Antigen selection based on expression levels during infection facilitates vaccine development for an intracellular pathogen

Claudia Rollenhagen*, Meike Sörensen*, Konstantin Rizos*, Robert Hurvitz†, and Dirk Bumann*‡§

*Department of Molecular Biology and †Core Facility Protein Analysis, Max Planck Institute for Infection Biology, Schumannstrasse 21/22, D-10117 Berlin, Germany; and ‡Department of Immunology, Hannover Medical School, D-30625 Hannover, Germany

Vaccines effective against intracellular pathogens could save the lives of millions of people every year, but vaccine development has been hampered by the slow largely empirical search for protective antigens. *In vivo* highly expressed antigens might represent a small attractive antigen subset that could be rapidly evaluated, but experimental evidence supporting this rationale, as well as practical strategies for its application, is largely lacking because of technical difficulties. Here, we used *Salmonella* strains expressing differential amounts of a fluorescent model antigen during infection to show that, in a mouse typhoid fever model, CD4 T cells preferentially recognize abundant *Salmonella* antigens. To identify a large number of natural *Salmonella* antigens with high expression levels during infection, we used a quantitative *in vivo* screening strategy. Immunization studies with five particularly attractive candidates revealed two highly protective antigens that might permit the development of an improved typhoid fever vaccine. In conclusion, we have established a rationale and an experimental strategy that will substantially facilitate vaccine development for *Salmonella* and possibly other intracellular pathogens.

Infectious diseases represent a tremendous worldwide health problem. Effective vaccines could offer long-term cost-effective solutions but, despite intense efforts, sufficiently efficacious and safe vaccines are still not available for many major pathogens. One important bottleneck is the identification of the few protective antigens among thousands of candidates that can be predicted from genome sequences. For antibody-mediated protective immunity to extracellular pathogens, appropriate antigens must be surface-exposed, and this has been the basis of a highly successful strategy for identifying protective antigens among a small subset of preselected candidates (1). However, for many important pathogens that reside within infected host cells during infection, cellular immune responses are required for protection, and relevant antigen properties for this type of immunity are poorly characterized. As a consequence, identification of protective antigen remains a slow largely empirical process for these pathogens.

Antigen abundance during infection could represent a potentially relevant antigen property, because immune responses are generally dose-dependent (2). Protein species in microbial cells vary in abundance between a few molecules and up to several million molecules per cell (3, 4). The small minority of highly expressed antigens might be preferentially recognized by the host's immune system, and selective testing of this small subset of attractive antigens could permit rapid identification of protective antigens. However, there are two major problems with this potentially attractive approach.

First, direct experimental evidence for preferential T cell responses to abundant antigens is largely lacking because of technical difficulties in quantifying microbial antigen expression in infected animals, and the limited indirect *in vitro* evidence actually does not support an important role of antigen abundance for protective immunity. In particular, antigens that are abundant in pathogen *in vitro* cultures are mostly unsuitable for inducing protective immunity (1). However, *in vitro* cultures imperfectly reproduce the relevant pathogen gene expression patterns in infected hosts so that antigens that are abundant *in vitro* might be weakly expressed *in vivo*. On the other hand, recent *in vitro* studies demonstrated that a single MHC II–peptide complex on the surface of an antigen-presenting cell is sufficient for activating cognate CD4 T cells (5), suggesting that high antigen abundance is not required for potent responses. However, such *in vitro* experiments might not fully reproduce complex *in vivo* situations where phagocytes process a large number of pathogen antigens and present the resulting peptide pool to a diverse T cell repertoire in multicellular microenvironments. It thus remains unclear whether highly *in vivo* expressed antigens represent attractive candidates for vaccine development.

As a second major problem for abundance-based antigen identification, the identity of highly *in vivo* expressed antigens remains largely unknown for most pathogens. Quantitative gene expression analysis of intracellular pathogens in infected tissues is generally hampered by the large excess of host RNA and protein. As an approximation, pathogen transcriptomes (6) and proteomes (7) have been analyzed in cell culture infection models, but such *in vitro* models imperfectly reproduce conditions in infected animals (8–11). Various reporter genes such as chloramphenicol acetyltransferase and GFP can be used to qualitatively detect gene expression in infected animals, and this permitted identification of several *in vivo* expressed genes of *Salmonella* and other pathogens (12, 13). However, because of a lack of quantitative data, it remains unclear whether any of the identified genes belong to the small minority of highly expressed genes. Similarly, many genes that are known to be required for virulence must be expressed at least at some stage of infection, but their expression level remains unknown. Real-time RT-PCR and detection of epitope-tagged proteins using Western blotting (14) are currently the only quantitative methods for *in vivo* expression analysis of individual genes, but prohibitively high costs make these methods unattractive for screening thousands of genes.

