or mouse sequences. The combination of these new, redesigned, and existing primers sequences resulted in a total of 208,311 primer pairs (table S1)

[1]Ludwig Center for Cancer Genetics and Therapeutics, and Howard Hughes Medical Institute at Johns Hopkins Kimmel Cancer Center, Baltimore, MD 21231, USA. [2]Department of Pediatrics, Section of Hematology-Oncology, Baylor College of Medicine, Houston TX 77030, USA. [3]Department of Biomedical Engineering, Institute of Computational Medicine, Johns Hopkins Medical Institutions, Baltimore, MD 21218, USA. [4]Department of Neurosurgery, Johns Hopkins Medical Institutions, Baltimore, MD 21231, USA. [5]Department of Pathology, Pediatric Brain Tumor Foundation, and Preston Robert Tisch Brain Tumor Center at Duke University Medical Center, Durham, NC 27710, USA. [6]Vavilov Institute for General Genetics, Moscow 8333, 117809, Russia. [7]GeneGo, Inc., St. Joseph, MI 49085, USA. [8]J. Craig Venter Institute, Rockville, MD 20850, USA. [9]Agencourt Bioscience Corporation, Beverly, MA 01915, USA. [10]Department of Neurology, School of Medicine, University of São Paulo, São Paulo, Brazil.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: bertvog@gmail.com (B.V.); velculescu@jhmi.edu (V.E.V.); kinzlke@jhmi.edu (K.W.K.)

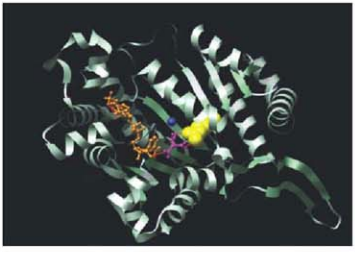

Fig. 1. Structure of the active site of IDH1. The crystal structure of the human cytosolic NADP(+)–dependent IDH is shown in ribbon format (PDBID: 1TOL) (*44*). The active cleft of IDH1 consists of a NADP-binding site and the isocitrate-metal ion-binding site. The alpha-carboxylate oxygen and the hydroxyl group of isocitrate chelate the Ca^{2+} ion. NADP is colored in orange, isocitrate in purple and Ca^{2+} in blue. The Arg^{132} residue, displayed in yellow, forms hydrophilic interactions, shown in red, with the alpha-carboxylate of isocitrate. Displayed image was created with UCSF Chimera software version 1.2422 (*50*).

that were successfully used for sequence analysis of the coding exons of these genes.

Twenty-two GBM samples (table S2) were selected for PCR sequence analysis, consisting of 7 samples extracted directly from patient tumors and 15 samples passaged in nude mice as xenografts. In the first stage of this analysis, called the Discovery Screen, the primer pairs were used to amplify and sequence 175,471 coding exons and adjacent intronic splice donor and acceptor sequences in the 22 GBM samples and in one matched normal sample. The data were assembled for each amplified region and evaluated using stringent quality criteria (*12*), resulting in successful amplification and sequencing of 95.0% of targeted amplicons in the 22 tumors (Table 1). A total of 689 Mb of sequence data was generated in this fashion. The amplicon traces were analyzed using automated approaches to identify changes in the tumor sequences that were not present in the reference sequences of each gene. Alterations present in the normal control sample and in single nucleotide polymorphism (SNP) databases were then removed from further analyses. The remaining sequence traces of potential alterations were visually inspected to remove false-positive mutation calls generated by the automated software. All exons containing putative mutations were then reamplified and sequenced in both the affected tumor and the matched normal DNA sample. This process allowed us to confirm the presence of the mutation in the tumor sample and determine whether the alteration was somatic (i.e., tumor-specific) or was present in the germline. All putative somatic mutations were examined computationally and experimentally to confirm that the alterations did not arise through the aberrant coamplification of related gene sequences (*12*).

Analysis of sequence alterations. Analysis of the identified somatic mutations revealed that one tumor (Br27P), from a patient previously treated with radiation therapy and temozolomide, had 17 times as many alterations as any of the other 21 patients (table S3). The mutation spectrum of this sample was also dramatically different from those of the other GBM patients (*12*) and was consistent with previous observations of a hypermutation phenotype in glioma samples of patients treated with temozolomide (*13*, *14*). After removing Br27P from consideration, we found that 685 genes (3.3% of the 20,661 genes analyzed) contained at least one nonsilent somatic mutation. The vast majority of these alterations were single-base substitutions (94%), whereas the others were small insertions, deletions, or duplications (Table 1). The 993 somatic mutations were observed to be distributed relatively evenly among the 21 remaining tumors (table S3), with a mean of 47 mutations per tumor, representing 1.51 mutations per Mb of GBM tumor genome sequenced. The six DNA samples extracted directly from patient tumors had smaller numbers of mutations than those obtained from xenografts, likely because of the masking effect of nonneoplastic cells in the former. It has previously been shown that cell lines and xenografts provide the optimal template DNA for cancer genome sequencing analyses (*15*) and that they faithfully represent the alterations present in the original tumors (*16*). Both the total number and the frequency of sequence alterations in GBMs were substantially smaller than the number and frequency of such alterations observed in colorectal or breast cancers and slightly less than in pancreatic cancers (*10*, *11*, *17*). The most likely explanation for this difference is the reduced number of cell generations in glial cells before the onset of neoplasia (*18*).

We further evaluated a set of 21 mutated genes identified in the Discovery Screen in a second screen, called a Prevalence Screen, comprising an additional 83 GBMs with well-documented clinical histories (table S2). The 21 genes selected were mutated in at least two Discovery Screen tumors and had mutation frequencies of >10 mutations per Mb of tumor DNA sequenced. Nonsilent somatic mutations were identified in 16 of these 21 genes in the additional tumor samples (table S4). The mutation frequency of all analyzed genes in the Prevalence Screen was 23 mutations per Mb of tumor DNA, markedly increased from the overall mutation frequency in the Discovery Screen of 1.5 mutations per Mb ($P < 0.001$). Additionally, the observed ratio of nonsilent to silent mutations among mutations in the Prevalence Screen was 14.5:1, substantially higher than the 3.1:1 ratio that was observed in the Discovery Screen ($P < 0.001$). The increased mutation frequency and higher fraction of nonsilent mutations suggested that genes mutated in the Prevalence Screen were enriched for genes that actively contributed to tumorigenesis.

In addition to the frequency of mutations in a gene, the type of mutation can provide information useful for evaluating its potential role in disease (*19*). The likely effect of missense mutations can be assessed through evaluation of the mutated residue by evolutionary or structural means. To evaluate missense mutations, we de-

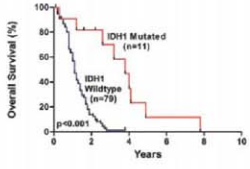

Fig. 2. Overall survival according to *IDH1* mutation status. The hazard ratio for death among patients with wild-type *IDH1* ($n = 79$), as compared to those with mutant *IDH1* ($n = 11$), was 3.7 (95 percent confidence interval, 2.1 to 6.5; $P < 0.001$). The median survival was 3.8 years for patients with mutated *IDH1*, as compared to 1.1 years for patients with wild-type *IDH1*.

Table 1. Summary of genomic analyses.

Sequencing analysis	
Number of amplicons attempted	219,229 (100%)
Number of amplicons passing quality control*	208,311 (95%)
Fraction of bases in passing amplicons with PHRED > 20	98.3%
Number of genes analyzed	20,661
Number of transcripts analyzed	23,219
Number of exons analyzed	175,471
Total number of nucleotides successfully sequenced	689,071,123
Number of somatic mutations identified ($n = 22$ samples)	2,325
Number of somatic mutations (excluding Br27P)	993
Missense	622
Nonsense	43
Insertion	3
Deletion	46
Duplication	7
Splice site or UTR	27
Synonymous	245
Average number of sequence alterations per sample	47.3
Copy number analysis	
Total number of SNP loci assessed for copy number changes	1,069,688
Number of copy number alterations identified ($n = 22$ samples)	281
Amplifications	147
Homozygous deletions	134
Average number of amplifications per sample	6.7
Average number of homozygous deletions per sample	6.1

*Passing amplicons were defined as having PHRED20 scores or better over 90% of the target sequence in 75% of samples analyzed [see (*12*) for additional information].

veloped an algorithm (LS-MUT) that employs machine learning of 58 predictive features based on evolutionary conservation and the physical-chemical properties of amino acids involved in the alteration (*12*). About 15% of the missense mutations evaluated were predicted to have a statistically significant effect on protein function when assessed by this method (table S3). We also were able to make structural models of 244 of the 870 missense mutations identified in this study (*20*). In each case, the model was based on x-ray crystallography or nuclear magnetic resonance spectroscopy of the normal protein or a closely related homolog. This analysis showed that 35 of the missense mutations are located close to a domain interface or substrate-binding site and thus are likely to affect protein function [links to structural models are available in (*12*)].

Analysis of copy number changes. The same tumors were then evaluated for copy number alterations through genomic hybridization of DNA samples to Illumina SNP arrays containing ~1 million probes (*21*). We have recently developed a sensitive and specific approach for the identification of focal amplifications resulting in 12 or more copies per nucleus (amplification by a factor of 6 or more compared with the diploid genome) as well as deletions of both copies of a gene (homozygous deletions) using such arrays (*22*). Unlike larger chromosomal aberrations, such focused alterations can be used to identify underlying candidate genes in these regions.

We identified a total of 147 amplifications (table S5) and 134 homozygous deletions (table S6) in the 22 samples used in the Discovery Screen (Table 1). Although the number of amplifications was similar in samples extracted from patient tumors and those that had been passaged as xenografts, the latter samples allowed detection of a larger number of homozygous deletions (average of 8.0 deletions per sample in the xenografts versus 2.2 per sample in the tumors). These observations are consistent with previous reports that document the difficulty of identifying homozygous deletions in samples containing contaminating normal DNA (*23*) and highlight the importance of using purified human tumor cells, such as those present in xenografts or cell lines, for genomic analyses.

Integration of sequencing, copy number, and expression analyses. Mutations that arise during tumorigenesis may provide a selective advantage to the tumor cell (driver mutations) or have no net effect on tumor growth (passenger mutations). The mutational data obtained from sequencing and analysis of copy number alterations were integrated to identify GBM candidate cancer genes (*CAN*-genes) that are most likely to be drivers and therefore worthy of further investigation. To determine whether a gene was likely to harbor driver mutations, we compared the number and type of mutations observed (including sequence changes, amplifications, and homozygous deletions) and determined the probability that these alterations would result from passenger mutation rates alone (*12*) (fig. S1).

The *CAN*-genes, together with their passenger probabilities, are listed in table S7. The *CAN*-genes included several with established roles in gliomas, including *TP53*, *PTEN*, *CDKN2A*, *RB1*, *EGFR*, *NF1*, *PIK3CA*, and *PIK3R1* (*24*–*34*). Of these genes, the most frequently altered were *CDKN2A* (altered in 50% of GBMs); *TP53*, *EGFR*, and *PTEN* (altered in 30 to 40%); *NF1*, *CDK4*, and *RB1* (altered in 12 to 15%); and *PIK3CA* and *PIK3R1* (altered in 8 to 10%) (Table 2). Overall, these frequencies, which are similar to or in some cases higher than those previously reported, validate the sensitivity of our approach for detecting somatic alterations.

Through analysis of additional gene members within cell signaling pathways affected by these genes, we identified alterations of critical genes in the *TP53* pathway (*TP53*, *MDM2*, and *MDM4*), the *RB1* pathway (*RB1*, *CDK4*, and *CDKN2A*), and the *PI3K/PTEN* pathway (*PIK3CA*, *PIK3R1*, *PTEN*, and *IRS1*). These alterations affected pathways in a majority of tumors (64%, 68%, and 50%, respectively), and in all cases but one, mutations within each tumor affected only a single member of each pathway in a mutually exclusive manner ($P < 0.05$) (Table 3).

Systematic analyses of functional gene groups and pathways contained within the well-annotated MetaCore database (*35*) identified enrichment of alterations in a variety of cellular processes in GBMs, including additional members of the *TP53* and *PI3K/PTEN* pathways. Many of the pathways identified were similar to core signaling pathways found to be altered in pancreas, colorectal, and breast tumors, such as those regulating control of cellular growth, apoptosis, and cell adhesion (*17*, *22*, *36*). However, several pathways were enriched only in GBMs. These included channels involved in transport of sodium, potassium, and calcium ions, as well as nervous system–specific cellular pathways such as synaptic transmission, transmission of nerve impulses, and axonal guidance (table S8). Mutations in these latter pathways may represent a subversion of normal glial cell processes to promote dysregulated growth and invasion.

Gene expression patterns can inform the analysis of pathways because they can reflect epigenetic alterations not detectable by sequencing or copy number analyses. To analyze the transcriptome of GBMs, we performed SAGE (*37*, *38*) on all GBM samples for which sufficient RNA was available (total of 16 samples), as well as on two independent normal brain RNA controls (table S9). When combined with sequencing-by-synthesis methods (*39*–*42*), SAGE provides a highly quantitative and sensitive measure of gene expression. We first used the transcript analysis to help identify previously uncharacterized target genes from the amplified and deleted regions that were revealed by our study. In tables S5 and S6, a candidate target gene could be identified within several of these regions

Table 2. Most frequently altered GBM *CAN*-genes. All *CAN*-genes are listed in table S7.

Gene	Point mutations*		Amplifications†		Homozygous deletions†		Fraction of tumors with any alteration (%)	Passenger probability‡
	No. of tumors	Fraction of tumors (%)	No. of tumors	Fraction of tumors (%)	No. of tumors	Fraction of tumors (%)		
CDKN2A	0/22	0	0/22	0	11/22	50	50	<0.01
TP53	37/105	35	0/22	0	1/22	5	40	<0.01
EGFR	15/105	14	5/22	23	0/22	0	37	<0.01
PTEN	27/105	26	0/22	0	1/22	5	30	<0.01
NF1	16/105	15	0/22	0	0/22	0	15	0.04
CDK4	0/22	0	3/22	14	0/22	0	14	<0.01
RB1	8/105	8	0/22	0	1/22	5	12	0.02
IDH1	12/105	11	0/22	0	0/22	0	11	<0.01
PIK3CA	10/105	10	0/22	0	0/22	0	10	0.10
PIK3R1	8/105	8	0/22	0	0/22	0	8	0.10

*Fraction of tumors with point mutations indicates the fraction of mutated GBMs out of the 105 samples in the Discovery and Prevalence Screens. CDKN2A and CDK4 were not analyzed for point mutations in the Prevalence Screen because no sequence alterations were detected in these genes in the Discovery Screen. †Fraction of tumors with amplifications and deletions indicates the number of tumors with these types of alterations in the 22 Discovery Screen samples. ‡Passenger probability indicates the probability obtained using the average of the lower and upper bound background mutation rates (*12*).

through the use of the mutational as well as transcriptional data. Second, we used the transcript analysis to help identify genes that were differentially expressed in GBMs compared to normal brain. A large number of genes (143) were expressed on average at levels 10 times as high in the GBMs. Among the overexpressed genes, 16 encoded proteins that are predicted to be secreted or expressed on the cell surface, suggesting new opportunities for diagnostic and therapeutic applications. Third, we used expression data to help assess the significance of genes containing missense mutations (table S3). Finally, we assessed whether the gene sets implicated in the pathways enriched for genetic alterations were also altered through expression changes. Notably, the gene sets in these pathways were more highly enriched for differentially expressed genes than the remaining sets ($P < 0.001$) (*12*). These expression data thus independently highlight the potential importance of these pathways in the development of GBMs.

High-frequency alterations of IDH1 in GBM. The *CAN*-gene list (table S7) included a number of individual genes that had not previously been linked to GBMs. The most frequently mutated of these genes, *IDH1* on chromosome 2q33, encodes isocitrate dehydrogenase 1, which catalyzes the oxidative carboxylation of isocitrate to α-ketoglutarate, resulting in the production of nicotinamide adenine dinucleotide phosphate (NADPH). Of the five isocitrate dehydrogenase proteins encoded in the human genome, at least three are localized to the mitochondria, while IDH1 is localized within the cytoplasm and peroxisomes (*43*). The IDH1 protein forms an asymmetric homodimer (*44*) and is thought to play a substantial role in cellular control of oxidative damage through generation of NADPH (*45*, *46*). None of the other IDH genes were found to be genetically altered in our analysis.

IDH1 was somatically mutated in 5 of the 22 GBM tumors in the Discovery Screen. Surprisingly, all 5 had the same heterozygous point mutation, a change of a guanine to an adenine at position 395 of the *IDH1* transcript (G395A), leading to the replacement of an arginine with a histidine at amino acid residue 132 of the protein (R132H). In our previous study of colorectal cancers, this same codon was mutated in a single case through alteration of the adjacent nucleotide, resulting in a R132C amino acid change (*10*). Five GBMs evaluated in our Prevalence Screen were found to have heterozygous somatic R132H mutations, and an additional two tumors had a third distinct somatic mutation affecting the same amino acid residue, R132S (fig. S2 and Table 4). In addition to the Discovery and Prevalence Screen samples, 44 other GBMs were analyzed for *IDH1* mutations, revealing six tumors with somatic mutations affecting R132. In total, 18 of 149 GBMs (12%) analyzed had alterations in *IDH1*. The R132 residue is conserved in all known species and is localized to the substrate binding site, where it forms hydrophilic interactions with the alpha-carboxylate of isocitrate (Fig. 1) (*44*, *47*).

Several important observations were made about *IDH1* mutations and their potential clinical importance. First, mutations in *IDH1* preferentially occurred in younger GBM patients, with a mean age of 33 years for *IDH1*-mutated patients, as opposed to 53 years for patients with wild-type *IDH1* ($P < 0.001$, t test) (Table 4). In patients under 35 years of age, nearly 50% (9 of 19) had mutations in *IDH1*. Second, mutations in *IDH1* were found in nearly all of the patients with secondary GBMs (mutations in 5 of 6 secondary GBM patients, as compared to 7 of 99 patients with primary GBMs) ($P < 0.001$, binomial test). Third, patients with *IDH1* mutations had a significantly improved prognosis, with a median overall survival of 3.8 years as compared to 1.1 years for patients with wild-type *IDH1* (Fig. 2) ($P < 0.001$, log-rank test). Although both younger age and mutated *TP53* are known to be positive prognostic factors for GBM patients, this associ-

Table 3. Mutations of the TP53, PI3K, and RB1 pathways in GBM samples. Mut, mutated; Amp, amplified; Del, deleted; Alt, altered.

	TP53 pathway				PI3K pathway					RB1 pathway			
Tumor sample	TP53	MDM2	MDM4	All genes	PTEN	PIK3CA	PIK3R1	IRS1	All genes	RB1	CDK4	CDKN2A	All genes
Br02X	Del			Alt				Mut	Alt			Del	Alt
Br03X	Mut			Alt	Mut				Alt				
Br04X	Mut			Alt	Mut				Alt	Mut			Alt
Br05X			Amp	Alt		Mut			Alt			Del	Alt
Br06X												Del	Alt
Br07X	Mut			Alt	Mut				Alt	Del			Alt
Br08X												Del	Alt
Br09P	Mut			Alt							Amp		Alt
Br10P	Mut			Alt									
Br11P	Mut			Alt									
Br12P	Mut			Alt			Mut		Alt				
Br13X	Mut			Alt								Del	Alt
Br14X							Mut		Alt			Del	Alt
Br15X										Mut		Del	Alt
Br16X		Amp		Alt							Amp		Alt
Br17X					Mut				Alt			Del	Alt
Br20P													
Br23X	Mut			Alt	Del				Alt				
Br25X					Mut				Alt			Del	Alt
Br26X						Mut			Alt			Del	Alt
Br27P	Mut			Alt							Amp		Alt
Br29P	Mut			Alt									
Fraction of tumors with altered gene/ pathway*	0.55	0.05	0.05	0.64	0.27	0.09	0.09	0.05	0.50	0.14	0.14	0.45	0.68

*Fraction of affected tumors in 22 Discovery Screen samples.

ation between *IDH1* mutation and improved survival was noted even in the subgroup of young patients with *TP53* mutations ($P < 0.02$, log-rank test).

