versions of SNAP-25 that mimic cleavage by BoNT/A (residues 1 to 197) were reconstituted into vesicles (with 15% PS). As a control, a truncated version of SNAP-25 that mimics cleavage by BoNT/E (corresponding to residues 1 to 180), was tested in parallel [this cleavage event results in a more profound block of exocytosis (*13*)]. Fusion was abolished by the "BoNT/E" truncation (Fig. 4B). In contrast, the "BoNT/A" truncation supported a low level of fusion that could be enhanced by increasing $[Ca^{2+}]$; the Ca^{2+} response was too impaired to determine the precise $[Ca^{2+}]_{1/2}$, but this value is >360 μM (Fig. 4B). Thus, the reconstituted system recapitulates the functional effect of BoNT/A and E treatment on neurons (*13*, *19*, *20*).

Membrane-embedded syt has been reported to stimulate membrane fusion in a Ca^{2+}-independent manner (*21*). We have repeated these experiments and observed the same phenomena. The lack of an effect of Ca^{2+} is surprising, because the ability of syt to interact with its targets in the reduced fusion assay is promoted by Ca^{2+}. The easiest explanation, however, is that the bacterially expressed full-length syt is not fully functional. Variants of a number of isoforms of syt, including syt I, that lack a transmembrane domain are expressed in cells where they may also regulate membrane traffic in vivo, supporting the idea that studies with the cytoplasmic domain of syt are physiologically relevant (*22–24*).

The data reported here indicate that a complex of syt, membranes, and SNARE proteins forms the core of the Ca^{2+}-triggered fusion apparatus. With this reconstitution approach, it should be possible to test additional factors to construct a Ca^{2+}-triggered membrane fusion complex that operates on the rapid (millisecond) time scale observed during synaptic transmission.

References and Notes

1. T. Weber *et al.*, *Cell* **92**, 759 (1998).
2. T. Sollner *et al.*, *Nature* **362**, 318 (1993).
3. J. A. McNew *et al.*, *Nature* **407**, 153 (2000).
4. F. Parlati *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 12565 (1999).
5. B. Katz, *The Release of Neural Transmitter Substances* (Charles C. Thomas, Springfield, IL, 1969).
6. G. J. Augustine, *Curr. Opin. Neurobiol.* **11**, 320 (2001).
7. N. Brose, A. G. Petrenko, T. C. Sudhof, R. Jahn, *Science* **256**, 1021 (1992).
8. W. C. Tucker, E. R. Chapman, *Biochem. J.* **366**, 1 (2002).
9. D. K. Struck, D. Hoekstra, R. E. Pagano, *Biochemistry* **20**, 4093 (1981).
10. Detailed materials and methods are available as supporting material on *Science* Online.
11. R. Jahn, T. C. Sudhof, *Annu. Rev. Neurosci.* **17**, 219 (1994).
12. C. Walch-Solimena *et al.*, *J. Cell Biol.* **128**, 637 (1995).
13. G. Schiavo, M. Matteoli, C. Montecucco, *Physiol. Rev.* **80**, 717 (2000).
14. X. Zhang, J. Rizo, T. C. Sudhof, *Biochemistry* **37**, 12395 (1998).
15. C. A. Earles, J. Bai, P. Wang, E. R. Chapman, *J. Cell Biol.* **154**, 1117 (2001).
16. I. Fernandez *et al.*, *Neuron* **32**, 1057 (2001).
17. E. R. Chapman, S. An, J. M. Edwardson, R. Jahn, *J. Biol. Chem.* **271**, 5844 (1996).
18. A. F. Davis *et al.*, *Neuron* **24**, 363 (1999).
19. H. Lundh, S. Leander, S. Thesleff, *J. Neurol. Sci.* **32**, 29 (1977).
20. S. G. Cull-Candy, H. Lundh, S. Thesleff, *J. Physiol.* **260**, 177 (1976).
21. L. K. Mahal, S. M. Sequeira, J. M. Gureasko, T. H. Sollner, *J. Cell Biol.* **158**, 273 (2002).
22. M. Craxton, M. Goedert, *FEBS Lett.* **460**, 417 (1999).
23. C. Bagala *et al.*, *Biochem. Biophys. Res. Commun.* **310**, 1041 (2003).
24. M. Fukuda, K. Mikoshiba, *J. Biol. Chem.* **274**, 31428 (1999).
25. We thank M. Dong for technical assistance with the BoNT/B experiments and A. Riccio for assistance with