In this study, we aimed at resolving these technical problems to test the hypothesis that preselection of highly *in vivo* expressed antigens might substantially facilitate vaccine development for intracellular pathogens. To directly determine the impact of antigen abundance on CD4 T cell activation, we used *Salmonella* expressing different amounts of a fluorescent model antigen. We then developed a quantitative screening strategy that selectively identifies *Salmonella* antigens with high expression levels during infection. Finally, we tested particularly attractive candidate antigens for protective immunity against a lethal *Salmonella* challenge infection.

MICROBIOLOGY

Materials and Methods

Bacterial Strains, Plasmids, and Library Construction. *Salmonella enterica* serovar Typhimurium strain SL3261 *aroA* (15) was used for experiments on CD4 T cell induction to allow for extended survival of infected mice. The virulent wild-type strain SL1344 (15) was used for identification of highly expressed genes and for challenge infection in immunization studies. *Escherichia coli* Electro10 (Stratagene) was used for cloning. Both species were grown in LB medium containing 30 μg·ml^{-1} kanamycin and 100 μg·ml^{-1} ampicillin (*E. coli* transformants) or 90 μg·ml^{-1} streptomycin and 100 μg·ml^{-1} ampicillin or 50 μg·ml^{-1} kanamycin (*Salmonella* transformants). The construction of *Salmonella* SL3261 *aroA* strains with different *in vivo* expression levels of the model antigen GFP_OVA (16) or a superagonistic derivative GFP_OVA^{EA} is described in *Supporting Materials and Methods*, which is published as supporting information on the PNAS web site.

A *Salmonella* promoter-trap library was constructed by inserting fragments (500–700 bp) of sonified SL1344 genomic DNA upstream of a promoterless *gfp_ova* reporter gene on the pBR322-derived plasmid pGFP_OVA (17). Translational readthrough into *gfp_ova* was prevented by stop codons in all three reading frames. Electroporation of the ligation mixture into *E. coli* yielded 1.1×10^6 independent clones, 80% of which carried an insert. A plasmid preparation from the *E. coli* library was electroporated into SL1344 yielding 10^6 clones that represent a >20-fold coverage of the *Salmonella* genome based on insert size, dual orientations, genome size, and Poisson statistics. To minimize clonal competition during *in vitro* growth, the complete library and all sorted sublibraries were grown on plates. To minimize detrimental effects of an enhanced osmosensitivity caused by high GFP levels, LB medium containing a reduced amount of NaCl (4g·liter^{-1}) was used.

Overexpression and Purification of *Salmonella* Antigens. Ten selected *S. enterica* serovar Typhimurium antigens (AroQ, IicA, Mig-14, SsaJ, SsaV, SseB, SifA, SifB, Stm4065, and VirK) were PCR-amplified from genomic DNA of strain SL1344, fused to an N-terminal His-6-tag, and overexpressed in *E. coli* by using the pQE-30 system (Qiagen, Chatsworth, CA). Five antigens (IicA, Mig-14, SseB, SsaJ, and SifB) could be obtained in sufficient quantities and were purified by cobalt affinity chromatography followed by ion exchange chromatography.

Animals, T Cell Adoptive Transfer, Infection, and T Cell Analysis. BALB/c mice and transgenic DO11.10 mice were bred and maintained in a specific pathogen-free animal facility. Ovalbumin-specific DO11.10 CD4$^+$ T cells were adoptively transferred into age- and sex-matched BALB/c mice, as described (16). Alternatively, DO11.10 cells were stimulated *in vitro* with 1 μM OVA peptide (amino acids 323–339) for 4 days before adoptive transfer. One day after transfer of unstimulated DO11.10 T cells or 4 days after transfer of prestimulated DO11.10 T cells, mice were orally infected with $\approx 5 \times 10^8$ colony-forming units (cfu) of *Salmonella*, as described (18). At various time intervals after infection, mice were killed and Peyer's patches were prepared. Aliquots of Peyer's patches single cell suspensions were analyzed for DO11.10 T cell blast formation by using four-color flow cytometry, as described (16). The number of DO11.10 CD4 T cell blasts in the combined Peyer's patches of each mouse was calculated from the total number of lymphocytes and the concentration of large DO11.10 CD4 T cells based on forward scatter (16). Separate aliquots of the same samples were treated with 0.1% Triton X-100 to liberate intracellular *Salmonella* and plated on selective media.