Discussion. The data resulting from this integrated analysis of mutations and copy number alterations have provided a novel view of the genetic landscape of glioblastomas. Like all large-scale genetic analyses, our study has limitations. We did not assess certain molecular alterations, including chromosomal translocations and epigenetic changes. However, our large-scale expression studies should have identified any genes that were differentially expressed through these mechanisms (table S9). Additionally, we focused on copy number changes that were focal amplifications or homozygous deletions, because these have historically been most useful in identifying cancer genes. The array data we have generated can also be analyzed to determine loss of heterozygosity (LOH) or low-amplitude regions of copy number gains, but such changes cannot generally be used to pinpoint new candidate cancer genes. Finally, the samples directly extracted from patient tumors contained small amounts of contaminating normal tissue, which limited our ability to detect homozygous deletions and, to a lesser extent, somatic mutations, in those specific tumors.

Despite these limitations, our study provides a number of important genetic and clinical insights into GBMs. First, it revealed that some of the pathways known to be altered in GBMs affect a larger fraction of genes and patients than previously anticipated. A majority of the tumors analyzed had alterations in genes encoding components of each of the *TP53*, *RB1*, and *PI3K* pathways. The fact that all but one of the cancers with mutations in members of a pathway did not have alterations in other members of the same pathway suggests that such alterations are functionally equivalent in tumorigenesis. Second, these results have identified a variety of new genes and signaling pathways not previously implicated in GBMs (table S7 and S8). Some of these pathways were found to be altered in previous genome-wide analyses of pancreatic, breast, and colorectal cancers and may represent core processes that underlie human tumorigenesis (*17*, *22*, *36*). A number of the signaling pathways mutated or altered through expression differences in GBMs appear to be involved in nervous system signaling processes and represent novel and potentially useful aspects of GBM biology.

The comprehensive nature of our study allowed us to identify *IDH1* as an unexpected target of genetic alteration in patients with GBM. All mutations in this gene resulted in amino acid substitutions at position 132, an evolutionarily conserved residue located within the isocitrate binding site (*44*). The recurrent nature of the mutations is reminiscent of activating alterations in oncogenes such as *BRAF*, *KRAS*, and *PIK3CA*. Our speculation that this sequence change is an activating mutation is strengthened by the absence of inactivating changes (e.g., frameshift or stop mutations), the absence of other alterations in key residues of the active site, and the fact that all mutations observed to date were heterozygous (without any evidence of loss of the second allele through LOH). Interestingly, enzymatic studies have shown that in vitro engineered substitution of arginine at residue 132 with a different amino acid (glutamate) than that observed in patients results in a catalytically inactive enzyme, suggesting a critical role for this residue (*48*). Further biochemical and molecular analyses will be needed to determine the effect of alterations of IDH1 on enzymatic activity and cellular phenotype.

Regardless of the specific molecular consequences of *IDH1* alterations, detection of mutations in *IDH1* is likely to be clinically useful. Although considerable effort has focused on the identification of characteristic genetic lesions in primary and secondary GBMs, the altered genes identified to date are not optimal for this purpose (*5*). Our study revealed IDH1 mutation to be a novel and potentially more specific marker for secondary GBM. One hypothesis is that *IDH1* alterations identify a biologically specific subgroup of GBM patients, including both patients who would be classified as having secondary GBMs and a subpopulation of primary GBM patients with a similar tumor biology and a more protracted clinical course (Table 4). Interestingly, patients with *IDH1* mutations had a very high frequency of *TP53* mutation and a very low frequency of mutations in other commonly altered GBM genes (Table 4). Patients with mutated

Table 4. Characteristics of GBM patients with IDH1 mutations

Patient ID	Patient age (years)*	Sex	Recurrent GBM†	Secondary GBM‡	Overall survival (years)§	IDH1 mutation		Mutation of TP53	Mutation of PTEN, RB1, EGFR, or NF1
						Nucleotide	Amino acid		
Br10P	30	F	No	No	2.2	G395A	R132H	Yes	No
Br11P	32	M	No	No	4.1	G395A	R132H	Yes	No
Br12P	31	M	No	No	1.6	G395A	R132H	Yes	No
Br104X	29	F	No	No	4.0	C394A	R132S	Yes	No
Br106X	36	M	No	No	3.8	G395A	R132H	Yes	No
Br122X	53	M	No	No	7.8	G395A	R132H	No	No
Br123X	34	M	No	Yes	4.9	G395A	R132H	Yes	No
Br237T	26	M	No	Yes	2.6	G395A	R132H	Yes	No
Br211T	28	F	No	Yes	0.3	G395A	R132H	Yes	No
Br27P	32	M	Yes	Yes	1.2	G395A	R132H	Yes	No
Br129X	25	M	Yes	Yes	3.2	C394A	R132S	No	No
Br29P	42	F	Yes	Unknown	Unknown	G395A	R132H	Yes	No
IDH1 mutant patients (n=12)	33.2	67% M	25%	42%	3.8	100%	100%	83%	0%
IDH1 wild-type patients (n=93)	53.3	65% M	16%	1%	1.1	0%	0%	27%	60%

*Patient age refers to age at which the sample was obtained. †Recurrent GBM designates a GBM which was resected >3 months after a prior diagnosis of GBM. ‡Secondary GBM designates a GBM which was resected > 1 year after a prior diagnosis of a lower grade glioma (WHO I-III). §Overall survival was calculated using date of GBM diagnosis and date of death or last patient contact: Patients Br10P and Br11P were alive at last contact. Median survival for IDH1 mutant patients and IDH1 wild-type patients was calculated using logrank test. Previous pathologic diagnoses in secondary GBM patients were oligodendroglioma (WHO grade II) in Br123X, low grade glioma (WHO grade I-II) in Br237T and Br211T, anaplastic astrocytoma (WHO grade III) in Br27P, and anaplastic oligodendroglioma (WHO grade III) in Br129X. Mean age and median survival are listed for the groups of IDH1-mutated and IDH1-wild-type patients.

IDH1 also had distinct clinical characteristics, including younger age and a considerably improved clinical prognosis (Table 4). It is conceivable that new treatments could be designed to take advantage of IDH1 alterations in these patients, because inhibition of a different IDH enzyme (IDH2) has recently been shown to result in increased sensitivity of tumor cells to a variety of chemotherapeutic agents (*49*). In summary, the discovery of IDH1 and other genes previously not known to play a role in human tumors (table S7) validates the utility of genome-wide genetic analysis of tumors in general and opens new avenues of basic and clinical brain tumor research.

References and Notes

1. D. N. Louis *et al.*, *Acta Neuropathol.* **114**, 97 (2007).
2. R. Stupp *et al.*, *N. Engl. J. Med.* **352**, 987 (2005).
3. H. Scherer, *Am. J. Cancer* **40**, 159 (1940).
4. P. Kleihues, H. Ohgaki, *Neuro-oncol.* **1**, 44 (1999).
5. H. Ohgaki, P. Kleihues, *Am. J. Pathol.* **170**, 1445 (2007).
6. H. Ohgaki *et al.*, *Cancer Res.* **64**, 6892 (2004).
7. I. K. Mellinghoff *et al.*, *N. Engl. J. Med.* **353**, 2012 (2005).
8. E. A. Maher *et al.*, *Cancer Res.* **66**, 11502 (2006).
9. C. L. Tso *et al.*, *Cancer Res.* **66**, 159 (2006).
10. T. Sjöblom *et al.*, *Science* **314**, 268 (2006).
11. L. D. Wood *et al.*, *Science* **318**, 1108 (2007).
12. Materials and methods are available as supporting material on *Science* Online.
13. D. P. Cahill *et al.*, *Clin. Cancer Res.* **13**, 2038 (2007).
14. C. Hunter *et al.*, *Cancer Res.* **66**, 3987 (2006).
15. J. M. Winter, J. R. Brody, S. E. Kern, *Cancer Biol. Ther.* **5**, 360 (2006).
16. S. Jones *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 4283 (2008).
17. S. Jones *et al.*, *Science* **321**, 1801 (2008).
18. R. Kraus-Ruppert, J. Laissue, H. Burki, N. Odartchenko, *J. Comp. Neurol.* **148**, 211 (1973).
19. P. C. Ng, S. Henikoff, *Nucleic Acids Res.* **31**, 3812 (2003).
20. R. Karchin, Structural models of mutants identified in glioblastomas (2008); http://karchinlab.org/Mutants/CAN-genes/brain/GBM.html
21. F. J. Steemers *et al.*, *Nat. Methods* **3**, 31 (2006).
22. R. J. Leary *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
23. P. Cairns *et al.*, *Nat. Genet.* **11**, 210 (1995).
24. J. M. Nigro *et al.*, *Nature* **342**, 705 (1989).
25. J. Li *et al.*, *Science* **275**, 1943 (1997).
26. K. Ueki *et al.*, *Cancer Res.* **56**, 150 (1996).
27. A. J. Wong *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 6899 (1987).
28. M. Mizoguchi, C. L. Nutt, G. Mohapatra, D. N. Louis, *Brain Pathol.* **14**, 372 (2004).
29. L. Frederick, X. Y. Wang, G. Eley, C. D. James, *Cancer Res.* **60**, 1383 (2000).
30. Y. Li *et al.*, *Cell* **69**, 275 (1992).
31. G. Thiel *et al.*, *Anticancer Res.* **15**, 2495 (1995).
32. Y. Samuels *et al.*, *Science* **304**, 554 (2004).
33. D. K. Broderick *et al.*, *Cancer Res.* **64**, 5048 (2004).
34. G. L. Gallia *et al.*, *Mol. Cancer Res.* **4**, 709 (2006).
35. S. Ekins, Y. Nikolsky, A. Bugrim, E. Kirillov, T. Nikolskaya, *Methods Mol. Biol.* **356**, 319 (2007).
36. J. Lin *et al.*, *Genome Res.* **17**, 1304 (2007).
37. V. E. Velculescu, L. Zhang, B. Vogelstein, K. W. Kinzler, *Science* **270**, 484 (1995).
38. S. Saha *et al.*, *Nat. Biotechnol.* **20**, 508 (2002).
39. M. Sultan *et al.*, *Science* **321**, 956 (2008).
40. R. Lister *et al.*, *Cell* **133**, 523 (2008).
41. A. Mortazavi, B. A. Williams, K. McCue, L. Schaeffer, B. Wold, *Nat. Methods* **5**, 621 (2008).
42. R. Morin *et al.*, *Biotechniques* **45**, 81 (2008).
43. B. V. Geisbrecht, S. J. Gould, *J. Biol. Chem.* **274**, 30527 (1999).
44. X. Xu *et al.*, *J. Biol. Chem.* **279**, 33946 (2004).
45. S. M. Lee *et al.*, *Free Radic. Biol. Med.* **32**, 1185 (2002).
46. S. Y. Kim *et al.*, *Mol. Cell. Biochem.* **302**, 27 (2007).
47. A. Nekrutenko, D. M. Hillis, J. C. Patton, R. D. Bradley, R. J. Baker, *Mol. Biol. Evol.* **15**, 1674 (1998).
48. G. T. Jennings, K. I. Minard, L. McAlister-Henn, *Biochemistry* **36**, 13743 (1997).
49. I. S. Kil, S. Y. Kim, S. J. Lee, J. W. Park, *Free Radic. Biol. Med.* **43**, 1197 (2007).
50. E. F. Pettersen *et al.*, *J. Comput. Chem.* **25**, 1605 (2004).

51. Contact information for the authors who directed the major components of this project is as follows: H. Yan, G. J. Riggins (clinical and sample coordination), yan00002@mc.duke.edu, griggin1@jhmi.edu; D. D. Bigner (pathological and clinical review), bigne001@mc.duke.edu; R. Karchin (bioinformatic analysis), karchin@jhu.edu; N. Papadopoulos (gene expression analysis), npapado1@jhmi.edu; G. Parmigiani (statistical analysis), gp@jhu.edu; B. Vogelstein, V. E. Velculescu, K. W. Kinzler (sequencing and copy number analysis), bertvog@gmail.com, velculescu@jhmi.edu, kinzler@jhmi.edu. We thank N. Silliman, J. Ptak, L. Dobbyn, and M. Whalen for assistance with PCR amplification; D. Lister, L. J. Ehinger, D. L. Satterfield, J. D. Funkhouser, and P. Killela for assistance with DNA purification; T. Sjöblom for assistance with database management; the Agencourt sequencing team for assistance with automated sequencing; and C.-S. Liu and the SoftGenetics team for their assistance with mutation detection analyses. This project was carried out under the auspices of the Ludwig Brain Tumor Initiative and was supported by the Virginia and D. K. Ludwig Fund for Cancer Research, NIH grants CA121113, N5052507, CA43460, CA57345, CA62924, CA09547, 5P50-NS-20023, CA108786, and CA11898, the Pew Charitable Trusts, the Pediatric Brain Tumor Foundation Institute, the Hirschhorn Foundation, Alex's Lemonade Stand Foundation, the American Brain Tumor Association, the American Society of Clinical Oncology, the Brain Tumor Research Fund and Beckman Coulter Corporation. Under separate licensing agreements between the Johns Hopkins University and Genzyme, Beckman Coulter, and Exact Sciences Corporations, B.V., V.E.V., and K.W.K. are entitled to a share of royalties received by the university on sales of products related to research described in this paper. These authors and the university own Genzyme and Exact Sciences stock, which is subject to certain restrictions under university policy. The terms of these arrangements are managed by the Johns Hopkins University in accordance with its conflict-of-interest policies.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1164382/DC1
Materials and Methods
Figs. S1 and S2
Tables S1 to S9
References

7 August 2008; accepted 27 August 2008
Published online 4 September 2008;
10.1126/science.1164382
Include this information when citing this paper.

REPORTS

Quantum Communication with Zero-Capacity Channels

Graeme Smith[1]* and Jon Yard[2]

Communication over a noisy quantum channel introduces errors in the transmission that must be corrected. A fundamental bound on quantum error correction is the quantum capacity, which quantifies the amount of quantum data that can be protected. We show theoretically that two quantum channels, each with a transmission capacity of zero, can have a nonzero capacity when used together. This unveils a rich structure in the theory of quantum communications, implying that the quantum capacity does not completely specify a channel's ability to transmit quantum information.

Noise is the enemy of all modern communication links. Cellular, Internet and satellite communications all depend crucially on active steps taken to mitigate and correct for noise. The study of communication in the presence of noise was formalized by Shannon (*1*), who simplified the analysis by making probabilistic assumptions about the nature of the noise. By modeling a noisy channel $\mathcal{N}$ as a probabilistic map from input signals to output signals, the capacity $C(\mathcal{N})$ of $\mathcal{N}$ is defined as the number of bits that can be transmitted per channel use, with vanishing errors in the limit of many transmissions. This capacity is computed via the formula $C(\mathcal{N}) = \max_X I(X;Y)$, where the maximization is over random variables X at the input of the channel, Y is the resulting output of the channel, and the mutual information $I(X;Y) = H(X) + H(Y) - H(X,Y)$ quantifies the correlation between input and output. $H(X) = -\sum_x p_x \log_2 p_x$ denotes the Shannon entropy, which quantifies the amount of randomness in X. The capacity, measured in bits per channel use, is the fundamental bound between communication rates that are achievable in principle and those that are not. The capacity formula guides the design of practical error-correction techniques by providing a benchmark against which engineers can test the performance of their systems. Practical implementations guided by the capacity result now come strikingly close to the Shannon limit (*2*).

A fundamental prediction of the capacity formula is that the only channels with zero capacity are precisely those for which the input

and output are completely uncorrelated. Furthermore, suppose one is given simultaneous access to two noisy channels $\mathcal{N}_1$ and $\mathcal{N}_2$. The capacity of the product channel $\mathcal{N}_1 \times \mathcal{N}_2$, where the channels are used in parallel, takes the simple form $\mathcal{C}(\mathcal{N}_1 \times \mathcal{N}_2) = \mathcal{C}(\mathcal{N}_1) + \mathcal{C}(\mathcal{N}_2)$; that is, the capacity is additive. Additivity shows that capacity is an intrinsic measure of the information-conveying properties of a channel.

Quantum data are an especially delicate form of information and are particularly susceptible to the deleterious effects of noise. Because quantum communication promises to allow unconditionally secure communication (*3*), and a quantum computer could dramatically speed up some computations (*4*), there is tremendous interest in techniques to protect quantum data from noise. A quantum channel $\mathcal{N}$ models a physical process that adds noise to a quantum system via an interaction with an unobservable environment (Fig. 1), generalizing Shannon's model and enabling a more accurate depiction of the underlying physics. In this setting, it is natural to ask what the capacity of a quantum channel is for transmitting quantum-mechanical information (*5*) and whether it has a simple formula in analogy with Shannon's.

Just as any classical message can be reversibly expressed as a sequence of bits, a quantum message (that is, an arbitrary state of a given quantum system) can be reversibly transferred to a collection of two-level quantum systems, or "qubits," giving a measure of the size of the system. The goal of quantum communication is to transfer the joint state of a collection of qubits from one location to another (Fig. 2). The quantum capacity $\mathcal{Q}(\mathcal{N})$ of a quantum channel $\mathcal{N}$ is the number of qubits per channel use that can be reliably transmitted via many noisy transmissions, where each transmission is modeled by $\mathcal{N}$. Although noiseless quantum communication with a noisy quantum channel is one of the simplest and most natural communication tasks one can imagine for quantum information, it is not nearly as well understood as its classical counterpart.

An analog for mutual information in the quantum capacity has been proposed (*6*) and called the "coherent information"

$$\mathcal{Q}^{(1)}(\mathcal{N}) = \max_{\rho^A} [H(B) - H(E)] \qquad (1)$$

The entropies are measured on the states induced at the output and environment of the channel (Fig. 1) by the input state ρ^A, where $H(B)$ is the von Neumann entropy of the state ρ^B at the output. Coherent information is rather different from mutual information. This difference is closely related to the no-cloning theorem (*7*), which states that quantum information cannot be copied, because the coherent information roughly measures how much more information B holds than E. The no-cloning theorem itself is deeply tied to the fundamentally quantum concept of entanglement, in which the whole of a quantum system can be in a definite state while the states of its parts are uncertain.

The best-known expression for the quantum capacity $\mathcal{Q}$ is given (*8*–*10*) by the "regularization" of $\mathcal{Q}^{(1)}$: $\mathcal{Q}(\mathcal{N}) = \lim_{n\to\infty} \frac{1}{n} \mathcal{Q}^{(1)}(\mathcal{N}^{\otimes n})$. Here $\mathcal{N}^{\otimes n}$ represents the parallel use of n copies of $\mathcal{N}$. The asymptotic nature of this expression prevents one from determining the quantum capacity of a given channel in any effective way, while also making it difficult to reason about its general properties. In contrast to Shannon's capacity, where regularization is unnecessary, here it cannot be removed in general (*11*, *12*). Consequently, even apparently simple questions, such as determining from a channel's description whether it can be used to send any quantum information, are currently unresolved. We find that the answer to this question depends on context; there are pairs of zero-capacity channels that, used together, have a positive quantum capacity (Fig. 3). This shows that the quantum capacity is not additive, and thus the quantum capacity of a channel does not completely specify its capability for transmitting quantum information.

Although a complete characterization of zero-capacity channels is unknown, certain classes of zero-capacity channels are known. One class consists of channels for which the joint quantum state of the output and environment is symmetric under interchange. These symmetric channels are quite different from Shannon's zero-capacity channels, because they display correlations between the input and output. However, they are useless by themselves for quantum communication because their symmetry implies that any capacity would lead to a violation of the no-cloning theorem (*7*, *13*). Another class of zero-capacity channels is entanglement-binding channels (*14*, *15*), also called Horodecki channels, which can only produce very weakly entangled states, satisfying a condition called positive partial transposition (*16*).

Even though channels from one or the other of these classes cannot be combined to faithfully transmit quantum data, we find that when one combines a channel from each class, it is sometimes possible to obtain a positive quantum capacity. We do this by proving a new relationship between two further capacities of a quantum channel: the private capacity (*10*) and the assisted capacity (*17*).

The private capacity $\mathcal{P}(\mathcal{N})$ of a quantum channel $\mathcal{N}$ is the rate at which it can be used to send classical data that is secure against an eavesdropper with access to the environment of the channel. This capacity is closely related to quantum key distribution protocols (*3*) and was shown (*10*) to equal the regularization of the private information

$$\mathcal{P}^{(1)}(\mathcal{N}) = \max_{X,\rho_i^A} I(X;B) - I(X;E)) \qquad (2)$$

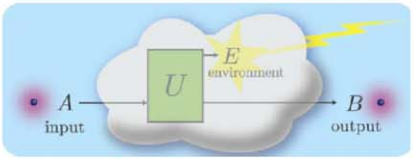

Fig. 1. Representation of a quantum channel. A channel reversibly transfers the state of a physical system in the laboratory of the sender to the combination of the system possessed by the receiver and an environment that is inaccessible to the users of the channel. Discarding the environment results in a noisy evolution of the state. The input and output denote separate places in space and/or time, modeling, for example, a leaky optical fiber or the irreversible evolution of the state of a quantum dot.