Integration of Word Meaning and World Knowledge in Language Comprehension

Peter Hagoort,[1,2,3]* Lea Hald,[1] Marcel Bastiaansen,[1] Karl Magnus Petersson[1]

Although the sentences that we hear or read have meaning, this does not necessarily mean that they are also true. Relatively little is known about the critical brain structures for, and the relative time course of, establishing the meaning and truth of linguistic expressions. We present electroencephalogram data that show the rapid parallel integration of both semantic and world knowledge during the interpretation of a sentence. Data from functional magnetic resonance imaging revealed that the left inferior prefrontal cortex is involved in the integration of both meaning and world knowledge. Finally, oscillatory brain responses indicate that the brain keeps a record of what makes a sentence hard to interpret.

Language is used, among other things, to exchange information about the world. This entails that, during online comprehension, the meaning of a phrase or sentence is derived and, in many cases, its truth is verified. For this to be possible, usually information about the words of a language and about the facts of the world need to be retrieved from memory.

At least since Frege (*1*, *2*), theories of meaning have made a distinction between the semantics of an expression and its truth value in relation to our mental representation of the state of affairs in the world (*3*, *4*). For instance, the sentence "the present queen of England is divorced" has a coherent semantic interpretation, but it contains a proposition that is false in the light of our knowledge in memory that she is married to Prince Phillip. The situation is different for the sentence "the favorite palace of the present queen of England is divorced." Under default interpretation conditions, this sentence has no coherent semantic interpretation, because the predicate *is-divorced* requires an animate argument. This sentence mismatches with our representation of the world in memory, because the descriptive features of the purported state of affairs are inherently in conflict. The difference between these two sentences suggests the distinction that can be made between facts of the world and facts of the words of our language, including their meaning. Although theories of semantic memory usually do not make this distinction (*5*), accounts of online language processing often do, and they distinguish between the retrieval and usage of world knowledge and of knowledge of word meaning.

Relative to the distinction between facts of the world and facts of the words of one's language, some aspects of word meaning might be characterized as linguistic in nature, whereas other aspects relate to world knowledge. In linguistic theory, the latter is referred to as the domain of pragmatics, and the former as the domain of semantics. Based on

[1]F.C. Donders Centre for Cognitive Neuroimaging, [2]Nijmegen Institute for Cognition and Information, University of Nijmegen, Nijmegen, Netherlands. [3]Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands.

*To whom correspondence should be addressed. E-mail: peter.hagoort@fcdonders.kun.nl

this distinction between semantics and pragmatics (*6–8*), the semantic interpretation of a sentence is thought to be separate from and to precede the integration of pragmatic or world knowledge information (*9*). However, a number of researchers (*3*, *10*) have pointed out that the distinction between linguistic meaning and world knowledge is problematic because many words are polysemous, and their meaning can only be fully established by invoking world knowledge (*11*).

We decided to contribute to settling this issue by providing neurophysiological evidence on the integration of semantic and world knowledge information. The underlying idea is that if a principled distinction can be made between linguistic meaning and world knowledge, concomitant processing differences should be obtained during the interpretation of a sentence.

