Success and Virulence in *Toxoplasma* as the Result of Sexual Recombination Between Two Distinct Ancestries

Michael E. Grigg,[1] Serge Bonnefoy,[1]* Adrian B. Hehl,[1]†
Yasuhiro Suzuki,[2,3]‡§ John C. Boothroyd[1]§

Toxoplasma gondii is a common human pathogen causing serious, even fatal, disease in the developing fetus and in immunocompromised patients. Despite its ability to reproduce sexually and its broad geographic and host range, *Toxoplasma* has a clonal population structure comprised principally of three lines. We have analyzed 15 polymorphic loci in the archetypal type I, II, and III strains and found that polymorphism was limited to, at most, two rather than three allelic classes and no polymorphism was detected between alleles in strains of a given type. Multilocus analysis of 10 nonarchetypal isolates likewise clustered the vast majority of alleles into the same two distinct ancestries. These data strongly suggest that the currently predominant genotypes exist as a pandemic outbreak from a genetic mixing of two discrete ancestral lines. To determine if such mixing could lead to the extreme virulence observed for some strains, we examined the F_1 progeny of a cross between a type II and III strain, both of which are relatively avirulent in mice. Among the progeny were recombinants that were at least 3 logs more virulent than either parent. Thus, sexual recombination, by combining polymorphisms in two distinct and competing clonal lines, can be a powerful force driving the natural evolution of virulence in this highly successful pathogen.

Most parasitic protozoa and many prokaryotic pathogens possess a "clonal" population genetic structure consisting of independently propagating, genetically quite divergent clonal lineages (*1–4*). For *Toxoplasma gondii*, a widespread zoonotic pathogen with a well-described sexual cycle in cats, the vast majority of strains (>94%) fall into one of three distinct clonal lines (referred to as "type" I, II, or III) rather than showing the mixing expected of a sexual population (*5, 6*). Meiotic recombination in natural populations of *Toxoplasma* has thus not been considered a major force driving strain diversity and variation.

[1]Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA 94305–5124, USA. [2]Department of Immunology and Infectious Diseases, Research Institute, Palo Alto Medical Foundation, Palo Alto, CA 94301, USA. [3]Division of Infectious Diseases and Geographic Medicine, Department of Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA.

*Present address: Unité d'Immunologie Moléculaire des Parasites, CNRS URA 1960, Institut Pasteur, 25 Rue du Dr Roux, 75724 Paris Cedex 15, France.
†Present address: Institute of Parasitology, University of Zürich, Winterthurerstrasse 266a, CH-8057, Zürich, Switzerland.
‡Present address: Department of Biomedical Sciences and Pathobiology, Virginia Polytechnic Institute, Blacksburg, VA 24061, USA.
§To whom correspondence should be addressed. E-mail: John.Boothroyd@stanford.edu (J.C.B.); ysuzuki@vt.edu (Y.S.)

The three clonal types are apparently not minor or random polymorphic states of no phenotypic consequence: type I lineage strains are highly virulent in outbred mice and perhaps humans (*7*), whereas type II and III lineage strains are relatively avirulent (*8–11*). These two avirulent lines do, however, predominate and readily establish chronic infections in animals and humans (*5, 11*). Occasionally, however, unusual strains are isolated that, based on limited restriction fragment length polymorphism (RFLP) analyses, appear to possess a shuffled combination of alleles similar to those found in the three major lineages (*5, 7, 10*).

The true extent of polymorphism between the three types ("intertypic") and within a given type ("intratypic") could not previously be estimated, because most polymorphism data were from isoenzyme or RFLP analyses, which do not give this level of detail. We have sequenced two loci (*BSR4* and *SAG3*) from six representatives of each of the three major types as well as 13 loci from one representative for each type. The 18 strains come from a variety of hosts (seven species) and three continents, although the majority of isolates are of pan-American origin (*12*). *BSR4* and *SAG3* are single-copy genes encoding surface antigens that are immunogenic in natural infections (*13, 14*) and therefore are expected

to possess higher levels of genetic polymorphism. We have chosen loci encoding antigenic proteins to increase the probability of detecting significant differences. Our analyses, and those from the literature, show that "housekeeping" genes possess much less polymorphism and are thus not so informative when looking for relatedness within the species (*15–19*).

For *BSR4*, all 18 strains have one or the other of only two sequences that differ at 44 positions over the 1.2 kilobases (kb) sequenced, with types I and II possessing one allele and type III the other (Fig. 1A). The complete lack of intratypic polymorphism is remarkable given the considerable breadth of geographic and host-species sampling. For the *SAG3* locus, no intratypic variation and only two sequence classes were found among these same 18 isolates (*20*). In this case, however, types I and III shared essentially the same allele, whereas the type II allele was the outlier. Types I and III did show minor intertypic variation with three unique polymorphisms (relative to the consensus) in the type I allele and four in the type III allele; in contrast, the type II allele possesses 24 polymorphisms.