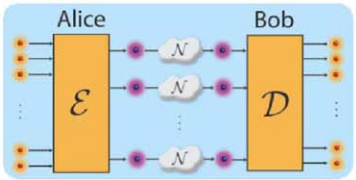

Fig. 2. The quantum capacity of a quantum channel. Quantum data are held by a sender (traditionally called Alice), who would like to transmit it to a receiver (Bob) with many parallel uses of a noisy quantum channel $\mathcal{N}$. Alice encodes the data with a collective encoding operation $\mathcal{E}$, which results in a joint quantum state on the inputs of the channels $\mathcal{N}^{\otimes n}$. The encoded state is sent through the noisy channels. When Bob receives the state, he applies a decoding operation $\mathcal{D}$, which acts collectively on the many outputs of the channels. After decoding, Bob holds the state that Alice wished to send. The quantum capacity is the total number of qubits in the state Alice sends divided by the number of channel uses.

[1]IBM T. J. Watson Research Center, 1101 Kitchawan Road, Yorktown Heights, NY 10598, USA. [2]Quantum Institute, Center for Nonlinear Studies (CNLS), Computer, Computational and Statistical Sciences (CCS-3), Los Alamos National Laboratory, Los Alamos, NM 87545, USA.

*To whom correspondence should be addressed. E-mail: gsbsmith@gmail.com

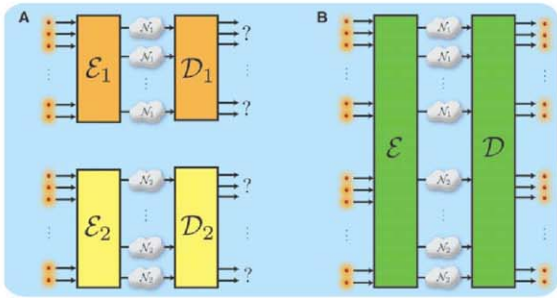

Fig. 3. **(A)** Alice and Bob attempt to separately use two zero-capacity channels $\mathcal{N}_1$ and $\mathcal{N}_2$ to transfer quantum states. Alice uses separate encoders $\mathcal{E}_1$ and $\mathcal{E}_2$ for each group of channels, and Bob uses separate decoders $\mathcal{D}_1$ and $\mathcal{D}_2$. Any attempt will fail because the capacity of each channel is zero. **(B)** The same two channels being used in parallel for the same task. Alice's encoder $\mathcal{E}$ now has simultaneous access to the inputs of all channels being used and Bob's decoding $\mathcal{D}$ is also performed jointly. Noiseless communication is nonetheless possible because $\mathcal{Q}$ is not additive.

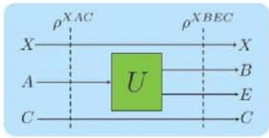

Fig. 4. Relating the private capacity and the assisted capacity. A straightforward proof of Eq. 3 uses the expression (*17*) $\mathcal{Q}_A(\mathcal{N}) = \frac{1}{2}\max_{\rho^{XAC}} I(X;B|C) - I(X;E|C)$. Here, $I(X;B|C)$ is the conditional mutual information: $H(XC) + H(BC) - H(XBC) - H(C)$. It is evaluated on the state obtained by putting the $\mathcal{A}$ part of a state ρ^{XAC} into the channel $\mathcal{N}$, which can be thought of as mapping $A \to BE$ as in Fig. 2. The maximization here is similar in form to Eq. 2, but is over a less-constrained type of state. Therefore, $\frac{1}{2}\mathcal{P}^{(1)}(\mathcal{N}) \leq \mathcal{Q}_A(\mathcal{N})$. This bound holds for the associated regularized quantities, and because regularization does not change $\mathcal{Q}_A$, Eq. 3 follows.

where the maximization is over classical random variables X and quantum states ρ_x^A on the input of $\mathcal{N}$ depending on the value x of X.

In order to find upper bounds on the quantum capacity, an "assisted capacity" was recently introduced (*17*), in which one allows the free use of arbitrary symmetric channels to assist quantum communication over a given channel. Letting $\mathcal{A}$ denote a symmetric channel of unbounded dimension (the strongest such channel), the assisted capacity $\mathcal{Q}_A(\mathcal{N})$ of a quantum channel $\mathcal{N}$ satisfies (*17*) $\mathcal{Q}_A(\mathcal{N}) = \mathcal{Q}(\mathcal{N} \times \mathcal{A}) = \mathcal{Q}^{(1)}(\mathcal{N} \times \mathcal{A})$.

Because the dimension of the input to $\mathcal{A}$ is unbounded, we cannot evaluate the assisted capacity in general. Nonetheless, the assisted capacity helps to reason about finite-dimensional channels.

Although Horodecki channels have zero quantum capacity, examples of such channels with nonzero private capacity are known (*18*, *19*). One of the two zero-capacity channels we will combine to give positive joint capacity is such a private Horodecki channel $\mathcal{N}_H$ and the other is the symmetric channel $\mathcal{A}$. Our key tool is the following new relationship between the capacities of any channel $\mathcal{N}$ (Fig. 4)

$$\frac{1}{2}\mathcal{P}(\mathcal{N}) \leq \mathcal{Q}_A(\mathcal{N}) \qquad (3)$$

A channel's assisted capacity is at least as large as half its private capacity. It follows that any private Horodecki channel $\mathcal{N}_H$ has a positive assisted capacity, and thus the two zero-capacity channels $\mathcal{N}_H$ and $\mathcal{A}$ satisfy $\mathcal{Q}_A(\mathcal{N}_H) = \mathcal{Q}(\mathcal{N}_H \times \mathcal{A}) > 0$.

Although our construction involves systems of unbounded dimension, one can show that any private Horodecki channel can be combined with a finite symmetric channel to give positive quantum capacity. In particular, there is a private Horodecki channel acting on a four-level system (*19*). This channel gives positive quantum capacity when combined with a small symmetric channel—a 50% erasure channel $\mathcal{A}_e$ with a four-level input, which half of the time delivers the input state to the output, otherwise telling the receiver that an erasure has occurred. We show (*20*) that the parallel combination of these channels has a quantum capacity greater than 0.01.

We find this "superactivation" to be a startling effect. One would think that the question, "can this communication link transmit any information?" would have a straightforward answer. However, with quantum data, the answer may well be "it depends on the context." Taken separately, private Horodecki channels and symmetric channels are useless for transmitting quantum information, albeit for entirely different reasons. Nonetheless, each channel has the potential to activate the other, effectively canceling the other's reason for having zero capacity. We know of no analog of this effect in the classical theory. Perhaps each channel transfers some different but complementary kind of quantum information. If so, can these kinds of information be quantified in an operationally meaningful way? Are there other pairs of zero-capacity channels displaying this effect? Are there triples? Does the private capacity also display superactivation? Can all Horodecki channels be superactivated, or just those with positive private capacity? What new insights does this yield for computing the quantum capacity in general?

Besides additivity, our findings resolve two open questions about the quantum capacity. First we find (*20*) that the quantum capacity is not a convex function of the channel. Convexity of a capacity means that a probabilistic mixture of two channels never has a higher capacity than the corresponding average of the capacities of the individual channels. Violation of convexity leads to a counterintuitive situation in which it can be beneficial to forget which channel is being used. We also find (*20*) channels with an arbitrarily large gap between $\mathcal{Q}^{(1)}$ [the so-called "hashing rate" (*8*–*10*)] and the quantum capacity. It had been consistent with previous results (*11*, *12*) to believe that $\mathcal{Q}$ and $\mathcal{Q}^{(1)}$ would be equal up to small corrections. Our work shows that this is not the case and indicates that the hashing rate is an overly pessimistic benchmark against which to measure the performance of practical error-correction schemes. This could be good news for the analysis of fault-tolerant quantum computation in the very noisy regime.

Forms of this sort of superactivation are known in the multiparty setting, where several separated parties communicate via a quantum channel with multiple inputs or outputs (*21*–*24*), and have been conjectured for a quantum channel assisted by classical communication between the sender and receiver (*25*, *26*). Because these settings are rather complex, it is perhaps unsurprising to find exotic behavior. In contrast, the problem of noiseless quantum communication with a noisy quantum channel is one of the simplest and most natural communication tasks imaginable in a quantum-mechanical context. Our findings uncover a level of complexity in this simple problem that had not been anticipated and point toward several fundamentally new questions about information and communication in the physical world.

References and Notes

1. C. E. Shannon, *Bell Syst. Tech. J.* **27**, 379 (1948).
2. T. Richardson, R. Urbanke, *IEEE Commun. Mag.* **41**, 126 (2003).
3. C. H. Bennett, G. Brassard, in *Proceedings of the IEEE International Conference on Computers, Systems and Signal Processing* (IEEE, New York, 1984), pp. 175–179.

4. P. W. Shor, *Proceedings of the 35th Annual Symposium on Foundations of Computer Science* (IEEE Computer Society Press, Los Alamitos, CA 1994), pp. 124–134.
5. C. H. Bennett, P. W. Shor, *Science* **303**, 1784 (2004).
6. B. Schumacher, M. A. Nielsen, *Phys. Rev. A* **54**, 2629 (1996).
7. W. Wootters, W. Zurek, *Nature* **299**, 802 (1982).
8. S. Lloyd, *Phys. Rev. A* **55**, 1613 (1997).
9. P. W. Shor, lecture notes from the Mathematical Sciences Research Institute Workshop on Quantum Computation, Berkeley, CA, 2002. Available online at www.msri.org/publications/ln/msri/2002/quantumcrypto/shor/1/.
10. I. Devetak, *IEEE Trans. Inf. Theory* **51**, 44 (2005).
11. D. DiVincenzo, P. W. Shor, J. A. Smolin, *Phys. Rev. A* **57**, 830 (1998).
12. G. Smith, J. A. Smolin, *Phys. Rev. Lett.* **98**, 030501 (2007).
13. C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, *Phys. Rev. Lett.* **78**, 3217 (1997).
14. M. Horodecki, P. Horodecki, R. Horodecki, *Phys. Lett. A* **223**, 1 (1996).
15. P. Horodecki, *Phys. Lett. A* **232**, 333 (1997).
16. A. Peres, *Phys. Rev. Lett.* **77**, 1413 (1996).
17. G. Smith, J. Smolin, A. Winter, *IEEE Trans. Inf. Theory* **54**, 4208 (2008).
18. K. Horodecki, M. Horodecki, P. Horodecki, J. Oppenheim, *Phys. Rev. Lett.* **94**, 160502 (2005).
19. K. Horodecki, L. Pankowski, M. Horodecki, P. Horodecki, *IEEE Trans. Inf. Theory* **54**, 2621 (2008).
20. Further details can be found in the supporting online material on *Science* Online.
21. P. Shor, J. Smolin, A. Thapliyal, *Phys. Rev. Lett.* **90**, 107901 (2003).
22. W. Dur, J. Cirac, P. Horodecki, *Phys. Rev. Lett.* **93**, 020503 (2004).
23. R. Duan, Y. Shi, *Phys. Rev. Lett.* **101**, 020501 (2008).
24. L. Czekaj, P. Horodecki, preprint available at http://arxiv.org/pdf/0807.3977.
25. P. Horodecki, M. Horodecki, R. Horodecki, *Phys. Rev. Lett.* **82**, 1056 (1999).
26. P. Shor, J. Smolin, B. Terhal, *Phys. Rev. Lett.* **86**, 2681 (2001).

27. We are indebted to C. Bennett, C. Callaway, E. Timmermans, B. Toner, and A. Winter for encouragement and comments on an earlier draft. J.Y. is supported by the CNLS and the Quantum Institute through grants provided by the Laboratory Directed Research and Development program of the U.S. Department of Energy.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1162242/DC1
SOM Text
References

24 June 2008; accepted 13 August 2008
Published online 21 August 2008;
10.1126/science.1162242
Include this information when citing this paper.

Synthesis and Solid-State NMR Structural Characterization of ^{13}C-Labeled Graphite Oxide

Weiwei Cai,[1,2] Richard D. Piner,[1] Frank J. Stadermann,[3] Sungjin Park,[1] Medhat A. Shaibat,[4] Yoshitaka Ishii,[4] Dongxing Yang,[1] Aruna Velamakanni,[1] Sung Jin An,[5] Meryl Stoller,[1] Jinho An,[1] Dongmin Chen,[2] Rodney S. Ruoff[1]*

The detailed chemical structure of graphite oxide (GO), a layered material prepared from graphite almost 150 years ago and a precursor to chemically modified graphenes, has not been previously resolved because of the pseudo-random chemical functionalization of each layer, as well as variations in exact composition. Carbon-13 (^{13}C) solid-state nuclear magnetic resonance (SSNMR) spectra of GO for natural abundance ^{13}C have poor signal-to-noise ratios. Approximately 100% ^{13}C-labeled graphite was made and converted to ^{13}C-labeled GO, and ^{13}C SSNMR was used to reveal details of the chemical bonding network, including the chemical groups and their connections. Carbon-13–labeled graphite can be used to prepare chemically modified graphenes for ^{13}C SSNMR analysis with enhanced sensitivity and for fundamental studies of ^{13}C-labeled graphite and graphene.

Unlike crystalline materials, the structure of materials that are amorphous or that vary in chemical composition can be difficult to determine. Solid-state nuclear magnetic resonance (SSNMR) can provide important structural insights, but often requires very high enrichment of nuclei with NMR-active spins. One example of such a material that has proven difficult to characterize, despite having been first prepared almost 150 years ago (*1*), is graphite oxide (GO), which can be prepared by heating graphite in oxidizing chemicals. GO is a layered material containing interlamellar water. Materials derived from GO include its chemically functionalized (*2*), reduced (*3*), and thermally expanded forms (*4*), as well as chemically modified graphenes (*2*, *3*, *5*–*8*).

SSNMR has been done on GO but has not provided a complete understanding of the chemical structure of this material, although the detailed chemical structure has been actively researched for many years (*2*, *7*). One difficulty is that the spectra do not attain a high signal-to-noise (S/N) ratio for natural abundance ^{13}C. The lack of ^{13}C-labeled GO has prevented application of modern multidimensional SSNMR methods that can provide information on the bonding arrangements of atoms and their connectivities. Although a series of one-dimensional (1D) ^{13}C SSNMR studies for GO and reduced GO revealed signal assignments and the basic chemical compositions of each, there is sparse experimental evidence of the connectivities of the chemical groups such as sp^2-bonded carbons (C=C), epoxide, carbonyl, and carboxylic groups. Thus, a variety of structural models of GO are still debated (*2*). However, we found from Monte Carlo simulations that even at only 20% ^{13}C, the abundance of ^{13}C-^{13}C bonds will be 400 times that of an unlabeled sample, so that the time required for detecting ^{13}C-^{13}C pairs in SSNMR of such a

[1]Department of Mechanical Engineering and the Texas Materials Institute, University of Texas at Austin, Austin, TX 78712, USA. [2]Beijing National Laboratory for Condensed Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing 100080, China. [3]Laboratory for Space Sciences, Department of Physics, Washington University, St. Louis, MO 63130, USA. [4]Department of Chemistry, University of Illinois at Chicago, 845 West Taylor Street, Chicago, IL 60607, USA. [5]National Creative Research Initiative Center for Semiconductor Nanorods and Department of Materials Science and Engineering, Pohang University of Science and Technology, Pohang, Gyeongbuk 790-784, Korea.

*To whom correspondence should be addressed. E-mail: r.ruoff@mail.utexas.edu

Fig. 1. (**A**) Optical images of Ni, and SEM images of (**B**) ^{13}C-labeled synthetic graphite and (**C**) the wrinkles.

sample will be less by a factor of about 160,000 than that of unlabeled samples. Evaluation of detailed bonding networks (three neighboring ^{13}C's and so on) can be done with Monte Carlo modeling in a straightforward way for any percentage labeling of ^{13}C.

There have been a number of recent publications exploring the catalytic growth of graphite or graphitic structures (*9*–*20*). We have developed a thermal chemical vapor deposition (CVD) method for growing synthetic graphite from methane on a resistively heated Ni foil as a catalytic substrate.

A detailed description of our reactor is given in the supporting online material (*21*). Briefly, our reactor consists of a vacuum chamber with a metal foil supported in the center with high-current electrodes. The Ni foil can be resistively heated to near its melting point. During the deposition, the substrate temperature was held at ~1200° to 1300°C and the pressure at 1 atm. The gas used was a mixture of 10% methane and 90% Ar. The methane was 1%, 30%, 50%, 70%, and 99.95% $^{13}CH_4$, in different growth runs. A deposition rate of ~2 μm/hour (for thin films) was obtained, and the growth rate was slowed to ~0.2 μm/hour for thicker films. Atomic force microscopy (AFM), scanning electron microscopy (SEM), x-ray diffraction (XRD), and Raman spectroscopy confirmed that the as-deposited carbon was very high-quality graphite. The graphite films deposited onto the Ni foils appeared very smooth and continuous by optical microscopy (Fig. 1). The smooth surface areas are separated from each other by wrinkles that are likely caused by the different thermal expansion coefficients of the deposited graphite and the nickel substrate (*17*). A typical smooth surface region is about 2 μm across, much smaller than the substrate grain sizes, and the typical wrinkle height as obtained by AFM is about 50 nm (fig. S2) (*21*). The ^{13}C content of the graphite samples was measured with a modified CAMECA ims3f secondary ion mass spectrometry (SIMS) instrument (CAMECA, Gennevilliers, France). The analytical error of these measurements is estimated to be ~5%, largely because the sample surfaces are uneven on a 10 to 100 μm scale. The ^{13}C content of each of the graphite films measured by SIMS was 1%, 19%, 41%, 54%, and 86% for samples prepared from 1%, 30%, 50%, 70%, and 99.95% $^{13}CH_4$, respectively. The ^{13}C contents measured by SIMS (19%, 41%, 54%, and 86%) differed significantly from the percentages of $^{13}CH_4$ used because for these four experimental runs the chamber had accumulated considerable residual carbon-containing material from many trial experiments with unlabeled methane. A thorough cleaning of the chamber before another run with 99.95% $^{13}CH_4$ yielded 99.5% ^{13}C-content synthetic graphite measured by SIMS. Use of a clean chamber is thus important for the ^{13}C content of the graphite to more closely match the ^{13}C content of the input $^{13}CH_4$ and $^{12}CH_4$.

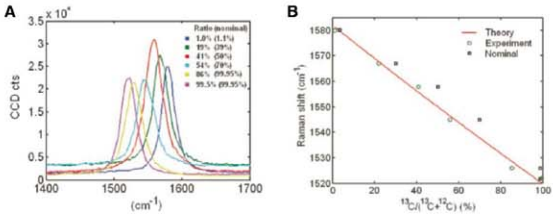

Fig. 2. (**A**) The G band of various ^{13}C-labeled synthetic graphites. (**B**) Shift of the G band frequency as a function of percentage of ^{13}C, as determined by SIMS. The theoretical curve was obtained as described in (*21*).

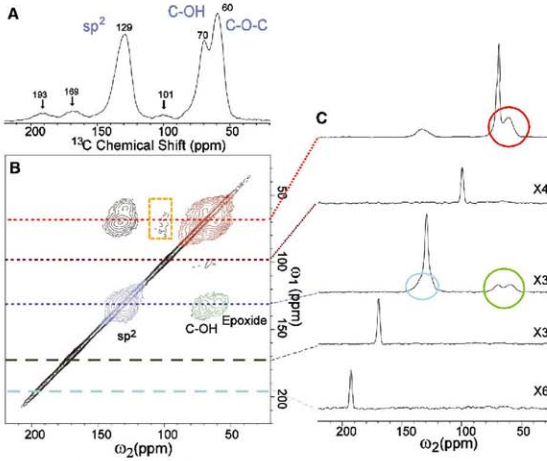

Fig. 3 (**A**) 1D ^{13}C MAS and (**B**) 2D $^{13}C/^{13}C$ chemical-shift correlation solid-state NMR spectra of ^{13}C-labeled graphite oxide with (**C**) slices selected from the 2D spectrum at the indicated positions (70, 101, 130, 169, and 193 ppm) in the ω_1 dimension. All the spectra were obtained at a ^{13}C NMR frequency of 100.643 MHz with 90 kHz ^1H decoupling and 20 kHz MAS for 12 mg of the sample. In (A), the ^{13}C MAS spectrum was obtained with direct ^{13}C excitation by a $\pi/2$-pulse. The recycle delay was 180 s, and the experimental time was 96 min for 32 scans. In (B), the 2D spectrum was obtained with cross polarization and fpRFDR ^{13}C-^{13}C dipolar recoupling sequence (*24*). The experimental time is 12.9 hours with recycle delays of 1.5 s and 64 scans for each real or imaginary t_1 point. A Gaussian broadening of 150 Hz was applied. The green, red, and blue areas in (B) and circles in (C) represent cross peaks between sp^2 and C-OH/epoxide (green), those between C-OH and epoxide (red), and those within sp^2 groups (blue), respectively.