This study presents electroencephalogram (EEG) and functional magnetic resonance imaging (fMRI) data that speak to this issue. While participants' brain activity was recorded, they read three versions of sentences such as "the Dutch trains are *yellow/white/sour* and very crowded" (the critical words are in italics). It is a well-known fact among Dutch people that Dutch trains are yellow, and therefore the first version of this sentence is correctly understood as true. However, the linguistic meaning of the alternative color term *white* applies equally well to trains as the predicate *yellow.* It is world knowledge about trains in Holland that makes the second version of this sentence false. This is different for the third version. This version contains a violation of semantic constraints. The core meaning of *sour* is related to taste and food. Under standard interpretation conditions, a predicate requires an argument whose semantic features match that of its predicate. For our third sentence, this is clearly not the case, because semantic features related to taste and food do not apply to trains. One could thus argue that for semantic internal reasons, the third sentence is false or incoherent (*3*, *12*). It is our knowledge about the words of our language and their linguistic meaning that pose a problem for the interpretation of the third version of this sentence.

The increased interpretation load of semantic and world knowledge violations is assumed to have an effect on electrophysiological brain activity and on the hemodynamic response. If semantic interpretation precedes verification against world knowledge, the effects of the semantic violations should be earlier and might invoke other brain areas than the effects of the world knowledge violations.
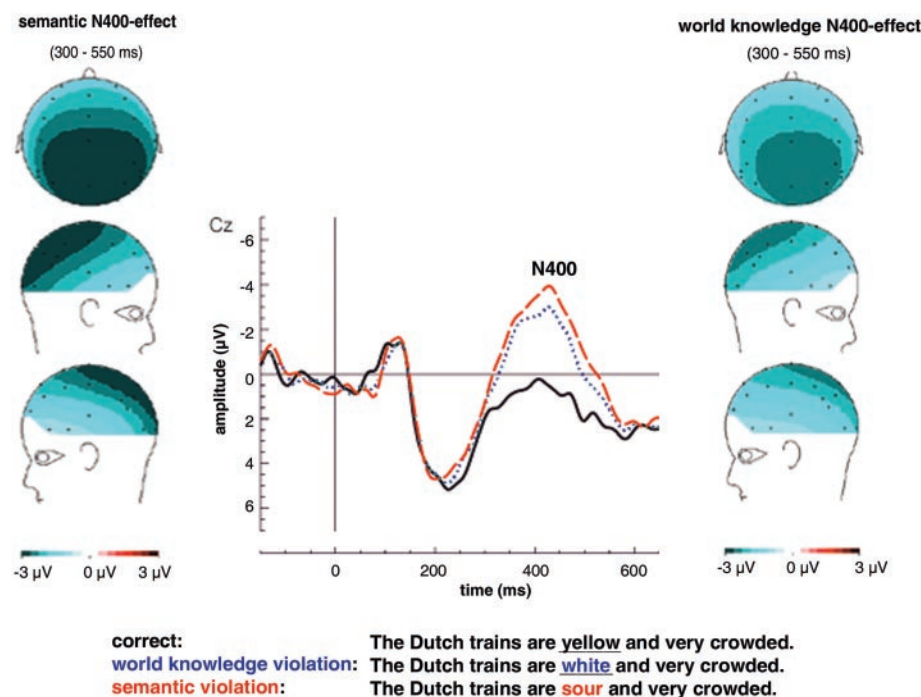
Based on EEG recordings from 29 electrode sites (*13*, *14*), event-related brain potentials (ERPs) were computed and time-locked to the onset of the critical words that embodied the semantic violation, the world knowledge violation, and their correct counterpart. ERPs reflect the summation of the postsynaptic potentials of a large ensemble of synchronously active neurons. They provide a sampling of the brain's electrical activity with a very high temporal resolution. We focused on one particular ERP effect, referred to as the N400. The amplitude of this negative-going ERP between roughly 250 and 550 ms, with a maximum at ~400 ms, is influenced by the processing of semantic information (*15*). The easier the match between the lexical semantics of a particular content word and the semantic specification of the context, the more reduced the N400 amplitude will be. The N400 is known to be very sensitive to semantic integration processes (*16*, *17*).