This total lack of intratypic variation agrees with several studies in which a smaller number of isolates and limited regions were analyzed for *SAG1*, *SAG2*, and *GRA4* (*21–24*). In the case of *SAG1* and *GRA4* coding regions, our expanded sequencing (Table 1) shows that type II and III strains share the identical allele, and type I is the clear outlier with 15 and 24 polymorphisms for the two genes, respectively. For *SAG2*, the situation is similar to that seen for *SAG3*, with the type II allele having virtually all of the polymorphisms (*n* = 12) and the type I and III sequences differing by only one nucleotide (Table 1). Recently, the coding region of *GRA6* was sequenced, and in those strains we believe to be types I, II, or III (23 in total), again, only two allelic classes were identified (*25*). Studies on introns and housekeeping genes in a number of natural isolates also concluded that within-lineage allelic diversity is virtually absent (*15–19*), with the exception of a deletion in the *ROP1* locus in a single lab strain (*26*). The overall lack of intratypic variation, even for the immunogens *BSR4* and *SAG3*, strongly suggests that these three clonal lineages of *Toxoplasma* have emerged as the dominant strains only relatively recently.

To determine the full extent of this dimorphism, we sequenced 15 mostly unlinked loci from archetypal type I, II, and III lineage strains (~65 kb of total sequence) (Table 1). Again, strict dimorphism was found at 7 of the 13 additional loci with only two sequences being found in the three types. For four of

the remaining six loci, the pattern was essentially dimorphic with two of the three sequences sharing an almost identical allele whereas the third was markedly different. This is identified by the subscript that denotes the number of polymorphisms different from consensus. For example, at *GRA3*, the type II allele has 20 differences from consensus, whereas types I and III each differ by just two nucleotides. We refer to these nearly identical alleles (defined as having <0.5% variation from consensus) as being of the same allelic class. By this definition, the remaining two loci, *SRS3* and *GRA2*, are of a single allelic class in all three strains. Thus, all 15 loci have, at most, only two allelic classes which we have designated as "A" and "E," where the A allele is defined as the allelic class shared by at least two of the parasite types. For those six antigenic loci where there is variation in the A allele between types, it is

Fig. 1. Allele relationships by sequence and phylogenetic analysis. (**A**) Polymorphic nucleotide sites at *BSR4* for 28 geographically disparate isolates of *Toxoplasma gondii* (37–39). The "A" allele was found in all six type I and six type II strains as well as ELG and P80. The "E" allele was found in all six type III strains as well as SSI, P89, and TONT. The annotated numerical positions refer to the numbered sites in the published sequence (*BSR4*, GenBank accession AF015290). Sites demarcated by an asterisk indicate unique polymorphic nucleotide positions; R = replacement for nonsynonymous polymorphic sites that result in an amino acid substitution within the protein coding region; S = silent for synonymous substitutions. Dashes indicate identity with the consensus sequence. (**B**) Analysis of 13 strains for *SAG1/SAG2/SAG4* and 28 strains for *SAG3/BSR4* using PAUP4.0b5. The set of 13 strains included a representative of each of the three archetypal strains [types I (RH), II (PRU), and III (CEP)], as well as five strains (P80, P89, SSI, TONT and ELG) that contain different combinations of the canonical A and E alleles and a further five strains representing the most divergent strains known. The set of 28 strains used for the *SAG3* and *BSR4* loci included the same 13 plus an additional five for each of types I, II, and III (none of which showed intratypic variation). Alignment was generated with Neighbor-Joining Distance program defaults (i.e., gap weight = 3.0; gap length weight = 0.1). Statistical support was generated by the Bootstrap Resampling method (1000 resamplings). Bootstrap values are shown above branch lines. Data is presented as a phylogram using midpoint rooting with horizontal distances between groups representing the percentage of nucleotide sequence divergence between samples. The bold line in each phylogram indicates where five or more strains are identical and highlights the dimorphism that predominates at each locus.