High-resolution SSNMR using magic angle spinning (MAS) has been used as a primary method to characterize GO at the molecular level (*2*, *22*, *23*). Figure 3A shows 1D ^{13}C MAS spectrum, and Fig. 3B shows 2D $^{13}C/^{13}C$ chemical-shift correlation SSNMR spectrum of ^{13}C-labeled GO (made from approximately 100% ^{13}C-labeled graphite) that was prepared using a modified Hummer's method (*2*) with ^{13}C-labeled graphite (*21*). The signal assignments for the three major

peaks at 59.7 (epoxide ^{13}C), 69.6 (^{13}C-OH), and 129.3 (sp^2 ^{13}C) parts per million (ppm) in Fig. 3A are based on studies by Lerf *et al.* (*23*). We performed additional analyses and confirmed that these assignments are likely correct (*21*). The obtained 1D spectrum shows similar features with those reported in (*23*) except for the relatively well-resolved minor peaks at 101, 169, and 193 ppm, which respectively yield only 12%, 15%, and 4% of the integrated intensity of the 70-ppm peak. The spectrum shows a considerably stronger sp^2 peak and a much weaker peak at 169 ppm compared with that by Szabo *et al.* (*22*), although the observed peak positions are similar. The peak at 169 ppm was previously attributed to ^{13}C=O (*22*). The results imply that their sample was subject to a higher level of oxidization than ours.

Because the natural abundance of ^{13}C is only 1%, attaining sufficient sensitivity in a 2D spectrum, as shown in Fig. 3B, is extremely difficult without the ^{13}C-labeled samples. For example, ^{13}C-^{13}C bonds exist only at 0.01% abundance without labeling; thus, obtaining an equivalent 2D spectrum for an unlabeled sample would require about 10^8 times as much time. The experiment in Fig. 3B was performed with a finite-pulse radio frequency–driven dipolar recoupling (fpRFDR) mixing sequence (*24*). With the labeled sample and a relatively short mixing time (1.6 ms), the experiment permitted us to identify ^{13}C-^{13}C pairs directly bonded or separated by two bonds. In Fig. 3B, there are several strong cross peaks. For example, cross peaks were observed at the positions (ω_1, ω_2) = (133 ppm, 70 ppm) and (130 ppm, 59 ppm) (green signals in Fig. 3B). These cross peaks represent spin polarization transfer from sp^2 carbons observed at ~130 ppm in ω_1, to C-OH and epoxide groups, which appear at 70 ppm and 59 ppm in ω_2, respectively. Unlike previous studies, these cross peaks directly present the connectivity between sp^2 ^{13}C and ^{13}C-OH, as well as that between sp^2 and epoxide ^{13}C through spin-spin dipolar couplings. The cross peak intensities are about 10%, compared with the diagonal signals, which represent signals for ^{13}C spins that had the same NMR frequencies in the two dimensions ($\omega_1 = \omega_2$). The relatively strong intensities of the cross peaks suggest that a large fraction of the sp^2 ^{13}C atoms are directly bonded to ^{13}C-OH and/or epoxide ^{13}C.

We also observed strong cross peaks between ^{13}C-OH and ^{13}C-epoxide (red signals). Again, the data suggest that a large fraction of C-OH and epoxide carbons are bonded to each other. The blue cross peaks indicate that there are sp^2 species having slightly different chemical shifts and that they are bonded with each other. Indeed, the sp^2 ^{13}C shifts for the cross peaks (green) are slightly different for the cross peaks to C-OH (133 ppm) and that to epoxide (130 ppm).

In the previous studies (*23*), the proximities of the chemical groups were tentatively assigned based on formation of the phenol group during the deoxidization of GO. In contrast, the present SSNMR data directly shows that these groups are chemically bonded. For the minor species, we found cross peaks only for the peak at 101 ppm (orange box). There are no visible cross peaks for the other minor components at 169 and 193 ppm, despite these minor peaks having comparable intensities to the 101-ppm peak. The results imply that these minor components at 169 and 193 ppm, which were previously attributed to the presence of C=O (*2*, *22*), are spatially separated from a majority of the sp^2, C-OH, and epoxide carbons.

Among six previously proposed models (*22*), only two, the Lerf-Klinowski model (*23*) and the Dékány model (*22*), present such a network. The model proposed by Dékány *et al.* may be correct for their more highly oxidized compound, because that structural model seems to call for a considerably higher level of oxidization to complete the modification of an sp^2 network into a network of linked cyclohexanes. Further studies would be needed to define all of the structural details of the system.

Chemically modified graphenes that will be of importance in a variety of new materials can now be ^{13}C-labeled and more effectively studied by SSNMR. High-quality ^{13}C-labeled graphite should find use for fundamental property measurements, including of ^{13}C-labeled graphene.

References and Notes

1. B. C. Brodie, *Annales des Chimie et des Physique* **59**, 466 (1860).
2. S. Stankovich *et al.*, *Carbon* **45**, 1558 (2007).
3. S. Stankovich *et al.*, *J. Mater. Chem.* **16**, 155 (2006).
4. M. J. McAllister *et al.*, *Chem. Mater.* **19**, 4396 (2007).
5. D. A. Dikin *et al.*, *Nature* **448**, 457 (2007).
6. S. Stankovich *et al.*, *Nature* **442**, 282 (2006).
7. S. Stankovich, R. D. Piner, S. T. Nguyen, R. S. Ruoff, *Carbon* **44**, 3342 (2006).
8. S. Watcharotone *et al.*, *Nano Lett.* **7**, 1888 (2007).
9. H. H. Angermann, G. Horz, *Appl. Surf. Sci.* **70-71**, 163 (1993).
10. R. Anton, *J. Mater. Res.* **20**, 1837 (2005).
11. A. Barbangelo, R. Sangiorgi, *Mater. Sci. Eng. A* **156**, 217 (1992).
12. F. Bonnet, F. Ropital, Y. Berthier, P. Marcus, *Mater. Corrosion Werkst. Korros.* **54**, 870 (2003).
13. D. V. Fedoseev, S. P. Vnukov, B. V. Derjaguin, *Carbon* **17**, 453 (1979).
14. D. Fujita, T. Homma, *Surf. Interface Anal.* **19**, 430 (1992).
15. D. Fujita, K. Yoshihara, *J. Vac. Sci. Technol. A* **12**, 2134 (1994).
16. C. Klink, I. Stensgaard, F. Besenbacher, E. Laegsgaard, *Surf. Sci.* **342**, 250 (1995).
17. A. N. Obraztsov, E. A. Obraztsova, A. V. Tyurnina, A. A. Zolotukhin, *Carbon* **45**, 2017 (2007).
18. H. Oudghiri-Hassani, S. Rakass, N. Abatzoglou, P. Rowntree, *J. Power Sources* **171**, 850 (2007).
19. R. Sindair, T. Itoh, R. Chin, *Microsc. Microanal.* **8**, 288 (2002).
20. M. Yudasaka, R. Kikuchi, Y. Ohki, S. Yoshimura, *J. Vac. Sci. Technol. A* **16**, 2463 (1998).
21. Materials and methods are available as supporting material on *Science* Online.
22. T. Szabo *et al.*, *Chem. Mater.* **18**, 2740 (2006).
23. A. Lerf, H. Y. He, M. Forster, J. Klinowski, *J. Phys. Chem. B* **102**, 4477 (1998).
24. Y. Ishii, *J. Chem. Phys.* **114**, 8473 (2001).

25. $^{13}CH_4$ was provided through a research grant from Cambridge Isotopes Laboratories, Inc. We (R.S.R.) wish to acknowledge The University of Texas for startup funds and for partial support by the *Defense Advanced Research Projects Agency* Center on Nanoscale Science and Technology for Integrated Micro/NanoElectromechanical Transducers (HR0011-06-1-0048). The WITec Micro-Raman instrument was acquired by an Air Force Office of Scientific Research, Defense University Research Instrumentation Program grant. Y.I. appreciates support from the Dreyfus Foundation Teacher-Scholar Award program and the NSF CAREER program (CHE 449952). S.J.A was financially supported by the Korea Science and Engineering Foundation under the National Creative Research Initiative Project (R16-2004-004-01001-0) of the Ministry of Science and Technology, Korea. J. Goodenough and A. Ruoff commented on an early version of this manuscript.

Supporting Online Material

www.sciencemag.org/cgi/content/full/321/5897/1815/DC1
Materials and Methods
Figs. S1 to S5
References

26 June 2008; accepted 15 August 2008
10.1126/science.1162369

Linear Response Breakdown in Solvation Dynamics Induced by Atomic Electron-Transfer Reactions

Arthur E. Bragg, Molly C. Cavanagh, Benjamin J. Schwartz*

The linear response (LR) approximation, which predicts identical relaxation rates from all nonequilibrium initial conditions that relax to the same equilibrium state, underlies dominant models of how solvation influences chemical reactivity. We experimentally tested the validity of LR for the solvation that accompanies partial electron transfer to and from a monatomic solute in solution. We photochemically prepared the species with stoichiometry Na^0 in liquid tetrahydrofuran by both adding an electron to Na^+ and removing an electron from Na^-. Because atoms lack nuclear degrees of freedom, ultrafast changes in the Na^0 absorption spectrum reflected the solvation that began from our two initial nonequilibrium conditions. We found that the solvation of Na^0 occurs more rapidly from Na^+ than Na^-, constituting a breakdown of LR. This indicates that Marcus theory would fail to describe electron-transfer processes for this and related chemical systems.

Solvent-solute interactions play an integral role in solution-phase chemical reactivity and particularly in electron-transfer (ET) reactions (*1*), in which solvation dynamics—the response of the solvent to changes in solute size and/or electronic charge distribution (*2*)—help drive the motion of charge from donor to acceptor. Current theoretical understanding of how

solvation dynamics influence chemical reactivity [for example, through the Marcus theory of ET (*1*)] is based largely on the idea of linear response (LR) (*3*–*6*). If the Hamiltonian governing solute-solvent interactions is assumed to change linearly with respect to a perturbation that moves the system out of equilibrium, then the relaxation of the perturbed nonequilibrated state should be identical to the regression of the spontaneous fluctuations of the system at equilibrium; this assumption is commonly called the LR approximation (*3*). In this work, we performed an experimental test of LR in which we investigated the solvation dynamics associated with elementary steps that play a role in many common chemical transformations, specifically the spontaneous partial transfer of an electron to and from an atomic solute in solution. We demonstrate that the LR approximation fails to underlie the solvation processes associated with this particular atomic ET reaction, leading us to predict that LR may break down more broadly in solution-phase chemical dynamics than previously appreciated.

A convenient way to characterize the evolution of solute-solvent interactions is to use the framework of time correlation functions (TCFs). At equilibrium, the evolution of solvent-solute interactions can be quantified by Eq. 1

$$C(t) = \frac{\langle \delta E(t) \delta E(0) \rangle}{\langle \delta E(0)^2 \rangle} \quad (1)$$

which autocorrelates the fluctuations of the solute-solvent interaction energy E about its average value, $\delta E(t) = E(t) - \langle E \rangle$; the angled brackets represent equilibrium ensemble averages. This TCF can be viewed as describing the time scale of the intrinsic fluctuations of an equilibrium solvent-solute system: if we artificially prepared a specific solvent-solute configuration that could be reached via thermal fluctuations at equilibrium, then $C(t)$ would reflect the time dependence of the subsequent relaxation away from this configuration. Within the limit of LR, the nonequilibrium relaxation of the solvent-solute interaction energy after perturbation of the system from equilibrium, $S(t)$, should match the TCF of the system at equilibrium (*3*, *7*, *8*)

$$S(t) = \frac{\overline{E}(t) - \overline{E}(\infty)}{\overline{E}(0) - \overline{E}(\infty)} \approx C(t) \quad (2)$$

where the overbars indicate nonequilibrium ensemble averages. Although the LR approximation (Eq. 2) can be derived from statistical mechanical perturbation theory {and thus is applicable when the perturbation to a system is near thermal energies [Boltzmann's constant times the temperature ($\sim k_B T$)] (*3*)}, it has recently been shown that LR arises whenever the fluctuations of a system are Gaussian, regardless of the magnitude of the deviation from equilibrium (*5*, *7*, *8*). It is these Gaussian statistics, in turn, that give rise to the familiar parabolic potentials along the generalized solvent coordinate that lie at the foundation of the Marcus theory of ET (*1*). Thus, when valid, the LR approximation provides a useful means of estimating and interpreting theoretically the time-dependent relaxation that follows the preparation of many (*1*, *3*–*5*, *8*), but not all (*6*, *7*, *9*), highly perturbed solvent-solute configurations.

How can we test the validity of LR experimentally, particularly for the solvation dynamics tied to chemical reactivity? To test LR directly, we must not only track the dynamics of quantity associated with the time-evolving solute-solvent interaction as the system relaxes from a well-defined nonequilibrium configuration [$S(t)$], but also measure the solvent-solute fluctuations at equilibrium [$C(t)$]. Unfortunately, the equilibrium TCF is difficult to access, although a recently developed approach that measures the difference in solvent spectral density after the photoexcitation of a probe solute shows great promise toward making this direct comparison (*10*). Here, we describe an alternative approach to assess the LR approximation: We compare the relaxation dynamics that begin from two different well-defined nonequilibrium configurations but end in the same final equilibrium state. If LR holds, then the time dependence of the solvent-solute relaxation from both nonequilibrium states should be identical to the equilibrium TCF. If the two nonequilibrium configurations relax on different time scales, as we discuss below, then the LR prediction must fail to describe one or both of the relaxation pathways.

Although our approach should apply to any solute-solvent system, we have chosen to survey the relaxation of atoms in solution because these solutes lack internal nuclear degrees of freedom. Thus, unlike the large molecular chromophores that are typically used as solvation probes, the atomic relaxation dynamics that we observe directly reflect the motions of the surrounding solvent and are unobscured by competing intramolecular processes (*11*). Specifically, we ex-

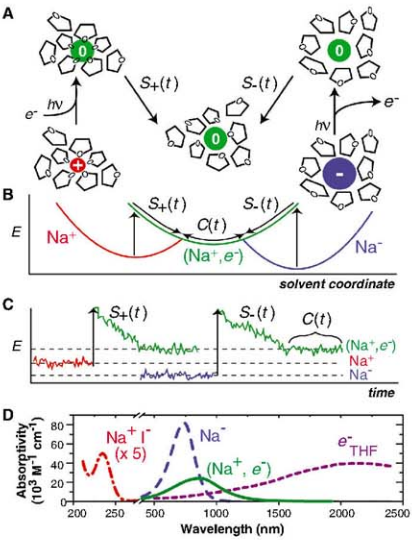

Fig. 1. (A to C) Illustration of how LR can be tested by examining the dynamic solvation of the same solute prepared from two different well-defined initial conditions. **(A)** Schematic depiction of the local solvent reconfiguration required to solvate the nonequilibrium (Na^+, e^-) TCP (green circles labeled 0) prepared after both electron detachment from Na^- (right, blue circle with minus sign) and electron attachment to Na^+ (left, red circle with plus sign). In both cases, the solvent must relax to accommodate changes in solute size and charge that result from partial ET reactions. **(B)** Another representation of the two solvent relaxation pathways shown in (A), presented using the picture derived from the Marcus theory of ET. **(C)** A schematic presentation of the time dependence of the two nonequilibrium solvent relaxation processes shown in (A) and (B); the ordinate in both (B) and (C) is the solute-solvent interaction energy. **(D)** The steady-state equilibrium spectra of the chemical species examined in this work: the absorption spectra of the equilibrated THF-solvated (Na^+, e^-) TCP [green solid curve (*18*)]; the THF-solvated electron [e^-_{THF}, purple dashed curve (*19*)]; the THF-solvated Na^- [blue dashed curve (*23*)]; and the CTTS transition of the Na^+-I^- ion pair in THF (red dash-dotted curve). We used both Na^- and Na^+-I^- as precursors for the photochemical production of nonequilibrium (Na^+, e^-) TCPs in liquid THF via electron detachment and attachment, respectively.

Department of Chemistry and Biochemistry, University of California, Los Angeles Los Angeles, CA 90095-1569, USA.

*To whom correspondence should be addressed. E-mail: schwartz@chem.ucla.edu

amined the nonequilibrium relaxation of neutral atomic sodium (Na^0) in liquid tetrahydrofuran (THF) that was prepared both by photoinduced ET to sodium cations (Na^+) and by electron photodetachment from sodium anions (Na^-, or sodide). Figure 1A schematically illustrates the solvent reconfiguration that occurs after each of these photochemical preparations; Fig. 1B presents this same relaxation in terms of the picture derived from the Marcus theory of ET, in which the relaxation of the solvent-solute interaction energy, E, is expressed as a function of a collective solvent coordinate; and Fig. 1C illustrates the time-dependent solute-solvent interaction energy as would be typically examined in a molecular dynamics (MD) simulation. The red and blue curves in Fig. 1C reflect fluctuations of the ion-solvent interactions about their equilibrium average values before each photochemical process. If we promptly add or remove an electron to or from one of the initially equilibrated ions through photoinduced ET (vertical black arrows in Fig. 1, A to C), the solvent interaction energy associated with the newly created neutral solute will be outside the range of the equilibrium fluctuations [green curves labeled $C(t)$ in Fig. 1, B and C], because the initial solvent structure is equilibrated for an ion and thus poorly accommodates the neutral Na. The nearby solvent molecules will then relax by both translating to accommodate the change in solute size and rotating so that the solvent dipoles become favorably oriented for the neutral Na. We label the nonequilibrium relaxation of the neutral Na that begins with the initial configuration characteristic of the Na anion "$S_-(t)$" (right side of Fig. 1, A to C), and we label the relaxation that begins from the initial configuration associated with the Na cation "$S_+(t)$" (left side of Fig. 1, A to C). If LR holds, then we expect $S_+(t) = S_-(t)$ because both should be equal to $C(t)$.

Atomic Na in THF is a convenient system for testing the LR approximation through this approach for several reasons. The precursors we used to create neutral Na photochemically (Na^- and the Na^+-I^- ion pair) are stable in liquid THF. In addition, the solvent structure and the nature of the delocalized solvent-supported electronic states of liquid THF (*12*, *13*) allow us to rapidly shuttle electrons to and from these ion precursors via photoinduced ET (*14*, *15*). Finally, the spectrum of atomic Na in THF is highly sensitive to the nature of the local solvent environment (*16*) and spans visible and near-infrared wavelengths (Fig. 1D, green solid curve), so we can easily track the solvation dynamics that follow each photochemical preparation, using transient absorption spectroscopy.

The extraordinary sensitivity of the Na atom's absorption spectrum to the local solvent environment arises from the chemical nature of Na in THF: Rather than a solvated atom, the equilibrated neutral Na solute in THF is better thought of as a (Na^+, e^-) tight-contact pair (TCP), in which the electron has substantial interactions with the solvent as well as a partial-valence interaction with the Na cation core (*16*, *17*). The steady-state absorption spectrum of the (Na^+, e^-) TCP peaks at 870 nm [Fig. 1D, green solid curve (*18*)], a wavelength that lies between the narrow gas-phase Na D line (peaking at 590 nm) and the broad absorption of the THF-solvated electron, e^-_{THF} [Fig. 1D, purple dotted curve, peaking at 2160 nm (*19*)], reflecting the fact that the chemical nature of this solute lies between the extremes of pure cation-electron and solvent-electron interactions. As a result, the TCP is formed from either a solvent-separated $Na^+ + e^-_{THF}$ pair or a weakly solvated, gas-phase–like Na^0 atom by partial transfer of an electron toward or away from the Na^+ core, respectively

$$(Na^+\cdots e^-) \xrightarrow{\text{partial ET}} (Na^+, e^-) \xleftarrow{\text{partial ET}} Na^0 \quad (3)$$

In our experiments, we used two different photoinduced ET processes (Fig. 1, A to C) to create the species on both the left and right of Eq. 3 and then track the solvation dynamics associated with the spontaneous partial ET reactions that generate (Na^+, e^-) as a result of the subsequent solvent-solute relaxation.