Sorting of Promoter-Trap Library and Clone Characterization. Female 8- to 12-week-old BALB/c mice were systemically infected by tail vein injection with an aliquot of the GFP_OVA promoter-trap library containing $\approx 10^7$ cfu of *Salmonella*. This large systemic inoculum dose was chosen to retain most of the libraries' diversity. More physiological infection conditions were used for validating identified promoters (see below). One day after infection with the library, mice were killed, and spleen was prepared, homogenized, and treated with 0.1% Triton X-100 to liberate intracellular *Salmonella*. GFP-expressing *Salmonella* were sorted with a flow cytometer (FacsDIVA, BD Biosciences, Palo Alto, CA), as described (17), by using 488-nm excitation and green and orange emission channels. To identify promoters with high *in vivo* activity, the *Salmonella* library was sorted twice for high fluorescence in infected spleen (first sort, >20,000 GFP copies per *Salmonella* cell; second sort, >100,000 copies). To obtain a more comprehensive promoter set, independent sort cycles were used to enrich constitutive promoters (first sort, >20,000 copies in spleen; second sort, >50,000 copies during logarithmic *in vitro* growth in LB; third sort, >100,000 copies in spleen) and strongly *in vivo* inducible promoters (first sort, >20,000 copies in spleen; second sort, <2,500 copies during logarithmic *in vitro* growth in LB; third sort, >100,000 copies in spleen). The flow cytometer was calibrated each time with previously established calibration standards (19).

Sorted clones were recovered on plates and characterized by PCR amplification of the inserts with flanking primers "up2" 5′-GTGATGTCGGCGATATAG and "do2" 5′-GAATTGG-GACAACTCCAG followed by restriction mapping with frequently cutting digestion enzymes *Tsp509*I, *Alu*I, and *Hpa*II. One hundred and ten nonredundant inserts were sequenced by using primer "do" 5′-TACTCATATGTATATCTCCTTCTTA, and associated *Salmonella* genes were identified by comparison with the genome sequence of *S. enterica* serovar Typhimurium LT2. Both the primary annotation (20) and the somewhat different annotation of The Institute for Genomic Research (TIGR) (available at www.tigr.org) were used. For seven clones, no associated genes could be identified, and these were excluded from further analysis.

GFP expression of individual clones in infected spleen was analyzed 3–4 days after low-dose (100–200 cfu) i.v. infection. Eight independent replicates revealed a coefficient of variation of 40% (95% confidence interval) for GFP levels of a P_{ssaG} promoter fusion, which was mostly due to variation between individual host animals. The flow cytometer (FacsSort, BD Biosciences) was calibrated each time with previously established calibration standards (19). For promoters present in multiple overlapping inserts, the clone with the smallest distance between the transcriptional fusion junction and the translation start of the first gene of the operon was analyzed.

For nine of the identified promoters (Table 2), the *in vivo* activity was also analyzed with single-copy chromosomal *gfp* fusions. To enhance sensitivity, we used a stable GFP variant [GFP.mut2 (21)] instead of degradable GFP_OVA in these experiments. *gfp.mut2* was PCR-fused to a kanamycin resistance cassette and integrated at the position of the start codon of the first annotated gene downstream of either the identified promoters using the Lambda Red method (22).