BSR4

A

Consensus

B
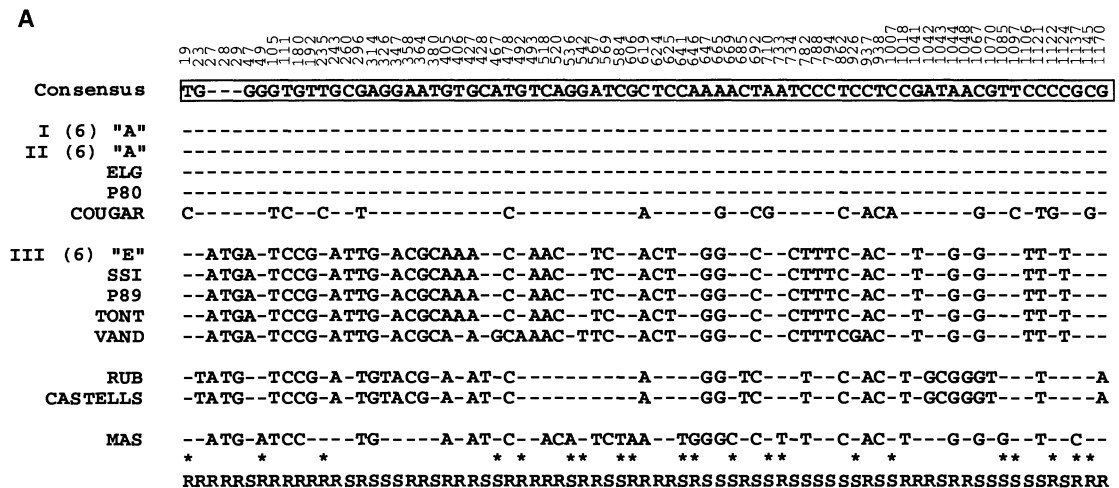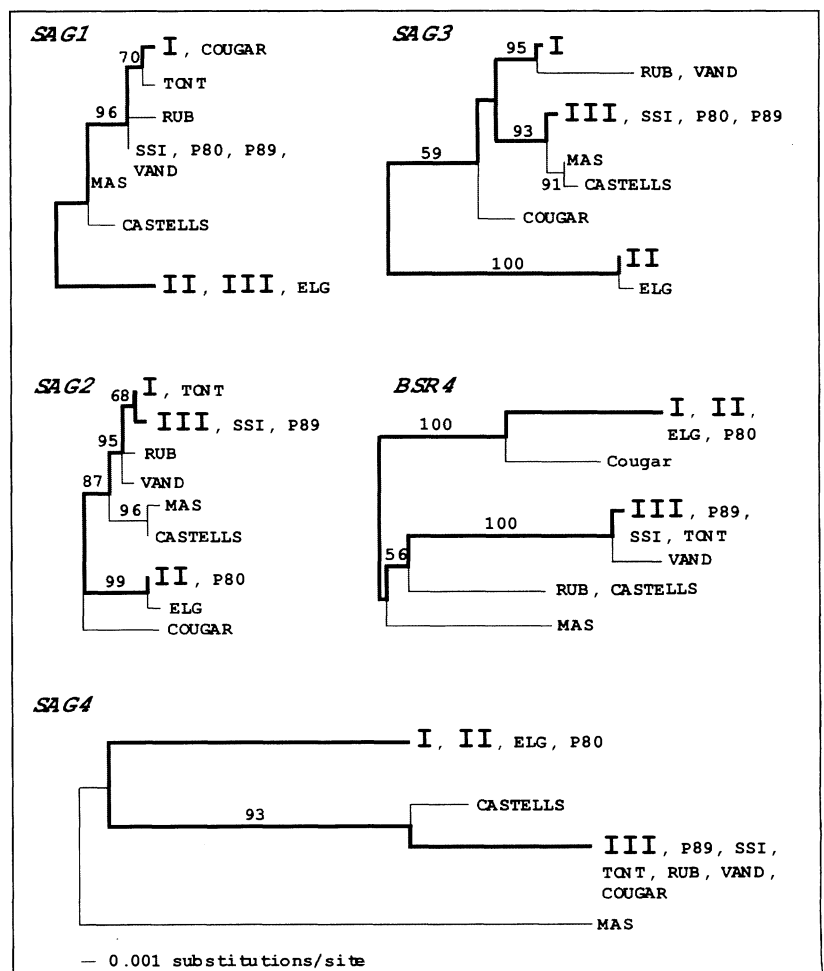
limited to an average of two nucleotides from consensus (<0.2%). In contrast, polymorphism between the A and E alleles is substantial, ranging from ~1% to ~5% for 11 of the loci. At two loci, *SRS1* and *SRS2*, the polymorphism is only just enough (0.5%) to warrant assigning an E allele, but this designation is reinforced by the total absence of variation in the A allele among the other two types. Note that the overall ratio of substitutions in the entire coding region that are silent versus replacements (81:157; Table 1) is about what would be expected by chance genetic drift without selection for or against change in the protein sequence. Overall, these results indicate that there were two, genetically distinct founding lines or populations in the evolution of the three currently predominant lines found at present.

In order to accurately estimate the number of crosses involved in the emergence of these three successful lines, extensive polymorphism data for all three types across most or all of an entire chromosome will be required. Previous analyses of RFLP data for 36 loci gave some suggestion of this dimorphism, although the resolution of those data precluded any firm conclusions except that type II and III strains are apparently more similar to each other than either is to type I (*27, 28*). The data we present do not support that latter association, although too few informative loci have been examined to strongly refute it. Our data do, however, clearly indicate that differences between strains are largely due to segregation rather than drift, and so phylogenetic trees for individual loci should not be used to infer their relatedness.