The THF-solvated (Na^+, e^-) TCP can be generated readily from its consituent parts, Na^+ and e^- (left side of Eq. 3) (*18*), but under most conditions it is challenging to measure the dynamic solvation of a solute produced through electron attachment, because the rate of attachment is generally limited by the mutual diffusion of the reactants. However, by taking advantage of the relatively strong ion-pair interactions that exist in weakly polar solvents such as THF, we can locate an electron-donating chromophore (here, iodide) close to the Na cation and thus photoinitiate prompt electron attachment. In previous work, we showed that electrons generated through charge-transfer–to–solvent (CTTS) excitation of counterion-free iodide in THF were ejected ~6 nm away from the iodine core (*20*), but that the presence of nearby Na^+ collapses the CTTS-electron ejection distribution within ~2 nm of the cation, so that >50% of the electrons attach to Na^+ to form (Na^+, e^-) TCPs within 2 ps (*21*). Consequently, the Na^+-I^- ion pair in THF provides an ideal precursor to create our (Na^+, e^-) TCP solvation probe from its constituent parts via excitation of the Na^+-I^- CTTS band (Fig. 1D, red dot-dashed curve) and to subsequently probe in detail the ultrafast solvation dynamics associated with electron attachment, $S_+(t)$.

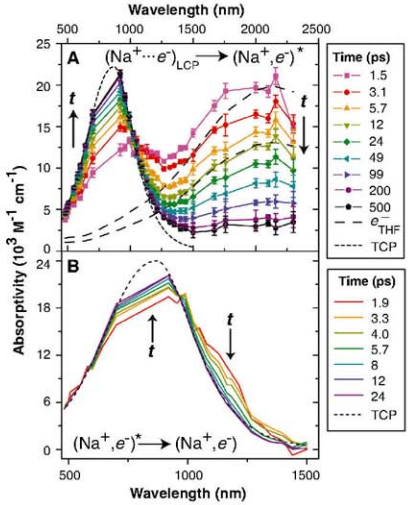

Fig. 2. Ultrafast transient absorption dynamics after the 263-nm CTTS excitation of Na^+-I^- in liquid THF. (**A**) Spectral reconstruction (SOM text) showing the dynamics of Na^+-e^- attachment to form the (Na^+, e^-) TCP; the symbols represent the data, which are connected with colored lines to guide the eye (the large linear region near 800 nm results from difficulty in probing near the laser fundamental with our transient absorption setup; error bars represent 95% confidence limits of measurements). The fact that the early-time data do not match the equilibrium spectrum of e^-_{THF} (shown as the long-dashed black curves that have been scaled for ease of comparison to the data at 1.5 and 12 ps) near ~1700 nm indicates that some of the CTTS-generated electrons rapidly attach to Na^+ to form Na cation-electron LCPs. The short-dashed black curve shows the equilibrium (Na^+, e^-) TCP spectrum for reference. (**B**) Spectral dynamics associated with the (Na^+, e^-) TCP created after electron attachment [colored lines, as in (A)]; the spectrum of this species has been isolated by subtracting the contributions of e^-_{THF} and $(Na^+\cdots e^-)_{LCP}$ (SOM text). The data show that after the LCP → TCP partial ET reaction, dynamic solvation results in a spectral blue shift of the nonequilibrium TCP absorption that occurs on a ~5-ps time scale. Different wavelength scales are used in (A) and (B).

Figure 2A presents the temporal evolution of the transient absorption that follows the 263-nm CTTS excitation of 20-mM NaI in THF [see the supporting online material (SOM) text for details (*22*)]. The data show that CTTS excitation rapidly induces an absorption in the 1200- to 2300-nm spectral region. This absorption then decays over the course of 200 to 300 ps concurrently with the appearance of a new absorption band in the 480- to 900-nm region. The absorption dynamics at intermediate wavelengths (900 to 1200 nm) evolve with more complexity, although a quasi-isosbestic point appears near 1130 nm. To first order, these spectral dynamics reflect the disappearance of CTTS-generated THF-solvated electrons (e^-_{THF}) as they diffusively attach to nearby Na^+ in solution to form the (Na^+, e^-) TCP (*21*).

A closer examination of Fig. 2A, however, reveals that the spectrum measured between 1200 and 2300 nm immediately after CTTS excitation of NaI (at 1.5 ps, pink trace) does not match that of the equilibrated THF-solvated electron (scaled and plotted as long-dashed curves). We have observed a similar spectrum (that is, slightly blue-shifted relative to the spectrum of e^-_{THF}) after the CTTS excitation of tetrabutylammonium iodide (t-BA^+-I^-) in THF (*21*), and we determined that the weak ion-pair interactions between I^- and t-BA^+ promote the formation of a loose cation-electron contact pair (LCP). The LCP is characterized by a weak Coulombic cation-electron attraction (not the stronger partial valence interactions that define the TCP), and its spectrum can be thought of as arising from a Stark shift of the e^-_{THF} absorption induced by the presence of the nearby cation. For the Na^+-I^- system studied here, the spectrum between 1200 and 2300 nm (Fig. 2A) rapidly decays in intensity over the initial few picoseconds as TCPs are formed, with the remaining feature matching the shape of the e^-_{THF} spectrum by ~12 ps (light green trace). Thus, the data in Fig. 2A demonstrate the transient formation of a sizeable Na^+-e^- LCP population after CTTS excitation of Na^+-I^- in THF (*21*, *22*)

$$I^- + Na^+ \xrightarrow{h\nu,\ CTTS} I + (Na^+\cdots e^-)_{LCP}\ (\sim 60\%)$$

$$\text{or } I + Na^+ + e^-_{THF}\ (\sim 40\%) \qquad (4a)$$

$$(Na^+\cdots e^-)_{LCP} \xrightarrow[2.3\ ps]{\text{partial ET}} (Na^+, e^-)^* \xrightarrow{S_+(t)} (Na^+, e^-) \qquad (4b)$$

$$Na^+ + e^-_{THF} \xrightarrow[63\ ps]{\text{diffuse}} (Na^+\cdots e^-)_{LCP} \xrightarrow{\text{via eq. 4b}} (Na^+, e^-) \qquad (4c)$$

$(Na^+, e^-)^*$ denotes a nonequilibrated TCP solute, and I is a spectator in Eqs. 4b and 4c [(SOM text) (*22*)]. As a result, CTTS excitation of NaI allows us to rapidly create a population of LCPs, which then spontaneously convert to a nonequilibrium $(Na^+, e^-)^*$ TCP population via partial ET to Na^+.

We can isolate the spectral evolution associated with TCP solvation by subtracting the spectral contributions from the LCP and e^-_{THF} [the latter formed in ~40% yield after CTTS ejection from iodide to locations far from a Na^+ cation (Eq. 4a)], and by normalizing for TCP population kinetics, as plotted in Fig. 2B [(SOM text) (*22*)]. Although the first ~2 ps of the TCP spectral dynamics are obscured by the LCP → TCP interconversion process as well as by experimental resolution effects (*21*), the data clearly show that the initially produced $(Na^+, e^-)^*$ species (at 1.9 ps) has a spectrum that is red-shifted and slightly broadened relative to the equilibrium absorption (black dashed curve). As the solvent relaxes, the transient spectrum of the TCP blue-shifts and narrows on a ~5-ps time scale, with solvation complete by 10 ps. Because the TCP solvation dynamics we measured may be rate-limited by the LCP → TCP reaction process (the first step of Eq. 4b), this ~5-ps time scale represents an upper limit for the decay of $S_+(t)$.

To examine the nonequilibrium solvation of the same (Na^+, e^-) TCP solute from a different initial solvent-solute configuration (Fig. 1, A to C, right panels), we also created the TCP in a solvent environment that initially accommodates its corresponding anion, Na^-. Like I^-, Na^- in liquid THF has a strong CTTS transition [Fig. 1D, blue dashed curve (*23*)], photoexcitation of which leads to subpicosecond electron ejection and spectral dynamics of this species have been isolated by subtracting the contribution from Na^0 and have been normalized to remove contributions from changes in TCP population [(*16*) and SOM text]. The data demonstrate that after the $Na^0 \rightarrow (Na^+, e^-)$ partial ET reaction, dynamic solvation results in a substantial red shift of the nonequilibrium TCP absorption that occurs on a ~10-ps time scale.

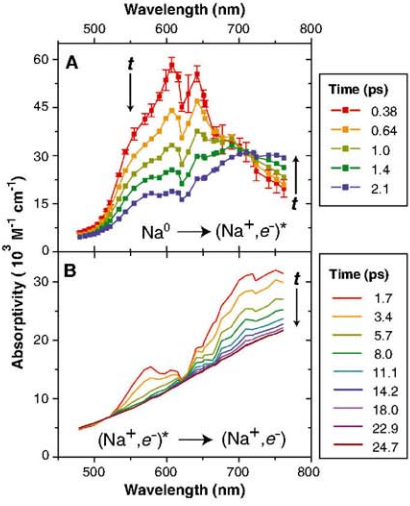

Fig. 3. Spectrally reconstructed ultrafast transient absorption dynamics of the neutral Na species created after 395-nm CTTS excitation of Na^- in liquid THF [(*16*) and SOM text (*22*)]. (**A**) Spectral dynamics (squares connected by colored lines to guide the eye; error bars represent 95% confidence limits of measurements) associated with the chemical conversion of a weakly solvated Na atom, characterized by the split and broadened Na D-line transition near 600 nm (0.38 ps), to become the THF-solvated (Na^+, e^-) TCP; the isosbestic point near 720 nm indicates that Na^0 and (Na^+, e^-) TCPs are chemically distinct species that interconvert via a spontaneous partial ET reaction. (**B**) Spectral dynamics associated with the (Na^+, e^-) TCP after electron detachment (colored lines); the

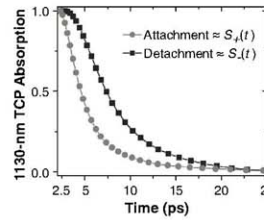

Fig. 4. Ultrafast spectral dynamics, probed at 1130 nm, of the (Na^+, e^-) TCP created after both electron attachment via 263-nm CTTS excitation of Na^+-I^- [circles, $S_+(t)$] and electron detachment via 395-nm CTTS excitation of Na^- [squares, $S_-(t)$]. The two spectral transients have been normalized at 2 ps and the detachment data have been inverted for ease of comparison. To the extent that probing at a single wavelength characterizes the entire solvent response (*26*), the data show that solvation of the nonequilibrium TCP occurs roughly twice as fast after electron attachment than after electron detachment (1/*e* time scales of 3.9 and 7.5 ps, respectively), thus demonstrating a breakdown of LR.

leaves behind a neutral atomic Na core. Because of the cavities that are inherent in the structure of liquid THF (*12*, *13*), the distance to which the CTTS-generated electrons are ejected from the neutral Na core increases with increasing excitation energy (*24*). Thus, by exciting the Na^- CTTS transition at 395 nm, we are able to move the CTTS-generated electron far from the Na^0 core, so that the spectral dynamics we measure for Na^0 are negligibly affected by the proximity of (or possible recombination with) the CTTS-ejected electron (*16*). The spectral dynamics associated with the Na^0 product generated via 395-nm excitation of the Na^- CTTS band, corrected for the modest Na^0-e^-_{THF} recombination that does take place, are shown in Fig. 3 (*16*, *22*).

The spectrum of the nascent neutral Na atom measured immediately after electron photodetachment from Na^- (at 0.38 ps) exhibits a broad, intense absorption band that peaks between 600 and 650 nm. We (*16*) and others (*25*) have assigned this feature to the sodium D-line transition, albeit broadened and split by the asymmetry of the surrounding solvent cavity. The similarity in the position of this band to that measured in the gas phase reflects the weak interaction of the nascent neutral atom with the surrounding solvent; this is expected because the initial solvent structure had accommodated equilibrated Na^-, which is much larger than the neutral Na atom (Fig. 1A, right) (*16*). Figure 3A further reveals that this initially produced D-line absorption feature decays substantially over the next ~2 ps and that the decay is accompanied by a concomitant increase in spectral intensity at longer wavelengths (>700 nm); a well-defined isosbestic point at ~720 nm indicates that these spectral changes result from a kinetic interconversion between two chemically distinct species. Careful analysis of this spectral progression reveals that the neutral Na solute starts from a state in which the 3*s* valence electron is bound almost exclusively by the Na^+ core but then undergoes a spontaneous partial ET reaction on a ~730-fs time scale to produce a nonequilibrated (Na^+, e^-)* TCP in which the electron is subject to substantial interactions with the solvent (*16*)

$$Na^- \xrightarrow{h\nu,\ CTTS} Na^0 + e^-_{THF} \qquad (5a)$$

$$Na^0 \xrightarrow[730\ fs]{partial\ ET} (Na^+, e^-)^* \xrightarrow{S_-(t)} (Na^+, e^-) \qquad (5b)$$

After this partial ET from the nascent neutral atom (which was produced via photoinduced CTTS detachment from Na^-), (Na^+, e^-)* relaxes with the surrounding solvent to reach equilibrium, $S_-(t)$.

Figure 3B highlights the spectral evolution of the TCP during the $S_-(t)$ relaxation, which has been isolated by removing contributions from the Na^0 absorption at 600 nm and normalized for its time-dependent population kinetics (*16*, *22*). The salvation induced evolution of the (Na^+, e^-) spectrum occurs more slowly than the partial ET reaction that forms this species from Na^0. Figure 3B shows that the time-dependent TCP spectrum red shifts (and broadens to maintain a constant oscillator strength) as it relaxes to equilibrium. The data show that this dynamic solvation process takes place on a ~10-ps time scale and that the $S_-(t)$ relaxation is not complete for at least 20 to 25 ps.

One way to approximately quantify the $S_+(t)$ and $S_-(t)$ TCFs inherent to the data in Figs. 2B and 3B is to compare the dynamics at a single wavelength (*26*). Thus, in Fig. 4, we compare the time-dependent TCP spectral dynamics at 1130 nm for both solvation processes. We chose this wavelength for two reasons. First, 1130 nm lies near the isosbestic point for the TCP and LCP/e^-_{THF} absorption bands; thus, for the photoinduced electron-attachment pathway, this wavelength probes primarily TCP solvation and not population dynamics, guaranteeing that the kinetic model we used to deconstruct our data does not influence the $S_+(t)$ solvation response we derive at this wavelength [(SOM text) (*22*)]. Second, the long-time TCP solvation after electron photodetachment is observed most cleanly at this wavelength, because there is no interference from Na^- bleach and the e^-_{THF} absorption contribution can be removed readily (*16*). To directly compare the 1130-nm spectral dynamics from the two pathways, we have inverted the $S_-(t)$ response from the photodetachment pathway, because the TCP spectral red shift results in a delayed rise (rather than a decay) at this wavelength (*16*). We also have normalized the 1130-nm spectral transients at 2 ps in order to focus on the long-time solvation dynamics and to exclude contributions from the faster ET interconversion kinetics observed along both relaxation pathways. Figure 4 indeed illustrates that, as estimated by the 1130-nm spectral dynamics, $S_+(t)$ decays roughly twice as fast as $S_-(t)$ (1/*e* time scales of 3.9- and 7.5-ps, respectively). Because we know that the $S_+(t)$ response is partly rate-limited by the LCP → TCP conversion process, whereas the $S_-(t)$ response occurs more slowly than the $Na^0 \rightarrow (Na^+, e^-)$ reaction time scale, this difference in the two responses represents a lower limit.

Together, Figs. 2B, 3B, and 4 prove that LR does not apply for the solvation dynamics that accompany the creation of the (Na^+, e^-) TCP through the two different partial ET reactions. The $S_+(t)$ solvation response that follows the creation of the TCP via partial ET from a LCP [which was created via electron photoattachment (Eq. 4)] starts from a solvent configuration that was optimized to accommodate the small Na^+, prompting a blue shift of the TCP spectrum that is complete within ~10 ps; this time scale is similar to the solvent relaxation measured after the excitation of a dye molecule in liquid THF (*27*). In contrast, the $S_-(t)$ solvation response that follows the creation of the TCP via partial ET from Na^0 (Eq. 5) starts from a solvent configuration that was optimized to accommodate the large Na^- anion, prompting a red shift of the TCP spectrum that is not complete for 20 to 25 ps. This large difference in time scale in the solvation dynamics of the same species approaching equilibrium from two different initial configurations is a clear indication that the LR approximation, which predicts identical $S_+(t)$ and $S_-(t)$ responses, does not describe the solvation dynamics that follow these simple partial electron-transfer reactions.

Why should we observe this breakdown of LR? Guided by results of MD simulations that have demonstrated a failure of LR similar to what we observed here (*6*), we believe that the changes in solute size involved in these partial charge-transfer reactions help promote the breakdown of LR. These simulations revealed that when the solute size decreases (as occurs upon the removal of an electron), the rate of the solvent response is essentially limited by the translational diffusion of first-shell solvent molecules into the void space left behind after the solute shrinks. Thus, we expect LR to break down in this case both because the initial solvent configuration about the smaller solute is never explored through equilibrium fluctuations and because the corresponding diffusional solvent motions that relax this configuration are slower than those at equilibrium. In contrast, the simulations also have shown that the solvent response that accompanies a solute size increase (as occurs upon the addition of an electron) is driven rapidly by strong, short-ranged repulsive interactions between the enlarged solute and the first-shell solvent molecules. The fact that $S_-(t)$ is substantially slower than $S_+(t)$ strongly suggests that the relative changes in size of the initial Na^0 and LCP as they become the TCP are primarily responsible for the breakdown of LR that we observed in our experiments.

To what extent do the processes we have studied reflect the solute-solvent relaxation in more common chemical transformations? Is the breakdown of LR we observed in this model system the rule, and not an exception, for the nonequilibrium solvation that occurs in the course of typical chemical processes? In addition to our observations, a breakdown of LR has recently been observed to follow the photodissociation of the ICN molecule in ethanol (*28*), which produces a CN fragment whose excited rotational motion is only slowly quenched by interactions with the surrounding solvent. This behavior demonstrates an anticipated failure of LR in scenarios in which a chemical process partitions substantial internal energy (many k_BT) into a product fragment (*7*, *28*). Even though the photochemical preparation in our experiments involves photons with energies of hundreds of k_BT, most of this energy is used to detach an electron from either Na^- or I^- and move it through liquid THF; the bulk of this energy is not partitioned directly to the TCP product. Moreover, our experiments measure the TCP solvation only after spontaneous partial ET

reactions (Eqs. 3 to 5). Thus, although we may anticipate that the solvation dynamics measured involve substantial changes in the local solvation structure due to the solute size change, we expect that the energies associated with $(Na^+, e^-)^*$ formation are not chemically extreme and are representative of those associated with common solution-phase reactions.

This clear breakdown of LR implies that the solvent fluctuations coupled to the (Na^+, e^-) TCP are not Gaussian, and thus that the potential surfaces associated with these ET processes are highly nonparabolic. As a result, the Marcus theory of ET would poorly describe these ET processes. We anticipate that this could be an important consideration in many similar outer-sphere ET reactions, where substantial rearrangement of the local solvent structure could induce a similar LR breakdown. It is also important to note that the LR approximation is built on the idea that the same solvent-solute motions that underlie equilibrium fluctuations are also responsible for the nonequilibrium solvation dynamics (*3*). Although we observed a clear difference in the time dependence of two solvation pathways that reflects a breakdown of LR, observing an identical time dependence would not have guaranteed that the LR holds. This is because even when the specific molecular motions responsible for relaxing a nonequilibrium perturbation differ considerably from the solvent fluctuations active at equilibrium, LR may appear to be valid if the relevant nonequilibrium and equilibrium solvent motions happen to occur on similar time scales; what we have termed a hidden breakdown of LR (*29*, *30*). Overall, these findings demonstrate that an accurate assessment of solvation dynamics—and, by extension, our understanding of solution-phase chemical reactivity—must be considered directly at the molecular level in order to determine correctly how best to understand the solvent relaxation resulting from a given nonequilibrium perturbation.