As expected, the classic N400 effect was obtained for the semantic violations. For the world knowledge violations, we also observed a clear N400 effect. Crucially, this effect was identical in onset and peak latency and was very similar in amplitude and topographic distribution to the semantic N400 effect (Fig. 1). This finding is strong empirical evidence that lexical semantic knowledge and general world knowledge are both integrated in the same time frame during sentence interpretation, starting at ~300 ms after word onset.

In addition, we used the EEG data to investigate oscillatory brain activity in a wide frequency range (1 to 70 Hz) in relation to the semantic and world knowledge violations. Amplitude increases of EEG oscillations in specific frequency bands, such as theta (4 to 7 Hz) and gamma (~30 to 70 Hz), that are induced by a cognitive event are thought to reflect the dynamic recruitment of the relevant neuronal networks engaged in cognitive processing (*18*). A wavelet-based time-frequency representation (TFR) of EEG power changes (Fig. 2) revealed a clear gamma peak for the world knowledge violation that was not seen for the semantic violation (*19*). In contrast, relative to the other conditions, the semantic violation resulted in an increase in power in the theta frequency range. Both effects are visible within the latency range of the N400. Especially at lower frequencies (e.g., at theta frequencies), the temporal resolution of the wavelet transform is relatively poor. This implies that the relative onset difference between theta and gamma activity cannot be taken as a reliable indicator of onset differences in the underlying neurophysiological events. In particu-

semantic N400-effect
(300 - 550 ms)

world knowledge N400-effect
(300 - 550 ms)

-3 µV   0 µV   3 µV

-3 µV   0 µV   3 µV

correct: The Dutch trains are <u>yellow</u> and very crowded.
world knowledge violation: The Dutch trains are <u>white</u> and very crowded.
semantic violation: The Dutch trains are <u>sour</u> and very crowded.