The predominance of types I, II, and III among isolates collected worldwide could be the result of the species having undergone a recent population bottleneck out of which only a few highly related clones have emerged. For one of the most common protozoan parasites on Earth (at least in terms of warm-blooded vertebrates), it is hard to imagine such a bottleneck arising through chance environmental factors. Instead, our data support the theory that today's most prevalent strains comprise remarkably successful, recombinant genotypes that rapidly and effectively came to dominate the niches examined.

Strains possessing different genotypes from the three predominant lineages have been reported (*5, 6, 9, 29*). These latter "rare" strains, isolated primarily from exotic species or geographically remote regions, might collectively represent a panmictic, diverse gene pool on which the species relies for its ability to expand into new niches. To examine this hypothesis and to determine the extent of the genetic complexity embodied by these rare strains, we sequenced genes from 10 additional isolates, each of which possess novel genotypes as determined by isoenzyme analyses (*6, 29*). Five antigenic loci (*SAG1, SAG2A, SAG3, SAG4, BSR4*) were selected, and sequencing of polymerase chain reaction (PCR)–amplified genomic DNA from these latter strains (~50 kb in total) shows that for half of these rare strains (TONT, SSI, P80, P89, and ELG), the A and E allelic classes are the only ones identified, except that in these strains, they exist in some new, shuffled combination relative to types I, II, and III (Fig. 1B). This is consistent with these rare strains being recombinants between types I, II, and III, and/or the less successful sibling progeny or cousins from the mating(s) that gave rise to the three archetypal lineages.

For the remaining five strains (RUB, MAS, CASTELLS, VAND, and COUGAR), the A and E allelic classes again predominate, but examples of a truly novel allele (i.e., >0.5% polymorphism from either A or E) were also found at some of the loci examined (Fig. 1B). Inclusion of these more "exotic" isolates in the phylogenetic analyses, however, still clustered the vast majority of alleles (with the exception of some from MAS and COUGAR) into the same two distinct lineages even though their inclusion sometimes obscured the strict dichotomy. At the most polymorphic locus, *BSR4*, 21 different single-nucleotide polymorphisms were detected in these five strains (i.e., compared to the A or E alleles), and the majority of these polymorphisms (17 of 21) were possessed by the two most divergent strains: COUGAR and MAS (Fig. 1B). The COUGAR allele clearly shares a common ancestry with the A allele, and the VAND allele is closest to the E allele (Fig. 1B). The alleles in RUB and CASTELLS appear to be chimeric between A and E, but evolution of their sequence can be envisaged by genetic drift. For MAS, however, the most parsimonious explanation for the relationship between alleles at this locus requires meiotic recombination. In the related apicomplexan parasite *Plasmodium falciparum*, the demonstration of high-frequency intragenic recombination at antigenic loci exists as a powerful example of genetic sex driving the evolution of diversity among alleles (*30–32*). These results show that although drifted or chimeric alleles do exist in the most exotic strains at the most variable loci, the dimorphic trend is generally still observed and thus, even the exotic strains seem to derive from intermixing between the proposed two ancestral lineages.

Differences in virulence between the major strain types could be the result of a gradual selection for mutations and/or reassortment of existing alleles following a genetic cross (i.e., without any mutation of those alleles). The latter phenomenon would be analogous to, but very different from, hybrid vigor in diploids, whereby the impact of alleles that are deleterious when homozygous is reduced by out-breeding. In *Toxoplasma*, where all vegetatively growing forms are haploid, such a phenomenon cannot occur. Interactions between alleles at different loci, however, could confer altered biological

Table 1. Allelic dimorphism in the three major lineages of *Toxoplasma gondii* (*37–39*). Consensus is the nucleotide sequence common to at least two of the three archetypal strains. Each subscript identifies the number of unique polymorphisms relative to the consensus. For all 15 loci there are, at most, only two allelic classes, and these have been designated as "A" and "E" where the A allele is defined as the allelic class shared by at least two of the parasite types. Sequences where variation exists within the A allelic class were either reamplified for re-sequencing and/or checked against the *Toxoplasma* EST database for confirmation of the identified polymorphism.