References and Notes

1. R. A. Marcus, N. Sutin, *Biochim. Biophys. Acta* **811**, 265 (1985).
2. M. Maroncelli, *J. Mol. Liq.* **57**, 1 (1993).
3. D. Chandler, *Introduction to Modern Statistical Mechanics* (Oxford Univ. Press, New York, 1987).
4. R. M. Stratt, M. Maroncelli, *J. Phys. Chem.* **100**, 12981 (1996).
5. P. L. Geissler, D. Chandler, *J. Chem. Phys.* **113**, 9759 (2000).
6. D. Ahlrichs, V. Tran, B. J. Schwartz, *J. Phys. Chem. B* **104**, 5382 (2000).
7. G. Tao, R. M. Stratt, *J. Chem. Phys.* **125**, 114501 (2006).
8. B. B. Laird, W. H. Thompson, *J. Chem. Phys.* **126**, 211104 (2007).
9. T. Fonseca, B. M. Ladanyi, *J. Phys. Chem.* **95**, 2116 (1991).
10. D. F. Underwood, D. A. Blank, *J. Phys. Chem. A* **109**, 3295 (2005).
11. D. S. Larsen, K. Ohta, Q. H. Xu, M. Cyrier, G. R. Fleming, *J. Chem. Phys.* **114**, 8008 (2001).
12. M. J. Bedard-Hearn, R. E. Larsen, B. J. Schwartz, *J. Chem. Phys.* **122**, 134506 (2005).
13. D. T. Bowron, J. L. Finney, A. K. Soper, *J. Am. Chem. Soc.* **128**, 5119 (2006).
14. M. J. Bedard-Hearn, R. E. Larsen, B. J. Schwartz, *J. Chem. Phys.* **125**, 194509 (2006).
15. I. B. Martini, E. R. Barthel, B. J. Schwartz, *Science* **293**, 462 (2001).
16. M. C. Cavanagh, R. E. Larsen, B. J. Schwartz, *J. Phys. Chem. A* **111**, 5144 (2007).
17. R. Catterall, J. Slater, M. C. R. Symons, *J. Chem. Phys.* **52**, 1003 (1970).
18. B. Bockrath, L. M. Dorfman, *J. Phys. Chem.* **77**, 1002 (1973).
19. F.-Y. Jou, G. R. Freeman, *Can. J. Chem.* **57**, 591 (1979).
20. A. E. Bragg, B. J. Schwartz, *J. Phys. Chem. B* **112**, 483 (2008).
21. A. E. Bragg, B. J. Schwartz, *J. Phys. Chem. A* **112**, 3530 (2008).
22. Methods are detailed in supporting material available on *Science* Online.
23. M. T. Lok, J. L. Dye, F. J. Tehan, *J. Phys. Chem.* **76**, 2975 (1972).
24. E. R. Barthel, B. J. Schwartz, *Chem. Phys. Lett.* **375**, 435 (2003).
25. O. Shoshana, J. L. Pérez Lustres, N. P. Ernsting, S. Ruhman, *Phys. Chem. Chem. Phys.* **8**, 2599 (2006).
26. V. Nagarajan, A. M. Brearley, T.-J. Kang, P. F. Barbara, *J. Chem. Phys.* **86**, 3183 (1987).
27. L. Reynolds, J. A. Gardecki, S. J. V. Frankland, M. L. Horng, M. Maroncelli, *J. Phys. Chem.* **100**, 10337 (1996).
28. A. C. Moskun, A. E. Jailaubekov, S. E. Bradforth, G. Tao, R. M. Stratt, *Science* **311**, 1907 (2006).
29. M. J. Bedard-Hearn, R. E. Larsen, B. J. Schwartz, *J. Phys. Chem. A* **107**, 4773 (2003).
30. M. J. Bedard-Hearn, R. E. Larsen, B. J. Schwartz, *Phys. Rev. Lett.* **97**, 130403 (2006).

31. This research was funded by NSF under grant number CHE-0603766. The authors thank R. E. Larsen for useful discussions and a critical reading of the manuscript.

Supporting Online Material

www.sciencemag.org/cgi/content/full/321/5897/1817/DC1
SOM Text
Figs. S1 to S4
References

9 June 2008; accepted 18 August 2008
10.1126/science.1161511

Mars' Paleomagnetic Field as the Result of a Single-Hemisphere Dynamo

Sabine Stanley,[1]* Linda Elkins-Tanton,[2] Maria T. Zuber,[2] E. Marc Parmentier[3]

Mars' crustal magnetic field was most likely generated by dynamo action in the planet's early history. Unexplained characteristics of the field include its strength, concentration in the southern hemisphere, and lack of correlation with any surface features except for the hemispheric crustal dichotomy. We used numerical dynamo modeling to demonstrate that the mechanisms proposed to explain crustal dichotomy formation can result in a single-hemisphere dynamo. This dynamo produces strong magnetic fields in only the southern hemisphere. This magnetic field morphology can explain why Mars' crustal magnetic field intensities are substantially stronger in the southern hemisphere without relying on any postdynamo mechanisms.

The Mars Global Surveyor mission showed that Mars possesses remanent crustal magnetic fields from a dynamo that was operational for a short time in Mars' early history (*1*). Remanent crustal magnetism is observed in early Noachian (>3.9 billion years old) crust in both the northern and southern hemispheres, except for much of the Tharsis volcanic province and the large impact basins Hellas and Argyre in the southern hemisphere, and Isidis and Utopia in the northern hemisphere. There is a conspicuous difference in the magnetic field intensities in the two hemispheres: The northern hemisphere contains only weak magnetic fields, whereas the southern hemisphere contains both strong and weak fields (*2*).

The timing of the dynamo is constrained by the observations that the floors of the large impact basins formed during the Late Heavy Bombardment [~3.9 billion years ago (Ga)] are not magnetized (*1*) and that the ancient Martian meteorite ALH84001 contains a remanent magnetic field dated earlier than 3.9 Ga (*3*). Most likely, the dynamo was active sometime between core formation (~4.5 Ga) and the Late Heavy Bombardment. The driving force for the dynamo, the intensity and morphology of the generated field, and the cause of the dynamo's demise are not well understood.

Another ancient Martian crustal feature is the hemispheric dichotomy. The northern and southern hemispheres have similar-aged crusts (*4*) but different topographies, thicknesses, and sediment covers (*5*). The northern hemisphere crust is low, thin, and covered with volcanic flows and sediments, whereas the southern hemisphere crust is high, thick, and largely devoid of sedimentary or volcanic resurfacing. Cratering evidence and the dichotomy's long wavelength suggest that the dichotomy is an ancient feature, directly related to crustal formation sometime between 4.5 and 3.9 Ga (*6*, *7*).

Because the crustal magnetic field and the dichotomy are similar in age, it is possible that their formation processes are related. Several endogenic mechanisms could explain both dichotomy formation and a concurrent dynamo sometime between 4.5 and 3.9 Ga. A hemispheric-scale (degree-1) pattern of mantle circulation resulting from either mantle convection in the presence of

[1]Department of Physics, University of Toronto, Toronto, ON M5S1A7, Canada. [2]Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [3]Department of Geological Sciences, Brown University, Providence, RI 02912, USA.

*To whom correspondence should be addressed. E-mail: stanley@physics.utoronto.ca

radial viscosity variations (*8*, *9*), early magma ocean crystallization resulting in overturn (*10*, *11*), or superplumes resulting from destabilization of the mantle lower thermal boundary layer (*12*) provides degree-1 crustal structure along with sufficiently vigorous core convection to sustain a short-lived dynamo. Effects from a very large impact (*13*) or several large impacts (*14*) were also suggested early on but encountered difficulties (*15*, *16*). The probability of several large impacts forming the northern lowlands is low, and there is no evidence of individual basins. A single very large impact is statistically possible. Presently, the dichotomy boundary is not circular or elliptical; however, a recent analysis of crustal thickness (*17*) demonstrated that its original shape was elliptical. Therefore, the dichotomy boundary and northern lowlands could be the result of a giant low-angle impact.

Table 1. Nondimensional parameter values in the dynamo model. Model values are given for the Prandtl number ($Pr = \upsilon/\kappa$, where υ is the kinematic viscosity and κ is the thermal diffusivity); a magnetic Prandtl number ($q_\kappa = \kappa/\eta$ where η is the magnetic diffusivity); the Ekman number [$E = \upsilon/(2\Omega r_o^2)$, where Ω is the angular rotation rate of the planet and r_o is the core radius]; and the modified Rayleigh number [$Ra = \alpha g_o h_T r_o^2/(2\Omega\eta)$, where α, g_o, and h_T are the thermal expansion coefficient, the gravitational acceleration at the CMB, and the prescribed superadiabatic temperature gradient at the inner core boundary, respectively]. Using representative Mars values of $\alpha = 10^{-5}$ K^{-1}, $g_o = 3.5$ m s^{-2}, $r_o = 1700$ km, $\Omega = 7.1 \times 10^{-5}$ s^{-1}, and $\eta = 2$ m s^{-2}, combined with the chosen value for the Rayleigh number, implies a superadiabatic temperature gradient at the inner core boundary of 5×10^{-8} K m^{-1}, corresponding to a superadiabatic temperature gradient at the CMB of $h_T r_{io}^2 = 6 \times 10^{-9}$ K m^{-1} (where r_{io} is the inner to outer core radius ratio).

Parameter	Value
Pr	1
q_κ	1
E	2×10^{-5}
Ra	18,000

Although explanations for dichotomy formation with concurrent magnetic field generation appear feasible, a serious problem is the markedly different intensities of the crustal magnetic fields in the northern and southern hemispheres. If the dynamo produced an axial-dipole-dominated magnetic field, similar to the geomagnetic field, and the northern and southern crusts formed at similar times with similar magnetic mineral densities and magnetic layer thicknesses, then one would expect crustal fields of similar strength in both hemispheres. Efforts to explain the hemispheric magnetic intensity differences generally involve postdynamo mechanisms in the northern hemisphere, such as hydrothermal alteration (*6*) or demagnetization resulting from early large impacts (*15*).

Here, we show that dynamo generation can also explain the hemispheric magnetic intensity differences, thereby removing the requirement for a postdynamo solution. All endogenic mechanisms involve hemispheric-scale mantle circulation, which will necessarily produce a degree-1 temperature anomaly in the mantle and hence at the core-mantle boundary (CMB). Numerical models have demonstrated that an exogenic (giant impact) mechanism could also produce a degree-1 temperature anomaly in the mantle and at the CMB (*18*). Because the CMB is the outer-bounding surface of the dynamo region, this temperature anomaly will result in a hemispheric heat flux variation on the outer boundary of the core. We therefore imposed a degree-1 variable heat flux pattern at the CMB in a dynamo simulation.

We used the numerical dynamo model of Kuang and Bloxham (*19*–*21*) with the parameter values given in Table 1. We imposed a heat flux across the CMB that was lower in the northern hemisphere than in the southern hemisphere (fig. S1). The heat flux variation on the CMB is relatively large, with the root mean square of the lateral variations three times the average superadiabatic heat flux. This may be reasonable for Mars when considering the relative temperatures of downwellings and upwellings at the CMB (*11*).

This spatially variable heat flux boundary condition produces a stable one-hemisphere dynamo. The radial component of the magnetic field is strongest and concentrated in the southern hemisphere with only weak fields in the northern hemisphere (Fig. 1A). Oscillatory hemispheric dynamos have been found in certain parameter regimes (*22*), producing a strong field in each hemisphere periodically. The dynamo in our simulation is different in that fields are actively gen-

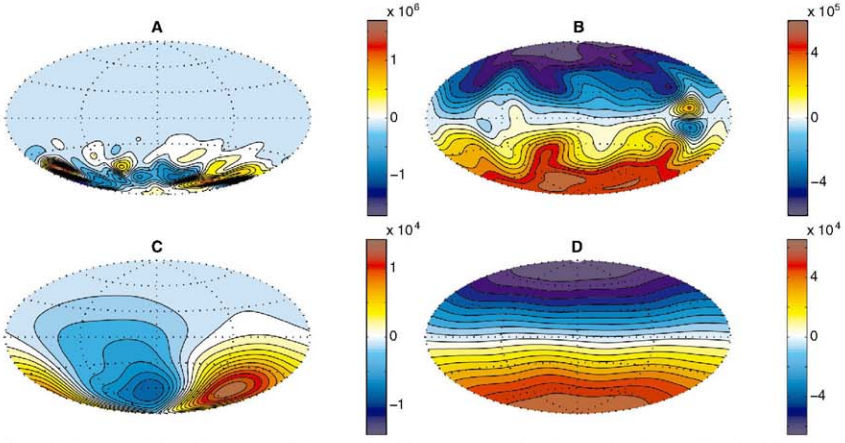

Fig. 1. Filled contours of the radial component of the magnetic field. For a model with a degree-1 heat flux outer-boundary condition, the field is plotted at the CMB in (**A**) and at the surface in (**C**). For a homogeneous heat flux outer-boundary condition, the field is plotted at the CMB in (**B**) and at the surface in (**D**). The CMB radius is 1700 km, the surface radius is 3400 km, and the units are nT.

erated in only a single hemisphere. The fields are smaller-scale and, although highly time-variable in morphology, remain strong in only the southern hemisphere (fig. S2). A dynamo model with a similar variable heat flux boundary condition as our model but with a different choice of velocity boundary condition does not produce a single-hemisphere dynamo (*23*), most likely because of the different force balances in the models [supporting online material (SOM) text].

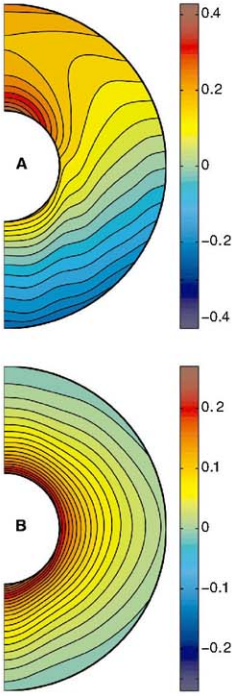

Fig. 2. Filled contours of the axisymmetric part of the nondimensional superadiabatic temperature. The profile for a model with a degree-1 heat flux outer-boundary condition is shown in **(A)** and for a model with a homogeneous heat flux outer-boundary condition in **(B)**. The temperature plotted is with respect to the average CMB temperature, so positive values are hotter than the average CMB temperature and negative values are colder than the average CMB temperature. To dimensionalize in Kelvin, nondimensional values should be multiplied by 0.085.

The radial field at the CMB in our model was more intense than that in a model with the same parameter values but with a homogeneous heat flux boundary condition that produced an axial-dipolar dynamo (Fig. 1B). However, because the power in the field components fell off faster with distance for smaller scales, the magnetic field at the surface was slightly weaker in our variable heat flux model than in the homogeneous heat flux model but was the same order of magnitude (Fig. 1, C and D).

The hemispheric boundary condition in the model changed the equatorial symmetry of the superadiabatic temperature (Fig. 2), affecting the dominant force balance in the core. This had a substantial effect on the velocity fields of the core in our model (Fig. 3A). The zonal flows, mainly resulting from thermal winds, were equatorially antisymmetric and therefore did not adhere to the Taylor-Proudman constraint, which is expected to hold when Coriolis and pressure forces dominate in the core (*24*). In addition, the meridional circulation pattern was concentrated in the southern hemisphere rather than filling the whole core, as in a model with homogeneous heat flux boundary conditions (Fig. 3B). The convection rolls were concentrated in the hemisphere with the colder CMB temperature and generated the strongest dynamo action in this region.

Our results suggest that the concentration of strong crustal fields in the southern hemisphere of Mars could result from a dynamo that produced a magnetic field concentrated in the southern hemisphere. In this scenario, no postdynamo process is required to remove a strong crustal field in the northern hemisphere. Although large basins are demagnetized in the northern hemisphere (as they are in the southern hemisphere), this mechanism can explain why none of the magnetized regions

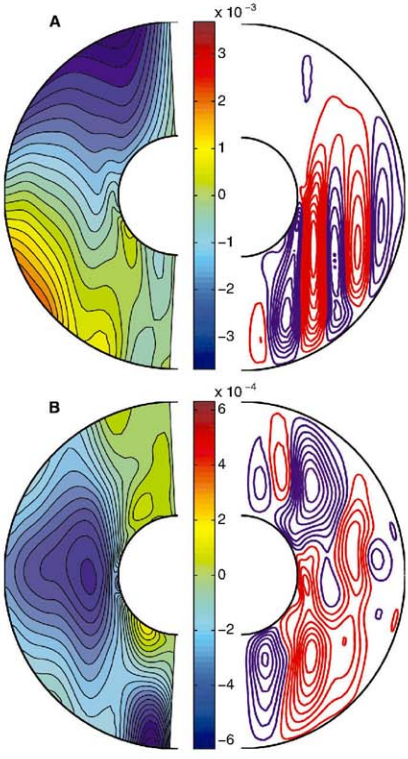

Fig. 3. Axisymmetric component of the nondimensional velocity field. Filled contours of the zonal velocity are shown on the left, and stream lines of the meridional circulation are shown on the right for a model with a degree-1 heat flux outer-boundary condition **(A)** and a model with a homogeneous heat flux outer-boundary condition **(B)**. For the zonal velocity, red indicates prograde zonal flow, and blue, retrograde zonal flow. For the meridional circulation, red indicates prograde circulation, and blue, retrograde circulation. The color bars apply to the zonal velocity figures only, and the units are m s^{-1}. The intensity of the meridional circulation is prescribed by the spacing of the stream lines.

in the northern hemisphere are as strong as the regions in the south.

In addition to explaining the occurrence of a strong field in only one hemisphere, our model is also able to explain the conflicts between various Mars paleomagnetic studies and rotational stability studies. The inferred paleomagnetic pole positions vary in location depending on the individual crustal anomaly used (*25*, *26*). Some of the paleopoles are also located in equatorial regions near the Tharsis bulge and hence are far from the current geographic poles. This has been interpreted as evidence for a large true polar wander event that relocated Tharsis from polar to equatorial regions in early Mars history (*27*). However, large Tharsis-driven true polar wander is in conflict with rotational stability studies (*28*, *29*) that demonstrate that the present-day gravitational figure of Mars favors a small Tharsis-driven true-polar wander scenario.

An assumption made in the paleomagnetic studies is that the dynamo-generated magnetic field was axial-dipolar dominated. This assumption implies that the magnetic pole coincided with the rotation pole and is used extensively in Earth paleomagnetic studies. Our models would dictate that the Mars dynamo-generated field was not axial-dipolar dominated and hence that the magnetic poles would not coincide with the rotation poles, rendering paleopole interpretations useless. In addition, because our dynamo-generated fields are multipolar, individual crustal magnetic fields at different locations can point to different paleomagnetic poles, thereby explaining the discrepancies in the different paleomagnetic studies.

A single-hemisphere dynamo also has implications for evolution of the martian atmosphere. A strong dynamo-generated magnetic field can more easily explain the intense crustal magnetism. However, efficient atmospheric erosion, necessary to explain the loss of Mars' early thick atmosphere, favors a weak internal magnetic field (*30*). Our single-hemisphere dynamo may provide an elegant solution to this problem because the northern hemisphere would be prone to atmospheric removal early in solar system history when the young Sun was more active (*31*), but the southern hemisphere could still possess a strong magnetic field in which the crustal rocks could magnetize.

References and Notes

1. M. H. Acuna *et al.*, *Science* **284**, 790 (1999).
2. B. Langlais, M. E. Purucker, M. Mandea, *J. Geophys. Res.* **109**, E02008 (2004).
3. B. P. Weiss *et al.*, *Earth Planet. Sci. Lett.* **201**, 449 (2002).
4. L. A. Edgar, H. V. Frey, *Geophys. Res. Lett.* **35**, L02201 (2008).
5. M. T. Zuber *et al.*, *Science* **287**, 1788 (2000).
6. S. C. Solomon *et al.*, *Science* **307**, 1214 (2005) and references therein.
7. H. V. Frey, *J. Geophys. Res.* **111**, E08S91 (2006).
8. S. Zhong, M. T. Zuber, *Earth Planet. Sci. Lett.* **189**, 75 (2001).
9. J. H. Roberts, S. Zhong, *J. Geophys. Res.* **111**, E06013 (2006).
10. L. T. Elkins-Tanton, E. M. Parmentier, P. C. Hess, *Meteorit. Planet. Sci.* **38**, 1753 (2003).
11. L. T. Elkins-Tanton, S. E. Zaranek, E. M. Parmentier, P. C. Hess, *Earth Planet. Sci. Lett.* **236**, 1 (2005).
12. Y. Ke, V. S. Solomatov, *J. Geophys. Res.* **111**, E10001 (2006).
13. D. E. Wilhelms, S. W. Squyres, *Nature* **309**, 138 (1984).
14. H. Frey, R. A. Shultz, *Geophys. Res. Lett.* **15**, 229 (1988).
15. F. Nimmo, M. S. Gilmore, *J. Geophys. Res.* **106**, 12315 (2001).
16. T. R. Watters, P. J. McGovern, R. P. Irwin III, *Annu. Rev. Earth Planet. Sci.* **35**, 621 (2007).
17. J. C. Andrews-Hanna, M. T. Zuber, W. B. Banerdt, *Nature* **453**, 1212 (2008).
18. W. A. Watters, M. T. Zuber, B. H. Hager, *J. Geophys. Res.*, in press.
19. W. Kuang, J. Bloxham, *Nature* **389**, 371 (1997).
20. W. Kuang, J. Bloxham, *J. Comput. Phys.* **153**, 51 (1999).
21. Additional information on the numerical methods is available as supporting material on *Science* Online.
22. E. Grote, F. H. Busse, *Phys. Rev. E* **62**, 4457 (2000).
23. G. A. Glatzmaier, R. S. Coe, L. Hongre, P. H. Roberts, *Nature* **401**, 885 (1999).
24. M. Kono, P. H. Roberts, *Rev. Geophys.* **40**, 1013 (2002).
25. Y. Quesnel, B. Langlais, C. Sotin, *Planet. Space Sci.* **55**, 258 (2007), and references therein.
26. B. Langlais, M. Purucker, *Planet. Space Sci.* **55**, 270 (2007); and references therein.
27. K. F. Sprenke, L. L. Baker, A. F. Williams, *Icarus* **174**, 486 (2005); and references therein.
28. A. Daradich *et al.*, *Icarus* **194**, 463 (2008).
29. J. T. Perron, J. X. Mitrovica, M. Manga, I. Matsuyama, M. A. Richards, *Nature* **447**, 840 (2007).
30. V. Dehant *et al.*, *Space Sci. Rev.* **129**, 279 (2007).
31. Y. N. Kulikov *et al.*, *Space Sci. Rev.* **129**, 207 (2007).