Fig. 1. Grand average (*n* = 30 subjects) ERPs for a representative electrode site (Cz) for the correct condition (black line), world knowledge violation (blue dotted line), and semantic violation (red dashed line). ERPs were time-locked to the presentation of the critical words (underlined). Semantic violations resulted in a larger N400 amplitude between 300 and 550 ms than the control condition [$F(1,29) = 73.0$, $P < 0.0001$]. N400 amplitudes to world knowledge violations were also larger than for correct controls [$F(1,29) = 27.4$, $P < 0.0001$]. The size of the effect was slightly larger for semantic violations than for world knowledge violations ($P < 0.05$). The onset of the effects for semantic and world knowledge violations was not significantly different (*14*). Spline-interpolated isovoltage maps display the topographic distributions of the mean differences from 300 to 550 ms between semantic violation and control (left) and between world knowledge violation and control (right). Topographic distributions of the N400 effect were not significantly different between semantic and world knowledge violations ($P = 0.9$).
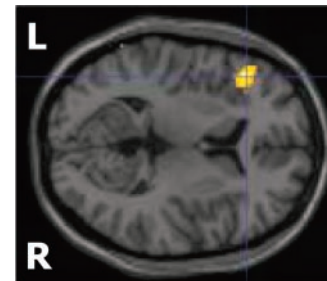
Time-frequency analysis
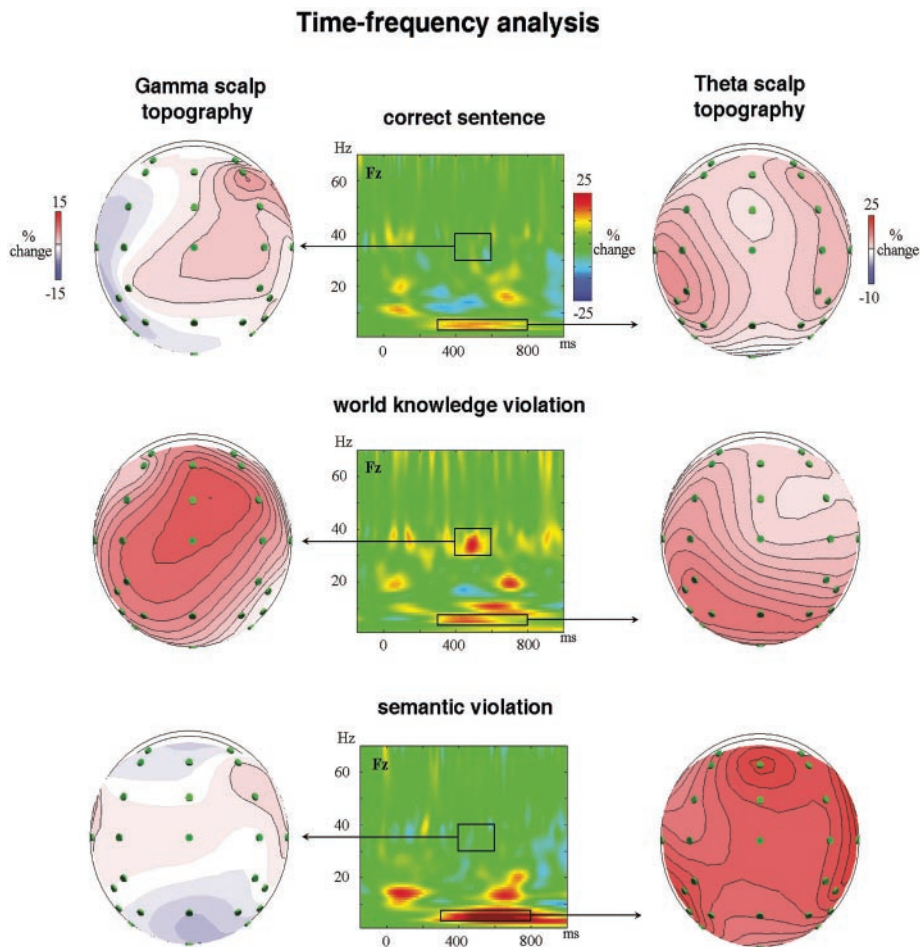
Fig. 2. Morlet wavelet–based TFRs of the induced activity for the three experimental conditions. The x axis represents time; the y axis, frequency. Boxes within the TFRs for a representative electrode (Fz) indicate the areas in which clear gamma and theta responses were observed in a latency range overlapping with the N400. Left of the TFRs are the topographic maps of the gamma activity within the indicated area. On the right, scalp topographies are shown for the theta response in the indicated area. The world knowledge violation showed a peak in the gamma frequency range with a frontocentral distribution, with significant increase in gamma power for eight electrodes (t test, two-tailed, $P < 0.05$). No increases in gamma power were seen in the other conditions. All conditions showed a significant theta response. The theta response was significantly stronger for the semantic violation than for both other conditions [$F(1,29) = 4.2$, $P < 0.05$], based on an analysis with left and right temporoparietal and frontal midline electrode sites as the regions of interest. In addition, only the semantic violation showed a frontal midline theta increase, as was found in a peak-surround analysis for midline electrode Fz [$F(1,29) = 11.2$, $P < 0.005$].

Fig. 3. The common activation for semantic and world knowledge violations compared to the correct condition, based on the results of a minimum T-field conjunction analysis (supporting online text). Both violations resulted in a single common activation ($P = 0.043$, corrected) in the left inferior frontal gyrus, in or in the vicinity of BA 45 (with coordinates [x, y, z] = [−44, 30, 8] and $Z = 4.87$) and BA 47 ([x, y, z] = [−48, 28, −12], $Z = 4.15$). The cross-hair indicates the voxel of maximal activation and has coordinates [x, y, z] = [−44, 30, 8] (left BA 45). L, left hemisphere; R, right hemisphere.

electrical and magnetoencephalogram recordings (27, 28). Our study provides the first evidence for a role of the LIPC in the integration of world knowledge that is represented in long-term memory, next to the integration of lexical semantic knowledge.