Locus	Chromosome*	Polymorphism from consensus			Sequence length (kb)	Amino acid replacement†
		RH(I)	PRU(II)	CEP(III)		
SAG1 (P30)	VIII	E_{15}	A_0	A_0	1.1	9/10
SAG2A (P22)	VIII	A_0	E_{12}	A_1	1.2	5/8
SAG3 (P43)	II	A_3	E_{24}	A_4	1.2	19/31
SAG4A (P18)	VII	A_0	A_0	E_{15}	0.5	13/15
SAG4B	VII	A_0	A_0	E_5	0.5	1/5
SRS1	VIII	E_{10}	A_0	A_0	2.0	3/7
SRS2	–	E_9	A_0	A_0	1.6	5/5
SRS3	–	A_3	A_0	A_0	1.0	2/3
SRS4	X	A_3	A_1	E_{19}	1.2	16/22
BSR4 (P36)	IV	A_0	A_0	E_{44}	1.2	21/44
GRA1	VIII	A_2	E_9	A_2	0.8	8/12
GRA2 (P28)	X	A_1	A_4	A_1	1.0	4/5
GRA3	X	A_2	E_{20}	A_2	0.7	20/24
GRA4	–	E_{24}	A_0	A_0	1.1	17/20
ROP1	VIII	E_{29}	A_1	A_0	0.6	14/27

*Mapped to chromosome by RFLP analysis using progeny from a II/III genetic cross as per method described (*33*). In order to assign chromosomal linkage, a polymorphism must exist between lineage II and III alleles; hence, *SRS2, SRS3*, and *GRA4* cannot be assigned based on this method. †Fraction of nucleotide polymorphisms within the coding region resulting in nonsynonymous changes.

properties to the progeny, including virulence. To test if this might be the explanation for why the archetypal strains are so different, despite sharing a very limited, common gene pool, we assessed the infectious properties of 16 F_1 progeny from a cross between the nonvirulent ME49 (type II) and CEP (type III) strains (33).

CBA/CaJ mice were infected intraperitoneally with 10^3 tachyzoites of each of the F_1 strains and the two parental lines. The inoculum used was below the median lethal dose (LD_{50}) for the parental strains, and so all mice were expected to survive the acute stage. Although this was indeed the case for the parental strains and most of the F_1 progeny, high mortality was observed during acute infection for the S23 and CL11 strains (Table 2). The S23 strain was, in fact, up to 3 logs more virulent than either parent, and even inocula containing, in theory, a single parasite were lethal. The CL11 strain was also significantly more virulent ($P < 0.005$) than either of the parents, although, with an LD_{50} of ~500, it is not so virulent as S23. There were no differences in invasion rate, growth rate, or time to lyse a culture between the virulent F_1 progeny and their avirulent siblings (20). Mice infected with S23 showed pathology similar to infections with the highly virulent type I strains: high parasite loads in the peritoneal cavity and various organs including the liver, spleen, and lung (12). No tachyzoites were detected in the peritoneal cavity, and neither was significant inflammation seen in all organs examined in mice infected with the nonvirulent sibling S22. The two seropositive S23 survivors possessed a normal cyst burden, thereby ruling out an inability of this strain to differentiate from the disseminating tachyzoite to the cystogenic bradyzoite form as a probable explanation for its heightened virulence.

Because Toxoplasma is haploid and neither parent showed the phenotype, it is highly unlikely that a single locus is responsible for the heightened virulence of S23 and CL11. This possibility cannot be excluded, however, because a hypothetical virulence gene that is phenotypically nonapparent in one parental background could be epistatic to differences in the genotype at one or more other loci. To gain some insight into which genes might be responsible for the virulence, we compared the published genotype of the virulent S23 and CL11 strains to those of the other F_1 progeny (33). No single region is clearly associated with virulence. Comparing S23 and CL11 to the avirulent S22 and CL12 strains, however, identified at least two regions, one on chromosome III and the other on chromosome IV, as possibly involved. The presence of only one of these two regions is clearly not sufficient because, for example, the CEP chromosome III found in S23 is also found in several avirulent F_1 progeny (i.e., CL19, S21, S25, S26, and S30). One or more genes in these regions, however, could interact with loci on chromosome IV derived from the ME49 parent to give the virulent phenotype.

Another possible explanation for the results presented here is that soluble factors secreted from the two parents interact to produce the virulence. To determine if this sort of trans-acting phenomenon is operating, an equal mix of the two parents was used as the inoculum (totaling 10^3 parasites) and the resulting infection was monitored in mice. No mortality was seen with this mix compared to the same number of either parent alone, indicating that it is not the mixture of strain types, per se, that leads to hypervirulence.

The results presented here indicate that through random genetic reassortment, two avirulent strains of Toxoplasma gondii can give rise to highly virulent progeny. The possibility that this is the result of a chance mutation arising in the progeny cannot be formally excluded, but several facts strongly argue against it. First, the recombinant progeny were frozen after their initial expansion and thawed only for these experiments. Second, a change in virulence was seen in 2 of 16 F_1 progeny and their phenotypes were different. And, third, continuous passage of nonvirulent strains over many years has never produced the magnitude of a change in virulence seen with S23. Instead, it would appear that virulence is the result of some combination of alleles at two or more loci. In combination with the population data described above, these results show that recombination can rapidly generate progeny possessing significantly altered biological properties. This suggests that type I strains and the other highly virulent natural recombinant strains so far isolated, including the recently described type IV strain (5–7, 9), do not possess discrete genes or "pathogenicity islands" that inevitably lead to heightened virulence, but rather they have some combination of alleles that cooperate to confer increased pathogenicity, each of which on their own is not an intrinsic virulence allele.