32. S.S. is partially funded by the National Science and Engineering Research Council (NSERC) of Canada. The numerical simulations in this study were performed on supercomputing resources partially funded by the Canadian Foundation for Innovation (CFI) and the Ontario Research Fund (ORF).

Supporting Online Material

www.sciencemag.org/cgi/content/full/321/5897/1822/DC1
Materials and Methods
SOM Text
Figs. S1 and S2

29 May 2008; accepted 19 August 2008
10.1126/science.1161119

The Structure and Dynamics of Mid-Ocean Ridge Hydrothermal Systems

D. Coumou,* T. Driesner, C. A. Heinrich

Sub-seafloor hydrothermal convection at mid-ocean ridges transfers 25% of the Earth's heat flux and can form massive sulfide ore deposits. Their three-dimensional (3D) structure and transient dynamics are uncertain. Using 3D numerical simulations, we demonstrated that convection cells self-organize into pipelike upflow zones surrounded by narrow zones of focused and relatively warm downflow. This configuration ensures optimal heat transfer and efficient metal leaching for ore-deposit formation. Simulated fluid-residence times are as short as 3 years. The concentric flow geometry results from nonlinearities in fluid properties, and this may influence the behavior of other fluid-flow systems in Earth's crust.

Hydrothermal convection at mid-ocean ridge spreading centers transports a major part of Earth's total heat flux, substantially affects the chemistry of crust and overlying ocean, and provides nutrients for chemosynthetic life on and beneath the sea floor. Mass, heat, and associated chemical fluxes from the crust to the ocean at mid-ocean ridge spreading centers are large (*1*, *2*). Fundamental features of this flow, such as the location of seawater recharge and the relative importance of off- versus along-axis convection, are still uncertain. Recent studies of active (*3*) and ancient (*4*) systems show that discharge can be highly focused in pipelike regions, possibly continuing to the base of the hydrothermal system (*3*). Recharge is often thought to occur over extensive areas (*5*, *6*), with off-axis faults guiding fluid pathways toward the base of the hydrothermal system. A common alternative view is that of fluid circulation being restricted to a high-permeability along-axis zone (*7*, *8*). Micro-earthquake data indicate that recharge can be focused close to the spreading center in some systems (*9*).

Recent two-dimensional (2D) numerical studies that included accurate thermodynamic properties of water have shown that the nonlinear dependence of fluid properties on pressure and temperature is a first-order control, determining the self-organization of convection cells (*10*–*12*). Quantitative 3D numerical models have been applied to low-permeability (*13*) or sedimented systems (*14*) and to a configuration with an along-axis high-permeability fracture (*15*, *16*) but not to the more highly permeable basaltic systems, which represent the greater and most active part of mid-ocean ridge spreading centers. Here, we describe a 3D model that represents the hydrothermal system without geological complexity so as to identify the first-order physical factors controlling the behavior of mid-ocean ridge convection cells.

The governing equations are an appropriate version of Darcy's law (*17*), conservation of mass and energy in an incompressible porous medium (*12*) and an accurate equation of state for pure water (*18*). Using pure water substantially reduces the computational complexity because pure water above the critical pressure (21.1 MPa) is always a single-phase fluid with properties closely resembling those of seawater. Two-dimensional across-axis simulations including the full-phase

Department of Earth Sciences, Eidgenössische Technische Hochschule–Zurich, Clausiusstrasse 25, Zurich 9082, Switzerland.

*To whom correspondence should be addressed. E-mail: coumou@pik-potsdam.de

relations of seawater have shown that the overall narrow shape of the upwelling plume remains similar in spite of phase separation (*19*), hence the simplification employed here is justified. The equations are solved within the 3D box shown in Fig. 1, which has been discretized by a total of ~2.5 million tetrahedral elements. The total amount of energy added to the system can be estimated from the spreading rate of the oceanic plates. Magma crystallization and cooling to ambient temperatures release energy up to ~120 MW/km of the spreading axis (*2*). Direct measurements from individual vent fields or ridge segments have given substantially larger values, indicating that magma supply is episodic and local (*20*–*22*). In the model, we set the heat flux at an intermediate value of 350 MW/km, which is distributed along the bottom in a Gaussian profile that mimics an elongated magma chamber with an across-axial width of roughly 1 km (Fig. 1) (*23*). We set a bulk permeability of $k = 5.0 \times 10^{-14}$ m^2. At these specific values of permeability and heat flux, the temperature near the bottom of the model establishes itself to be close to the expected magmatic temperatures of ~1200°C (*24*). At lower permeability, the system cannot remove heat fast enough, and the system heats up unrealistically. At higher permeabilities, heat is mined faster, suppressing bottom boundary temperatures to values much smaller than magmatic temperatures. The top boundary represents the sea floor and is kept at a pressure of 25 MPa, representing an ocean depth of roughly 2.5 km. To allow hot fluids to vent through the top, we used a mixed thermal boundary condition (*11*, *12*): In elements experiencing upflow, the thermal gradient is set to zero to represent unobstructed outflow to the ocean, whereas elements experiencing downflow take in 10°C water.

In the simulations, after a short initialization period, convection developed into pipelike upflow regions, spaced at regular distances of roughly 500 m (Fig. 2A). Although thermal instabilities periodically form and rise close to the axial center, the locations of the main upflow regions remain relatively fixed. Within these pipes, fluids of ~400°C move upward vigorously and vent at the sea floor, forming near-circular discharge areas. The surface area of these fields is ~2×10^4 m^2, which is in good agreement with the range observed for natural black smoker fields of 3×10^3 m^2 to 1×10^5 m^2 (*25*). Most of the downflow happens in concentric tube-shaped regions directly surrounding the upflow plumes (Fig. 2, A to C). Fluid temperatures in these regions ranged from 100° to 300°C at already shallow depth. At a radius of less than twice that of the upflow zone, the downward mass flux reached a value five times as large as the average downward mass flux and about half the maximum upward flux. At a depth of 500 m, the integrated upward mass flux through the whole model is ~800 kg/s, which implies ~200 kg/s per kilometer of ridge segment and ~100 kg/s per vent field. Roughly two-thirds of this mass is provided by downflow in the warm near-axial regions. Elevated temperatures of the downflowing fluids reaching 300°C imply that substantial heat is recirculated; roughly one-quarter of the heat traveling upward is lost conductively to the adjacent downflow zones, in which it is recirculated back to the bottom. This recirculation process therefore controls the thermal structure of the downflow zone, heating it to average temperatures of ~200°C. Relatively warm downflow has been observed in 2D models (*26*), in which downflow has to take place in between

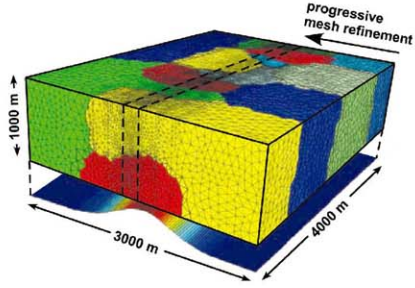

Fig. 1. 3D mesh consisting of 2.5-m tetrahedron elements. Resolution is refined toward the axial center, with the finest resolution between the dashed lines, and colors indicate computational domains assigned to separate processors (*23*). At the bottom boundary, the bell-shaped form of the heat flux profile is plotted, ranging from 10 W/m^2 (blue) to 300 W/m^2 (red).

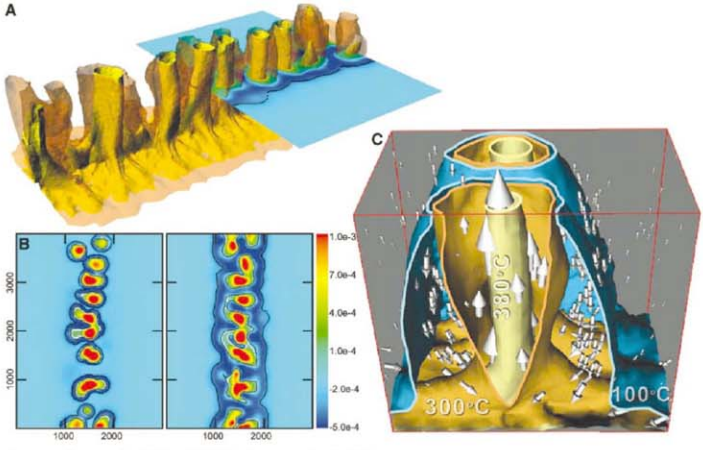

Fig. 2. Thermal and fluid-flow structure of the simulation after ~100 years. **(A)** The 3D contours of the 300°C (transparent) and 380°C (yellow) isotherms. The horizontal cross-section plots vertical mass fluxes [scale as plotted in (B)]. **(B)** Mass fluxes in kilograms per square meter per second through horizontal cross-sections at 100- and 500-m depth. Thin black lines plot the 100°C and 300°C isotherms, and the white line separates the upflowing from downflowing regions (x axis, meters; y axis, meters). **(C)** Cross-section through a thermal plume plotting 100°C (blue), 300°C (brown), and 380°C (yellow) isotherms as well as mass fluxes (arrows). High-hydraulic-conductivity downflow zones between the 100° and 300°C isotherms are clearly visible.

hot upflow zones, but is not naturally expected in 3D. Darcy fluxes in the system are ~5.0×10^{-6} m/s in the upflow zone and ~0.75×10^{-6} m/s in the surrounding near-axial downflow zone. By converting these values to actual pore velocities, a residence time of only ~3 years is calculated for a fluid particle entering the system through the proximal recharge zone, traveling to the base of the hydrothermal cell, and flowing out again through the black smoker. This short travel time is in agreement with recent estimates made with natural radionuclide tracers (*27*, *28*).

To understand why this system evolves this way, we analyzed the steady-state pressure gradient ($\partial p/\partial z$) in a simplified geometry, taking account of the nonlinearity of the fluid properties. In the near-axial region, the vertical pressure gradient must have a value between that of a cold and hot hydrostatic, so that cold fluids can be brought downward and hot fluids can be brought upward. As a first approximation, we assumed that recharge is limited to a region directly surrounding a pipe-shaped upflow zone (fig. S1). The mass balance between up- and downflowing water is

$$A_u \rho_u \frac{k_u}{\mu_u}\left[\frac{\partial p}{\partial z} - \rho_u g\right] = A_d \rho_d \frac{k_d}{\mu_d}\left[\rho_d g - \frac{\partial p}{\partial z}\right] \quad (1)$$

Here, A_u is the cross-sectional area of the upflow zone and A_d is the cross-sectional area of the downflow zone. Subscripts u and d indicate properties of the upwelling and the downwelling fluid, respectively; ρ is the water density; μ is the water viscosity; and g is the acceleration due to gravity (fig. S1). Equation 1 assumes that horizontal pressure gradients between up- and downflow areas can be neglected. This assumption is allowed in the near-axial region because there the fluid flow is dominantly vertical. The vertical pressure gradient can be expressed in terms of the density ρ_n of a fictitious neutrally buoyant fluid, according to $\partial p/\partial z = \rho_n g$. Equation 1 can now be written as

$$\frac{\rho_n - \rho_u}{\rho_d - \rho_u} = \frac{1}{1 + \gamma R} \quad (2)$$

Here, the geometric constant γ represents the ratio $A_u k_u / A_d k_d$ and R is the ratio of fluid properties $\mu_d \rho_u / \mu_u \rho_d$. Physically, the product γR can be seen as the ratio of the resistance against flow in the downflow region over that in the upflow region and, depending on pressure and temperature, can have values ranging from 0.1 to 10 γ. From Eq. 2, we can derive an expression for the fluxibility F, defined as the ability of a system to transport energy by buoyancy-driven convection

$$F = \frac{\rho_u (h_u - h_d)(\rho_d - \rho_u)}{\mu_u} \frac{1}{1 + \gamma R} \quad (3)$$

where h is the specific enthalpy of the fluid. Equation 3 is an extension of the ordinary definition of fluxibility (*29*) because it allows resistance in the downwelling limb as well as the upwelling limb. In the limit $\gamma \to 0$, in which the resistance of the downwelling limb is negligible as compared with that of the upwelling limb, Eq. 3 reduces to the original definition of fluxibility (*29*, *30*).

Heating water from 10° to 200°C decreases μ by approximately one order of magnitude, whereas ρ changes by only about 10% (Fig. 3A). As a consequence, 200°C fluids can be brought downward highly efficiently: The greatly reduced viscosity makes them mobile, whereas their only weakly decreased density maintains a substantial downward buoyancy force as compared with the upflow zone. Heating the fluid to even higher temperatures would cause the density to decrease substantially and hence lower the downward buoyancy force. Downward mass transport can therefore be optimized at temperatures of around 200°C.

In order to analyze our numerical results, we took γ as ~1 because k is homogeneous ($k_d = k_u$), and the area of upflow is roughly the same as the area of increased downflow (Fig. 2B). Figure 3B plots F against temperatures of the up- and downflowing fluid using $\gamma = 1$. It shows that energy transport can be maximized when upflowing water reaches temperatures of ~400°C and downflowing water is ~200°C, exactly like the thermal structure observed in the numerical simulations. The thermal structure of the numerical simulations featured enhanced downflow for fluids in the temperature range from 100° to 300°C, which fits with the peak of maximum fluxibility in Fig. 3B. At these conditions, a neutrally buoyant fluid has a density of ~600 kg/m^3 (Fig. 3C), implying that $\partial p/\partial z$ is ~6000 Pa/m, which is in agreement with simulation results. If upflowing fluids have either a lower or higher temperature than ~400°C, then their hydraulic resistance will increase, hence the pressure gradient has to increase to drive such fluids upwards, resulting in larger neutral buoyancies (Fig. 3C). Similarly, the resistance of the downflow zone increases if its temperature is either larger or smaller than ~200°C. Cooling a ~200°C fluid will increase its viscosity more rapidly than its density, whereas heating a ~200°C fluid will decrease its density faster than its viscosity (see Fig. 3A). Both effects result in a larger hydraulic resistance of the fluid. This analysis shows that the convection cells operate in a state of least hydraulic resistance in order to maximize the overall efficiency of energy transport. Thus, surprisingly, substantial recirculation of heat unexpectedly enhances mass fluxes through the system and thereby increases the overall heat transport of the cell. Previous studies indicated that black smoker systems could well be operating in a state of maximum energy transport (*30*–*33*), explaining why black smoker vent fields are often close to but never higher than ~400°C (*11*).

Convection cells evolve toward and remain in this state of least resistance by counteracting feedback mechanisms that self-regulate the flow in convection cells. For instance, increasing resistance by cooling the downflow zone to temperatures lower than ~200°C slows down the fluid flow through the system. This reduced flow increases conductive heat losses from the upflow area, heating up the downflow area again to temperatures of ~200°C. Furthermore, the upflow temperature cannot become hotter than ~400°C (*11*). Therefore, if convection stagnates because of external forces, the boundary layer between the magma chamber and hydrothermal cell will broaden, but upflow temperatures will stay at ~400°C.

The regular spacing between black smoker fields has previously been explained with a slot-convection model, in which convection dominantly takes place in a high-permeability axial plane and therefore could be considered a 2D system (*12*). Our simulations show that this restriction is not required and that, at high heat-flux conditions, convection naturally forms individual black smoker fields at regular spacing because of the nonlinear temperature and pressure dependence of the fluid's properties.

The Main Endeavour Field (MEF) on the Juan de Fuca ridge appears to be a natural exam-

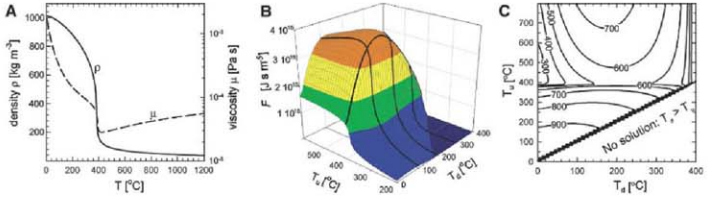

Fig. 3. **(A)** Density (solid) and viscosity (dashed) as a function of temperature at 30 MPa. **(B)** F plotted against upwelling and downwelling temperatures (Eq. 3). Solid black lines follow contours at downwelling temperature (T_d) = 100°C, T_d = 200°C, T_d = 300°C, and T_u = 400°C. Blue, 0 < F < 1e15; Green, 1e15 < F < 2e15; Yellow, 2e15 < F < 3e15; and Orange, 3e15 < F < 4e15. **(C)** Neutral density (in kilograms per cubic meter) plotted against upwelling and downwelling temperatures.

ple closely matching our simulation results. There, regularly spaced vent sites, with black smokers venting fluids close to 400°C, have been active for several years (*34*, *35*). Magnetic anomaly data show that upflow zones are narrow pipelike structures that reach to a depth of at least a few hundreds of meters and possibly to the base of the hydrothermal system with a regular spacing comparable with that in our simulations (*3*). Though natural systems are geologically much more complex than the numerical model described here, we suggest that the MEF is likely to operate in a state of maximum energy transport, with recharge occurring close to the vent sites. The optimal site for future in situ tracer injection experiments would therefore be about twice the radius of a black smoker vent field. The relatively short residence times resulting from our calculations also indicate that such an experiment can lead to a successful tracer test within a realistic time frame.

Massive sulfide ore deposits form when convecting seawater leaches metals from new basaltic crust and reprecipitates them as sulfides at the outflow points of active black smokers. The solubility of iron, zinc, and copper increases almost exponentially with temperature (*36*). Our simulations indicate that the average temperature of fluid-rock interaction is much higher than would be expected from dispersed seawater infiltration across the spreading axis (*37*). As a result, solubility-limited leaching is expected to be much more effective by including large parts of the hot downflow path and not being restricted to the basal reaction zone of the system. With an average copper content in mid-ocean ridge basalt of 25 parts per million, a typical small seafloor sulfide deposit of 0.2 million metric tons containing 3% copper (*38*) can be formed by leaching only the upflow zone and its immediately surrounding hot downflow zone. The copper solubility in this region varies from ~10^{-7} mol/kg at 200°C to ~10^{-4} mol/kg at 350°C, ensuring that the metals are quickly leached and a deposit can form within a period of 100 to 1000 years.