Both word meaning and world knowledge are recruited and integrated very rapidly, within some 400 ms, during online sentence comprehension. The LIPC seems to be critical both in the computation of meaning and in the verification of linguistic expressions. Although Frege (1) made an important distinction between the sense of a proposition and its reference, their processing consequences appear to be immediate and parallel. Our results provide evidence against a nonoverlapping two-step interpretation procedure in which first the meaning of a sentence is determined, and only then is its meaning verified in relation to our knowledge of the world. This is compatible with findings of the immediate influence of visual information (29) and of the preceding discourse on the interpretation of phrases and sentences (30). Semantic interpretation is not separate from its integration with nonlinguistic elements of meaning.

In conclusion, while reading a sentence, the brain retrieves and integrates word meanings and world knowledge at the same time. The LIPC is a crucial area for this integration process. Moreover, it does not take longer to discover that a sentence is untrue than to detect that it is semantically anomalous. However, the oscillatory brain responses suggest that the brain keeps a record of what makes a sentence hard to interpret, whether this is word meaning or world knowledge.

References and Notes
1. G. Frege, Z. *Philosophie Philosophische Kritik* **100**, 25 (1892).
2. P. A. M. Seuren, *Western Linguistics: An Historical Introduction* (Blackwell, Oxford, 1998).

lar, the latency of the gamma peak seems to be later than the onset and the peak of the N400 ERP response. This suggests that the oscillatory brain responses reflect post-integration processes.

Gamma oscillations have been associated with feature binding within and across modalities (20, 21). These oscillations are suggested to play an important role in the integration of activity in both local and distributed neural networks (18). It has been suggested that activity in the theta band might involve contributions from the hippocampal complex with, presumably, additional cortical contributions (22). Theta activity has been observed in relation to both episodic and working memory tasks (23). Although the integration of word meaning and world knowledge occurs rapidly and simultaneous-

ly, the different oscillatory responses for semantic and world knowledge violations indicate that the brain seems to keep a record of the nature of the integration problem.

Finally, we ran an event-related fMRI experiment using the same materials (24). The fMRI data, time-locked to the onset of the critical words, revealed an increase in activation in the left inferior prefrontal cortex (LIPC), as compared to correct sentences, that was common to both semantic and world knowledge violations (Fig. 3). The activation was observed in, or in the vicinity of, Brodmann's areas (BAs) 45 and 47. This region has previously been reported in connection to semantic processing (25, 26). This region is also known to contain one of the N400 generators, based on evidence from intracranial

3. R. Jackendoff, *Foundations of Language: Brain, Grammar, Evolution* (Oxford Univ. Press, Oxford, 2003).
4. Although for Frege (*1*), reference was established relative to objects in the world, here we follow Jackendoff's suggestion (*3*) that this is done relative to objects and the state of affairs as mentally represented.
5. S. Zola-Morgan, L. R. Squire, in *The Development and Neural Bases of Higher Cognitive Functions* (New York Academy of Sciences, New York, 1990), pp. 434–456.
6. N. Chomsky, *Reflections on Language* (Pantheon, New York, 1975).
7. J. Katz, *Semantic Theory* (Harper & Row, New York, 1972).
8. D. Sperber, D. Wilson, *Relevance* (Harvard Univ. Press, Cambridge, MA, 1986).
9. K. I. Forster, in *Sentence Processing*, W. E. Cooper, C. T. Walker, Eds. (Erlbaum, Hillsdale, NJ, 1989), pp. 27–85.
10. H. H. Clark, *Using Language* (Cambridge Univ. Press, Cambridge, 1996).
11. Often word meanings can only be fully determined by invoking world knowledge. For instance, the meaning of "flat" in a "flat road" implies the absence of holes. However, in the expression "a flat tire," it indicates the presence of a hole. The meaning of "finish" in the phrase "Bill finished the book" implies that Bill completed reading the book. However, the phrase "the goat finished the book" can only be interpreted as the goat eating or destroying the book. The examples illustrate that word meaning is often underdetermined and necessarily intertwined with general world knowledge. In such cases, it is hard to see how the integration of lexical meaning and general world knowledge could be strictly separated (*3*, *31*).
12. W. Marslen-Wilson, C. M. Brown, L. K. Tyler, *Lang. Cognit. Process.* **3**, 1 (1988).
13. ERPs for 30 subjects were averaged time-locked to the onset of the critical words, with 40 items per condition.