These results are similar to the situation seen with another haploid pathogen, the influenza virus, where once adapted to mice, a mixed infection with two nonvirulent strains can yield neurovirulent reassortants (34). This is the first example of this phenomenon, however, in a nonviral pathogen, and these data have important implications for the nature of virulence and the role of sexual recombination in the evolution of new strains able to take advantage of an ever-changing spectrum of hosts. This differs from the proposed clonal propagation theory held for certain other protozoa, which predicts that extensive genetic divergence exists between clonal lineages, as is the case in Trypanosoma cruzi (2) and perhaps Trypanosoma brucei rhodesiense (35). Instead, our data favor a selective sweep hypothesis whereby a few strains, drawing on a remarkably limited (essentially dimorphic) gene pool, emerge with a more potent assortment of alleles from rare crosses. These recombinants rapidly and effectively come to dominate amidst a background of strains with less optimal (for a given time and place) genotypes, similar to what has been described for Trypanosoma brucei brucei (35). Note that a mixed infection with different strains of Toxoplasma can produce in excess of 10^8 recombinant F_1 progeny from a single cat (36) and so occasionally such infections can have a dramatic effect on the population biology of the species.

Table 2. Mortality in CBA/CaJ mice after intraperitoneal infection with ME49, CEP or the F_1 progeny of a cross between them (33). F_1 (x14) indicates the cumulative results for 14 different F_1 progeny (CL12, CL13, CL16, CL18, CL19, S21, S22, and S25-30) none of which resulted in any death using a dose of 1000 tachyzoites and a total of 85 mice. F_1-S23 and F_1-CL11 correspond to the 2 of 16 F_1 progeny that showed a difference in virulence compared to their siblings and parents. na, not applicable.

Strain of T. gondii	Tachyzoites injected	% Mortality in mice (No. of mice died/ No. of mice injected)	Infected survivors
ME49	1000	0 (0/10)	10/10
CEP	1000	0 (0/15)	15/15
F_1 (×14)	1000	0 (0/85)	20/20*
F_1-S23	1000	100 (28/28)	na
F_1-S23	100	100 (10/10)	na
F_1-S23	10	83 (5/6)	1/1
F_1-S23	1	60 (3/5)	1/2
F_1-CL11	1000	60 (9/15)	6/6
F_1-CL11	100	20 (1/5)	4/4

*Twenty survivors were screened for infection and all were found to be seropositive.

References and Notes
1. J. M. Smith, N. H. Smith, M. O'Rourke, B. G. Spratt, Proc. Natl. Acad. Sci. U.S.A. **90**, 4384 (1993).
2. F. J. Ayala, Proc. Natl. Acad. Sci. U.S.A. **95**, 3346 (1998).
3. M. Tibayrenc, F. Kjellberg, F. J. Ayala, Proc. Natl. Acad. Sci. U.S.A. **87**, 2414 (1990).
4. M. Tibayrenc et al., Proc. Natl. Acad. Sci. U.S.A. **88**, 5129 (1991).