References and Notes

1. C. Stein, S. Stein, *J. Geophys. Res.* **99**, 3081 (1994).
2. A. Fisher, in *Energy and Mass Transfer in Marine Hydrothermal Systems*, P. E. Halbach, V. Tunnicliffe, J. R. Hein, Eds. (Dahlem Univ. Press, Berlin, 2003), vol. 89, pp. 29–52.
3. M. Tivey, H. Johnson, *Geology* **30**, 979 (2002).
4. L. Coogan *et al.*, *Am. J. Sci.* **306**, 389 (2006).
5. H. Johnson, K. Becker, R. Von Herzen, *Geophys. Res. Lett.* **20**, 1875 (1993).
6. S. Kelley, J. Baross, J. Delaney, *Annu. Rev. Earth Planet. Sci.* **30**, 385 (2002).
7. P. Nehlig, T. Juteau, *Mar. Geol.* **84**, 209 (1988).
8. W. Wilcock, A. Fisher, *Geophys. Monogr.* **144**, 51 (2004).
9. M. Tolstoy, F. Waldhause, D. Bohnenstiehl, R. Weekly, W. Kim, *Nature* **451**, 181 (2008).
10. S. Ingebritsen, D. O. Hayba, *Geophys. Res. Lett.* **21**, 2199 (1994).
11. T. Jupp, A. Schultz, *Nature* **403**, 880 (2000).
12. D. Coumou, T. Driesner, S. Geiger, C. Heinrich, S. Matthai, *Earth Planet. Sci. Lett.* **245**, 218 (2006).
13. B. Travis, D. Janecky, N. Rosenberg, *Geophys. Res. Lett.* **18**, 1441 (1991).
14. M. Rabinowicz, J. Boulegue, P. Genthon, *J. Geophys. Res.* **103**, 24045 (1998).
15. M. Rabinowicz, J. Sempéré, P. Genthon, *J. Geophys. Res.* **104**, 29275 (1999).
16. F. Fontaine, M. Rabinowicz, J. Boulegue, *Earth Planet. Sci. Lett.* **184**, 407 (2001).
17. S. Ingebritsen, W. Sanford, C. Neuzil, *Groundwater in Geologic Processes* (Cambridge Univ. Press, Cambridge, ed. 2, 2006).
18. G. Haar, Kell, *NBS/NRC Steam Tables* (Hemisphere Publishing, New York, 1984).
19. D. Coumou, thesis, Eidgenössische Technische Hochschule–Zurich, Switzerland (2008).
20. U. Ginster, M. Mottl, R. Von Herzen, *J. Geophys. Res.* **99**, 4937 (1994).
21. E. Baker, T. Urabe, *J. Geophys. Res.* **101**, 8685 (1996).
22. P. Ramondenc, L. Germanovich, K. Von Damm, R. Lowell, *Earth Planet. Sci. Lett.* **245**, 487 (2006).
23. Materials and methods are available as supporting material on *Science* Online.
24. J. Sinton, R. Detrick, *J. Geophys. Res.* **97**, 197 (1992).
25. R. Lowell, A. Rona, R. Von Herzen, *J. Geophys. Res.* **100**, 327 (1995).
26. F. Fontaine, W. Wilcock, *Geochem. Geophys. Geosyst.* **8**, Q07010 (2007).
27. D. Kadko, D. Butterfield, *Geochim. Cosmochim. Acta* **62**, 1521 (1998).
28. D. Kadko, G. K., D. Butterfield. *Geochim. Cosmochim. Acta* **71**, 6019 (2007).
29. C. Lister, *Geophys. J. Int.* **120**, 45 (1995).
30. T. Jupp, A. Schultz, *J. Geophys. Res.* **109**, 10.1029/2003JB002697 (2004).
31. W. Malkus, *Proc. R. Soc. London. Ser. A* **225**, 196 (1954).
32. F. Busse, D. Joseph, *J. Fluid Mech.* **54**, 521 (1972).
33. W. S. D. Wilcock, *J. Geophys. Res.* **103**, 2585 (1998).
34. E. M. Van Ark *et al.*, *J. Geophys. Res.* **112**, 10.1029/2005JB004210 (2007).
35. D. Glickson, D. Kelley, J. Delaney, *Geochem. Geophys. Geosyst.* **8**, Q06010 (2007).
36. J. Hemley, G. Cygan, J. Fein, G. Robinson, W. Angelo, *Econ. Geol.* **87**, 1 (1992).
37. J. Franklin, H. Gibson, I. Jonasson, A. Galley, in *Economic Geology, 100th Anniversary Volume*, J. W. Hedenquist, J. F. H. Thompson, R. J. Goldfarb, J. P. Richards, Eds. (Society of Economic Geologists, Littleton, CO, 2005), pp. 523–560.
38. M. Hannington, I. Jonasson, P. Herzig, S. Petersen, *Geophys. Monogr.* **91**, 115 (1995).

39. This work was supported by the Swiss National Science Foundation (grant 200020-107955). The authors thank S. Geiger for providing the mesh depicted in Fig. 1 and S. Ingebritsen, P. Weiss, L. Cathles, and three anonymous reviewers for useful discussions as well as careful proofreading of earlier versions of the manuscript.

Supporting Online Material

www.sciencemag.org/cgi/content/full/321/5897/1825/DC1
Materials and Methods
Fig. S1
References

24 April 2008; accepted 28 July 2008
10.1126/science.1159582

Neodymium-142 Evidence for Hadean Mafic Crust

Jonathan O'Neil,[1]* Richard W. Carlson,[2] Don Francis,[1] Ross K. Stevenson[3]

Neodymium-142 data for rocks from the Nuvvuagittuq greenstone belt in northern Quebec, Canada, show that some rock types have lower $^{142}Nd/^{144}Nd$ ratios than the terrestrial standard ($\varepsilon^{142}Nd = -0.07$ to -0.15). Within a mafic amphibolite unit, $^{142}Nd/^{144}Nd$ ratios correlate positively with Sm/Nd ratios and produce a ^{146}Sm-^{142}Nd isochron with an age of 4280^{+53}_{-81} million years. These rocks thus sample incompatible-element-enriched material formed shortly after Earth formation and may represent the oldest preserved crustal section on Earth.

The past decade has seen dramatic discoveries concerning the oldest rocks on Earth, with precise zircon ages pushing the terrestrial rock record back beyond 4 billion years ago (Ga) (*1*) and the detrital zircon record to beyond 4.3 Ga (*2*). Because zircon is a rare to nonexistent phase in most mafic rocks, prospecting for ancient crust through zircon analysis has focused the search on the more evolved rock types that likely, as today, do not represent the major volume of Earth's crust.

The short-lived ^{146}Sm-^{142}Nd isotopic system [half life ($T_{1/2}$) = 103 million years (My)] has proven useful for investigating the early differentiation of the silicate portion of Earth. Recent measurements of the ^{146}Sm-^{142}Nd system in Eoarchean (4.0 to 3.6 Ga) rocks, primarily from Greenland, show excesses in $^{142}Nd/^{144}Nd$ ratios of 10 to 20 parts per million (ppm) compared to modern terrestrial standards testifying to Earth differentiation events within a few tens of million years of Earth formation (*3*–*7*). The high $^{142}Nd/^{144}Nd$ measured for these rocks indicate that the Eoarchean crustal rocks were sourced in a mantle with high Sm/Nd ratio. We describe evidence from the Nuvvuagittuq greenstone belt that a complimentary, low Sm/Nd ratio, reservoir is also found in the terrestrial rock record and that these rocks may be the oldest yet discovered on Earth.

The recent discovery of the Nuvvuagittuq greenstone belt in Ungava, Québec, provides a new suite of Eoarchean rocks with which to further our understanding of the early crust-mantle system. The Nuvvuagittuq belt exposes volcanic and metasedimentary rocks in an isoclinal synform refolded into a more open south-plunging synform (Fig. 1) (*8*) and is surrounded by a 3.66-billion-

[1]Earth and Planetary Sciences Department, McGill University, 3450 University Street, Montreal, Quebec, H3A 2A7, Canada. [2]Department of Terrestrial Magnetism, Carnegie Institution of Washington, 5241 Broad Branch Road, NW, Washington, DC 20015, USA. [3]GEOTOP (Centre de recherche en géochemie et géodynamique), Université du Québec à Montréal, Post Office Box 8888, Succursale Centre-ville, 210, Président-Kennedy Avenue, Montreal, Quebec H3C 3P8, Canada.

*To whom correspondence should be addressed. E-mail: oneil_jo@eps.mcgill.ca

year (Gy) tonalite (*9*, *10*). Geochronological constraints for the belt come mainly from rare felsic bands (0.5 to 1 m in width) composed of plagioclase, biotite, and quartz that have yielded a discordant zircon age of 3817 ± 16 My (*9*). A minimum age of 3750 My has also been obtained from U-Pb ion microprobe analyses of zircons found in a similar lithology (*11*). Although no clear crosscutting relationship has been found, the felsic bands are commonly found within gabbroic sills that may suggest an intrusive nature for these rocks. The dominant lithology of the belt is a cummingtonite-amphibolite referred to as "faux-amphibolite" because of its unusual mineralogical composition in which the dominant amphibole is cummingtonite, in contrast to the hornblende-dominated amphibolites usually found in the Superior Province. The faux-amphibolite is composed of variable amounts of cummingtonite, plagioclase, biotite, and quartz plus or minus garnet (*8*), commonly with compositional layering defined by the alternation of biotite-rich and cummingtonite-rich laminations. These rocks are very heterogeneous and can be almost entirely composed of cummingtonite, giving them a light gray color, whereas some are reddish brown in color and mainly consist of biotite and garnet, with minor amounts of cummingtonite.

The faux-amphibolite is intruded in the western limb of the synform by ultramafic and gabbroic sills. Compared with these relatively undeformed gabbros, two larger gabbroic sills toward the center of the belt display pronounced gneissic textures, suggesting that they may be older than the less-deformed gabbroic sills. The faux-amphibolite in the western limb rarely contains garnet, whereas toward the center of the belt it has abundant garnet and commonly higher Al_2O_3 contents (*8*). The faux-amphibolite was originally interpreted to be a paragneiss (*10*) because of the abundance of garnet and compositional layering. These rocks, however, are more mafic than typical Archean shales and have a basaltic major element composition similar to those of the gabbroic sills but with lower CaO, slightly lower TiO_2, and commonly higher Al_2O_3 contents. These features suggested that they could be highly altered mafic pyroclastites comagmatic with the gabbro sills (*8*). The faux-amphibolite, however, is enriched in light rare earth elements (LREEs), which argues against a direct cogenetic relationship with the gabbros that intrude it, all of which have flat REE patterns (table S1).

Seven samples of the faux-amphibolite yielded low $^{147}Sm/^{144}Nd$ ratios (0.143 to 0.179) with correspondingly low measured $^{143}Nd/^{144}Nd$ ($\varepsilon^{143}Nd$ from −27.4 to −9.7, where $\varepsilon^{143}Nd = [(^{143}Nd/^{144}Nd)_{sample}/(^{143}Nd/^{144}Nd)_{CHUR} - 1] \times 10^4$ and CHUR is chondritic uniform reservoir) and $^{142}Nd/^{144}Nd$ ratios ($\varepsilon^{142}Nd = -0.07$ to −0.15, where $\varepsilon^{142}Nd = [(^{142}Nd/^{144}Nd)_{sample}/(^{142}Nd/^{144}Nd)_{standard}) - 1] \times 10^4$) relative to the terrestrial standard [Fig. 2, table S2, and supporting online material (SOM) text]. Two tonalite samples also show low $^{142}Nd/^{144}Nd$ ($\varepsilon^{142}Nd = -0.12$ to −0.16) as do both felsic band samples ($\varepsilon^{142}Nd = -0.09$ to −0.10), although the felsic bands overlap the terrestrial standard within the external error of 0.06 as determined by the 2σ-population reproducibility of the standards (SOM text). The gabbros have $^{147}Sm/^{144}Nd$ ratios (0.183 to 0.193) and measured $^{143}Nd/^{144}Nd$ ratios ($\varepsilon^{143}Nd = -4.7$ to 0.6) that are just slightly below chondritic values, with $^{142}Nd/^{144}Nd$ ratios overlapping the terrestrial standard ($\varepsilon^{142}Nd = -0.09$ to −0.02).

The Nuvvuagittuq rocks have $^{142}Nd/^{144}Nd$ ratios that fall between the $^{142}Nd/^{144}Nd$ ratios of chondrites and the terrestrial standard [e.g., (*12*)]. This is in contrast to Eoarchean rocks of Greenland that have $^{142}Nd/^{144}Nd$ values higher than the terrestrial standard (*3*–*7*). Also, unlike the Greenland rocks, the $^{142}Nd/^{144}Nd$ ratios of the gabbros and faux-amphibolite correlate positively with their Sm/Nd ratios, producing a statistically significant slope corresponding to a $^{146}Sm/^{144}Sm$ ratio of 0.00116 ± 0.00049 [mean square weighted deviation (MSWD) = 0.67, error with 95% confidence] and an initial $\varepsilon^{142}Nd = -0.02 \pm 0.15$ relative to the terrestrial standard (Fig. 3). This line is fit assigning a constant ±6 ppm error for $^{142}Nd/^{144}Nd$ for all samples to represent the external reproducibility of these isotope ratio determinations. For a solar system initial $^{146}Sm/^{144}Nd = 0.008$ (*13*) at 4567 Ga (*14*), this slope corresponds to an age of 4280^{+53}_{-81} My. Fitting just the faux-amphibolite data provides a slope of 0.0012 ± 0.0011 corresponding to an age of 4286^{+96}_{-370} My with an initial $\varepsilon^{142}Nd = +0.02 \pm 0.32$ relative to the terrestrial standard. Both the tonalites and the felsic bands fall off these correlations to the low Sm/Nd ratio side.

At the whole-rock scale, isochrons need not provide the crystallization ages of the rocks that define the isochron. Because 4.28-Gy-old the isochron shown in Fig. 3 goes through the value of the modern terrestrial mantle, the unfractionated rocks would plot close to the modern mantle value on Fig. 3 regardless of their age. For example, 12 ultramafic to gabbroic samples from one sill cutting the faux-amphibolite give a ^{147}Sm-^{143}Nd isochron age of 3840 ± 280 My (MSWD = 3.8) with initial $\varepsilon^{143}Nd = +0.9$ (Fig. 4A), but the data for some of the same samples lie close to the 4.28-Gy isochron in Fig. 3. The old age in Fig. 3 is dictated by the low Sm/Nd ratios of the faux-amphibolite. The faux-amphibolite defines a scattered ^{147}Sm-^{143}Nd correlation with an age of 3819 ± 270 My (MSWD = 5.5) and an initial $\varepsilon^{143}Nd = -1.4$. The negative initial $\varepsilon^{143}Nd$ of the faux-amphibolites is unusual for rocks of this age but is consistent with the low $^{142}Nd/^{144}Nd$ ratios measured for these samples.

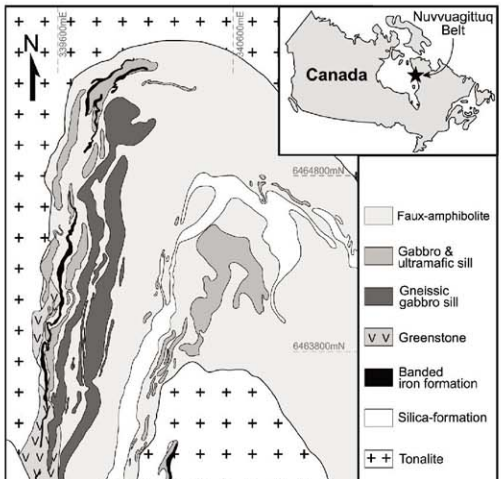

Fig. 1. Simplified geological map of the Nuvvuagittuq greenstone belt. The geology of the belt is described in more detail in (*8*). Coordinates are in universal transverse mercator zone 18 NAD 27.

Sm/Nd ratios as low as those measured in the faux-amphibolite and felsic bands would not produce $^{142}Nd/^{144}Nd$ ratios outside of measurement uncertainty from the terrestrial standard if this parent/daughter fractionation occurred later than ~4.1 to 4.2 Ga. Thus, even if the faux-amphibolite has crystallization ages of 3.8 Gy, the steepness of the $^{142}Nd/^{144}Nd$–Sm/Nd covariation requires that they sample a LREE-enriched material that is at least 4.28 Gy old. This LREE-enriched component could be either older crust that contaminated parental melts like the gabbros or a LREE-enriched mantle source that melted to produce the faux-amphibolite. Regardless of the nature of the LREE-enriched component, its low Sm/Nd ratio must have formed while ^{146}Sm was still extant, and the 4.28-Gy-old isochron provides the best indication of the age of this end member.

An alternate interpretation is that the 4.28-Gy-old isochron indeed dates the formation age of the faux-amphibolite, but this possibility is not supported by the 3.8-Gy-old ^{147}Sm–^{143}Nd age of this unit. For an isochron that passes through the terrestrial mantle point, however, a reduction in Sm/Nd ratio of the faux-amphibolite with the lowest Sm/Nd ratios by only 4.4% caused by metamorphism at 3.8 Ga would rotate a ^{147}Sm–^{143}Nd isochron of 4.28 Gy to 3.8 Gy. Increasing the Sm/Nd ratio of the low–Sm/Nd ratio faux-amphibolite by 4.4% would increase the ^{146}Sm–^{142}Nd isochron age by only 25 My, well within the uncertainty of the data, illustrating the potential of the ^{146}Sm–^{142}Nd system to see through later metamorphic events.

Obviously, other corroborative data would help resolve whether the 4.28-Gy age dates the rocks themselves or an older component involved in their genesis. In spite of attempts to do so, zircons have not yet been found in the faux-amphibolite. Whole-rock Pb isotope data for the faux-amphibolite (table S1) do not define a valid isochron. The best fit line through the $^{206}Pb/^{204}Pb$–$^{207}Pb/^{204}Pb$ data corresponds to an age of 2.4 ± 0.4 Gy, indicative of a late disturbance of the U-Pb system at the whole-rock scale. The faux-amphibolite is crosscut by the gabbro sills and therefore must be older than the gabbros. Although the undeformed gabbros give a ^{147}Sm–^{143}Nd isochron age of 3.84 Gy (Fig. 4A), an isochron constructed from nine samples of the more gneissic, presumably older, gabbros gives an age of 4023 ± 110 My (MSWD = 0.78) with initial $\varepsilon^{143}Nd = +1.7$ (Fig. 4B). All samples of the faux-amphibolite, except PC-129, yield negative $\varepsilon^{143}Nd$(3.8 Gy) values ranging from –3.2 to –1.0, compared with the mostly positive $\varepsilon^{143}Nd$(3.8 Ga) (–0.2 to +3.1) for the gabbro and ultramafic sills (table S2). When the $\varepsilon^{143}Nd$ values for the faux-amphibolite, except PC-129, are calculated for an age of 4.28 Gy, they range from –0.3 to +2.3 with an average value of 0.6. This average initial $\varepsilon^{143}Nd$ value is consistent with the mantle value at 4.28 Ga predicted by various depleted mantle evolution models (*5*, *15*, *16*), but

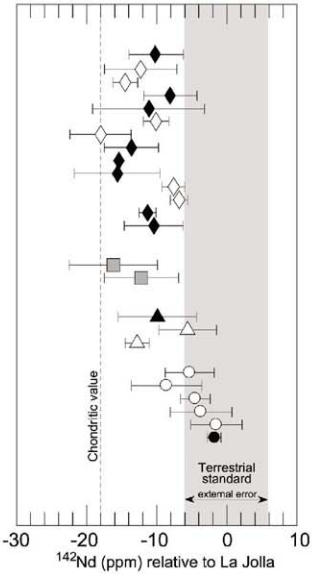

Fig. 2. The $^{142}Nd/^{144}Nd$ ratios for the Nuvvuagittuq rocks normalized to the La Jolla standard. Gray solid bar corresponds to the external (6 ppm) error obtained on the terrestrial standard. Error bars for individual samples correspond to either the 2σ-mean internal precision of the mass spectrometer analysis or the 2σ-mean of repeat analysis of the sample on the same mass spectrometer filament load. Solid circle indicates ultramafic sill; open circles, gabbro; diamonds, faux-amphibolite (alternating open and solid diamonds show data for replicate analyses of single samples); squares, tonalite; and triangles (alternating open and solid triangles show data for replicate analyses of single samples), felsic band.

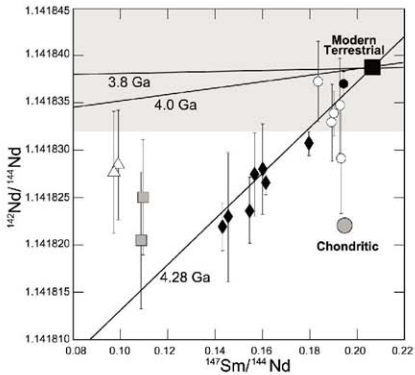

Fig. 3. $^{142}Nd/^{144}Nd$ versus $^{147}Sm/^{144}Nd$ isochron diagram. Symbols as described for Fig. 2. Only the average value of replicate analyses is plotted on this figure. The horizontal gray band shows the ±6 ppm external precision obtained on the terrestrial standard. Error bars on individual samples are either the 2σ mean of multiple analyses or the 2σ mean of the individual mass spectrometer run for samples run only once. The best fit line through the faux-amphibolite and gabbro data corresponding to an age of 4.28 Gy is shown, as are 3.8- and 4.0-Gy isochrons for reference. The gray circle shows the average value measured for ordinary and enstatite chondrites (*12*).