Sentences were presented word by word on the center of a computer screen, with a stimulus onset asynchrony of 600 ms. While subjects were reading the sentences, their EEG was recorded and amplified with a high-cutoff frequency of 70 Hz, a time constant of 8 s, and a sampling frequency of 200 Hz.
14. Materials and methods are available as supporting material on *Science* Online.
15. M. Kutas, S. A. Hillyard, *Science* **207**, 203 (1980).
16. C. Brown, P. Hagoort, *J. Cognit. Neurosci.* **5**, 34 (1993).
17. C. M. Brown, P. Hagoort, in *Architectures and Mechanisms for Language Processing*, M. W. Crocker, M. Pickering, C. Clifton Jr., Eds. (Cambridge Univ. Press, Cambridge, 1999), pp. 213–237.
18. F. Varela *et al.*, *Nature Rev. Neurosci.* **2**, 229 (2001).
19. We obtained TFRs of the single-trial EEG data by convolving complex Morlet wavelets with the EEG data and computing the squared norm for the result of the convolution. We used wavelets with a 7-cycle width, with frequencies ranging from 1 to 70 Hz, in 1-Hz steps. Power values thus obtained were expressed as a percentage change relative to the power in a baseline interval, which was taken from 150 to 0 ms before the onset of the critical word. This was done in order to normalize for individual differences in EEG power and differences in baseline power between different frequency bands. Two relevant time-frequency components were identified: (i) a theta component, ranging from 4 to 7 Hz and from 300 to 800 ms after word onset, and (ii) a gamma component, ranging from 35 to 45 Hz and from 400 to 600 ms after word onset.
20. C. Tallon-Baudry, O. Bertrand, *Trends Cognit. Sci.* **3**, 151 (1999).
21. W. H. R. Miltner *et al.*, *Nature* **397**, 434 (1999).
22. M. Bastiaansen, P. Hagoort, *Cortex* **39** (2003).
23. O. Jensen, C. D. Tesche, *Eur. J. Neurosci.* **15**, 1395 (2002).

24. Whole brain T2*-weighted echo planar imaging blood oxygen level–dependent (EPI-BOLD) fMRI data were acquired with a Siemens Sonata 1.5-T magnetic resonance scanner with interleaved slice ordering, a volume repetition time of 2.48 s, an echo time of 40 ms, a 90° flip angle, 31 horizontal slices, a 64 × 64 slice matrix, and isotropic voxel size of 3.5 × 3.5 × 3.5 mm. For the structural magnetic resonance image, we used a high-resolution (isotropic voxels of 1 mm³) T1-weighted magnetization-prepared rapid gradient-echo pulse sequence. The fMRI data were preprocessed and analyzed by statistical parametric mapping with SPM99 software (http://www.fil.ion.ucl.ac.uk/spm99).
25. S. E. Petersen *et al.*, *Nature* **331**, 585 (1988).
26. B. T. Gold, R. L. Buckner, *Neuron* **35**, 803 (2002).
27. E. Halgren *et al.*, *J. Psychophysiol.* **88**, 1 (1994).
28. E. Halgren *et al.*, *Neuroimage* **17**, 1101 (2002).
29. M. K. Tanenhaus *et al.*, *Science* **268**, 1632 (1995).
30. J. J. A. van Berkum *et al.*, *J. Cognit. Neurosci.* **11**, 657 (1999).
31. P. A. M. Seuren, *Discourse Semantics* (Basil Blackwell, Oxford, 1985).
32. We thank P. Indefrey, P. Fries, P. A. M. Seuren, and M. van Turennout for helpful discussions. Supported by the Netherlands Organization for Scientific Research, grant no. 400-56-384 (P.H.).