5. D. K. Howe, L. D. Sibley, *J. Infect. Dis.* **172**, 1561 (1995).
6. M. L. Darde, *Curr. Top. Microbiol. Immunol.* **219**, 27 (1996).
7. M. E. Grigg, J. Ganatra, J. C. Boothroyd, T. P. Margolis, *J. Infect. Dis.* **184**, 633 (2001).
8. L. D. Sibley, J. C. Boothroyd, *Nature* **359**, 82 (1992).
9. L. D. Sibley, D. K. Howe, *Curr. Top. Microbiol. Immunol.* **219**, 3 (1996).
10. D. K. Howe, B. C. Summers, L. D. Sibley, *Infect. Immun.* **64**, 5193 (1996).
11. D. K. Howe, S. Honore, F. Derouin, L. D. Sibley, *J. Clin. Microbiol.* **35**, 1411 (1997).
12. Supplementary material is available at www.sciencemag.org/cgi/content/full/294/5540/161/DC1
13. G. Couvreur, A. Sadak, B. Fortier, J. F. Dubremetz, *Parasitology* **97**, 1 (1988).
14. S. Tomavo et al., *Infect. Immun.* **59**, 3750 (1991).
15. K. Luton, M. Gleeson, A. M. Johnson, *Parasitol. Res.* **81**, 310 (1995).
16. W. L. Homan et al., *Parasitol. Res.* **83**, 285 (1997).
17. M. Binas, A. M. Johnson, *Int. J. Parasitol.* **28**, 1033 (1998).
18. A. Fazaeli, P. E. Carter, T. H. Pennington, *J. Parasitol.* **86**, 716 (2000).
19. T. Lehmann, C. R. Blackston, S. F. Parmley, J. S. Remington, J. P. Dubey, *J. Parasitol.* **86**, 960 (2000).
20. M. E. Grigg et al., data not shown.
21. R. Bulow, J. C. Boothroyd, *J. Immunol.* **147**, 3496 (1991).
22. H. Rinder, A. Thomschke, M. L. Darde, T. Loscher, *Mol. Biochem. Parasitol.* **69**, 123 (1995).
23. S. F. Parmley et al., *J. Parasitol.* **80**, 293 (1994).
24. R. Meisel et al., *Mol. Biochem. Parasitol.* **81**, 259 (1996).
25. A. Fazaeli, P. E. Carter, M. L. Darde, T. H. Pennington, *Int. J. Parasitol.* **30**, 637 (2000).
26. D. K. Howe, L. D. Sibley, *Exp. Parasitol.* **78**, 242 (1994).
27. J. C. Boothroyd, A. J. LeBlanc, L. D. Sibley, in *Toxoplasmosis*, J. Smith, Ed. (Springer-Verlag, Heidelberg, Germany, 1993), pp. 63–72.
28. J. C. Boothroyd, *Infect. Agents Dis.* **2**, 100 (1993).
29. M. L. Darde, I. Villena, J. M. Pinon, I. Beguinot, *J. Clin. Microbiol.* **36**, 324 (1998).
30. K. Tanabe, M. Mackay, M. Goman, J. G. Scaife, *J. Mol. Biol.* **195**, 273 (1987).
31. D. J. Conway et al., *Exp. Parasitol.* **73**, 469 (1991).
32. P. J. Kerr, L. C. Ranford-Cartwright, D. Walliker, *Mol. Biochem. Parasitol.* **66**, 241 (1994).
33. L. D. Sibley, A. J. LeBlanc, E. R. Pfefferkorn, J. C. Boothroyd, *Genetics* **132**, 1003 (1992).
34. C. Scholtissek, A. Vallbracht, B. Flehmig, R. Rott, *Virology* **95**, 492 (1979).
35. A. MacLeod et al., *Proc. Natl. Acad. Sci. U.S.A.* **97**, 13442 (2000).
36. L. C. Pfefferkorn, E. R. Pfefferkorn, *Exp. Parasitol.* **50**, 305 (1980).
37. For *BSR4*, the GenBank accession numbers are: AF394603 for RH, AF394604 for PRU, and AF394605 for CEP. For *SAG3*, the GenBank accession numbers are: AF340227 for RH, AF340228 for PRU, and AF340229 for CEP.
38. PCR amplification of all 15 single-copy, polymorphic loci was performed as described (8). Briefly, 35 reaction cycles consisting of 94°C for 30 s, 60°C for 30 s, and 72°C for 1 min was carried out using 1 μl of parasite DNA as template ($\sim10^4$ parasite equivalents). For the BSR4 primers, the annealing temperature was 52°C instead of 60°C. The majority of primers were selected to amplify the coding regions of the respective genes used in this study. The primers utilized for amplification are as follows (all sequences in 5'-3' orientation): *SAG1*, Forward (F): CAATGTGCACCTGTAGGAAGC, Reverse (R): GTGGA-ATTCCTTTGTCGATTTGAG; *SAG2*, F: GAAATGTT-TCAGGTTGCTGC, R: AACGTTTCACGAAGGCACA; *SAG3*, F: GACGAATTCACGAGGGAGCTTGCT, R: GC-GCGTTGTTAGACAAGACA; *SAG4A*, F: TACGATTTC-AAGAAGGCGCT, R: GTCTCGAGCTTCGCAGCATGAT-ACA; *SAG4B*, F: CGTCCTGGTACTCAACGACG, R: GC-CAACCGCAGTCGATTGGT; *SRS1*, F: TCAAGGCATTG-TGCGTGACC, R: TGTCCACTCACGTACCGAAA; *SRS2*, F: CGAGAATGGCGACGCGTGCGTCTT, R: TTCCCC-

ACTCAATAGGCAAGT; *SRS3*, F: CACAACGCGAAAT-GCGCTTA, R: CACATATTGCCATCACGCAT; *SRS4*, F: CTTTCTGGCCTGGTGTTTGT, R: TGCGGATCCCGT-CTGGACAGCTGAAAATT; *BSR4*, F: GACTACTCGAGG-GACGGT, R: CCCAAGGAACTAACAATGA; *GRA1*, F: CGGTTTGCTTGTGTTGTTTG, R: CATGGGGTACGA-TCACAACA; *GRA2*, F: CCTGCGAACTGATGACAGAA, R: CGGCTTTGTAGACCTTCAGC; *GRA3*, F: TACGC-GTCGAGTAACCAGTG, R: AGAGACTGGCACGATGC-TTT; *GRA4*, F: GGAACATGTAGCGTCCACTG, R: AA-TCGCATGCAACGTAACAG; *ROP1*, F: GCGATATG-GCTTGTCGTCAG, R: TTAACTTCGGAGGACCCCGC.
39. PCR products were gel-purified from low-melt agarose gels followed by recovery on glass beads using the UltraClean 15 DNA Purification Kit (MoBio Labs, Inc.). Sequencing was performed at the Stanford PAN Facility on 0.5 to 1 μg of purified DNA using 2 pmol of the forward or reverse PCR primers.
40. We thank P. Basch, J. Remington, and members of the Boothroyd and Falkow lab for helpful suggestions, T. Lehmann for sharing unpublished data, M.-L. Dardé, D. Sibley, and D. Howe for providing parasite strains and lysates for genotyping, L. Guethlein for assistance with the phylogenetic analyses, and S. Falkow, P. Parham, and B. Bohannan for critical review of the manuscript. Supported by grants from NIH (AI21423, AI41014, and AI04717) and from the California County of Alameda District Attorney's Office and the University of California Universitywide AIDS Research Program (R00-PAM-015).

Uniform Binding of Aminoacyl-tRNAs to Elongation Factor Tu by Thermodynamic Compensation

Frederick J. LaRiviere,* Alexey D. Wolfson, Olke C. Uhlenbeck†

Elongation factor Tu (EF-Tu) binds all elongator aminoacyl–transfer RNAs (aa-tRNAs) for delivery to the ribosome during protein synthesis. Here, we show that EF-Tu binds misacylated tRNAs over a much wider range of affinities than it binds the corresponding correctly acylated tRNAs, suggesting that the protein exhibits considerable specificity for both the amino acid side chain and the tRNA body. The thermodynamic contributions of the amino acid and the tRNA body to the overall binding affinity are independent of each other and compensate for one another when the tRNAs are correctly acylated. Because certain misacylated tRNAs bind EF-Tu significantly more strongly or weakly than cognate aa-tRNAs, EF-Tu may contribute to translational accuracy.

Elongation factor Tu (EF-Tu) is a guanine nucleotide binding protein that, when complexed with guanosine 5'-triphosphate (GTP), binds elongator aminoacyl-tRNAs (aa-tRNAs) and participates in the early steps of codon-directed peptide bond formation catalyzed by the ribosome. Tight binding of a tRNA by EF-Tu requires the presence of a cognate amino acid esterified to its 3' terminus by the appropriate aminoacyl–tRNA synthetase (aaRS) (*1*). EF-Tu is generally thought of as a nonspecific binding protein because it binds every elongator aa-tRNA with approximately the same affinity (*2–4*), despite a wide diversity of tRNA sequences, as well as substantial differences in the size, charge, and hydrophobicity of the esterified amino acid. However, EF-Tu binds certain tRNAs esterified with a noncognate amino acid quite differently than it binds the corresponding correctly aminoacylated tRNA. For example, the Su^{+7} suppressor tRNA binds *Escherichia*

coli EF-Tu about threefold as tightly when misacylated with glutamine than when correctly acylated with tryptophan, possibly explaining why glutamine is introduced at amber codons more efficiently than tryptophan (*5*). Another notable example comes from the large number of microorganisms that lack either AsnRS or GlnRS and instead use a nondiscriminating AspRS or GluRS to misacylate tRNA^{Asn} or tRNA^{Gln} followed by a tRNA amidotransferase to synthesize Asn-tRNA^{Asn} and Gln-tRNA^{Gln} (*6*). Two groups have shown that the misacylated Asp-tRNA^{Asn} and Glu-tRNA^{Gln} intermediates in this pathway do not bind to EF-Tu, potentially explaining why misincorporation of aspartate and glutamate does not occur in these organisms (*7, 8*). These results suggested that the correct combination of amino acid and tRNA body are required for efficient EF-Tu binding, although how this specificity is achieved remained unclear.

To further explore the contribution of the amino acid and the tRNA body to EF-Tu binding, variants of *E. coli* tRNA^{Ala}, tRNA^{Val}, tRNA^{Gln}, and yeast tRNA^{Phe}, lacking modified nucleotides, were each amino-acylated with alanine, valine, glutamine, and phenylalanine, creating 4 cognate aa-tRNAs and 12 misacylated aa-tRNAs (*9*). Because

Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309-0215, USA.

*Present address: Brandeis University, Department of Biochemistry, 415 South Street, Waltham, MA 02454-9110, USA.
†To whom correspondence should be addressed. E-mail: olke.uhlenbeck@colorado.edu