Complete Genome Sequence of the Apicomplexan, *Cryptosporidium parvum*

Mitchell S. Abrahamsen,[1,2]*† Thomas J. Templeton,[3]†
Shinichiro Enomoto,[1] Juan E. Abrahante,[1] Guan Zhu,[4]
Cheryl A. Lancto,[1] Mingqi Deng,[1] Chang Liu,[1]‡
Giovanni Widmer,[5] Saul Tzipori,[5] Gregory A. Buck,[6] Ping Xu,[6]
Alan T. Bankier,[7] Paul H. Dear,[7] Bernard A. Konfortov,[7]
Helen F. Spriggs,[7] Lakshminarayan Iyer,[8] Vivek Anantharaman,[8]
L. Aravind,[8] Vivek Kapur[2,9]

The apicomplexan *Cryptosporidium parvum* is an intestinal parasite that affects healthy humans and animals, and causes an unrelenting infection in immunocompromised individuals such as AIDS patients. We report the complete genome sequence of *C. parvum*, type II isolate. Genome analysis identifies extremely streamlined metabolic pathways and a reliance on the host for nutrients. In contrast to *Plasmodium* and *Toxoplasma*, the parasite lacks an apicoplast and its genome, and possesses a degenerate mitochondrion that has lost its genome. Several novel classes of cell-surface and secreted proteins with a potential role in host interactions and pathogenesis were also detected. Elucidation of the core metabolism, including enzymes with high similarities to bacterial and plant counterparts, opens new avenues for drug development.

Cryptosporidium parvum is a globally important intracellular pathogen of humans and animals. The duration of infection and pathogenesis of cryptosporidiosis depends on host immune status, ranging from a severe but self-limiting diarrhea in immunocompetent individuals to a life-threatening, prolonged infection in immunocompromised patients. A substantial degree of morbidity and mortality is associated with infections in AIDS patients. Despite intensive efforts over the past 20 years, there is currently no effective therapy for treating or preventing *C. parvum* infection in humans.

Cryptosporidium belongs to the phylum Apicomplexa, whose members share a common apical secretory apparatus mediating locomotion and tissue or cellular invasion. Many apicomplexans are of medical or veterinary importance, including *Plasmodium*, *Babesia*, *Toxoplasma*, *Neosprora*, *Sarcocystis*, *Cyclospora*, and *Eimeria*. The life cycle of *C. parvum* is similar to that of other cyst-forming apicomplexans (e.g., *Eimeria* and *Toxoplasma*), resulting in the formation of oocysts

[1]Department of Veterinary and Biomedical Science, College of Veterinary Medicine, [2]Biomedical Genomics Center, University of Minnesota, St. Paul, MN 55108, USA. [3]Department of Microbiology and Immunology, Weill Medical College and Program in Immunology, Weill Graduate School of Medical Sciences of Cornell University, New York, NY 10021, USA. [4]Department of Veterinary Pathobiology, College of Veterinary Medicine, Texas A&M University, College Station, TX 77843, USA. [5]Division of Infectious Diseases, Tufts University School of Veterinary Medicine, North Grafton, MA 01536, USA. [6]Center for the Study of Biological Complexity and Department of Microbiology and Immunology, Virginia Commonwealth University, Richmond, VA 23198, USA. [7]MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK. [8]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA. [9]Department of Microbiology, University of Minnesota, Minneapolis, MN 55455, USA.

*To whom correspondence should be addressed. E-mail: abe@umn.edu
†These authors contributed equally to this work.
‡Present address: Bioinformatics Division, Genetic Research, GlaxoSmithKline Pharmaceuticals, 5 Moore Drive, Research Triangle Park, NC 27009, USA.