Immunization and Challenge Infection. BALB/c mice were s.c. immunized with complete Freund's adjuvant mixed with PBS or 10 μg of either HP0231 (23), Mig-14, IicA, SseB, SsaJ, or SifB, respectively, or complete Freund's adjuvant with a mixture of 10 μg of Mig-14 and 10 μg of SseB. After 4 weeks, mice received a booster immunization with incomplete Freund's adjuvant. After an additional 2 weeks, mice were challenged either intravenously with 200 cfu or intragastrically with 2×10^7 cfu (500 LD$_{50}$) virulent wild-type *S. enterica* serovar Typhimurium
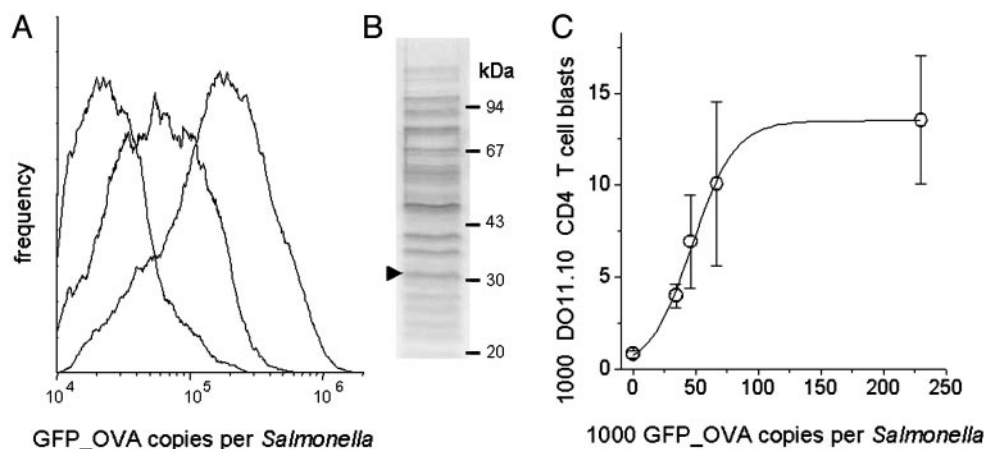
Fig. 1. Relationship between *Salmonella* antigen expression and cognate CD4 T cell activation. (*A*) GFP_OVA expression of different *Salmonella* strains in murine Peyer's patches 5 days after oral infection. (*B*) SDS/PAGE analysis of an *in vitro* culture of *Salmonella* expressing ≈130,000 GFP_OVA copies per cell. The arrowhead indicates the previously identified GFP_OVA band (16). (*C*) Blast formation of ovalbumin-specific DO11.10 CD4[+] T cells 7 days after infection with *Salmonella* expressing different GFP_OVA levels as determined by flow cytometry (16). Means and SEMs from three mice in each group are shown. The number of DO11.10 blasts saturates at GFP_OVA levels exceeding ≈100,000 copies per *Salmonella* cell. Different maximal numbers of DO11.10 T cell blasts were obtained in two additional independent experiments (means of 11,000 and 20,000), but the GFP_OVA expression level required for half-maximal responses was reproducibly ≈50,000 copies per *Salmonella* cell.

SL1344 (15). Five days after systemic challenge infection, mice were killed, and the bacterial load in the spleen was determined by plating. Survival after oral challenge infection was recorded daily for 50 days. Antibodies to Mig-14, IicA, SseB, SsaJ, and SifB in sera obtained before immunization or 2 weeks after the booster immunization were measured by ELISA in 96-well plates coated with 50 ng of antigen per well.

Results

Impact of *in Vivo* Antigen Abundance on CD4 T Cell Induction During *Salmonella* Infection. To directly determine the impact of antigen abundance on the induction of specific CD4 T cells during infection, we used the model antigen GFP_OVA consisting of the GFP at the N terminus and amino acids 319–343 of ovalbumin at the C terminus (16). In contrast to other antigens, microbial GFP_OVA expression can be quantified in infected animals based on its green fluorescence by using two-color flow cytometry (17). The ovalbumin sequence allows tracking of early antigen-specific CD4[+] T cell responses with high sensitivity and temporal resolution using a T cell receptor-transgenic DO11.10 adoptive transfer model (24). We have previously demonstrated in a mouse typhoid fever model that, after oral infection with *Salmonella* expressing GFP_OVA, DO11.10 CD4[+] T cells in the interfollicular regions of Peyer's patches become activated, form blasts, and divide (16).

Here, we used this system to determine how CD4[+] T cell responses to a *Salmonella*-encoded antigen depend on antigen availability. One day after adoptive transfer of 4×10^6 naive ovalbumin-specific DO11.10 CD4 T cells, mice were infected with *Salmonella* strains that express *in vivo* between 10,000 and 230,000 GFP_OVA copies per *Salmonella* cell (Fig. 1*A*) but have otherwise identical properties, including an unaltered colonization capability (19). Flow cytometric analysis of Peyer's patches at 7 days postinfection revealed moderate numbers of DO11.10 CD4 T cell blasts in mice infected with *Salmonella* expressing <35,000 GFP_OVA copies but strong responses to high-expression *Salmonella* strains (Fig. 1*C*). Saturating responses were observed for GFP_OVA expression levels exceeding 100,000 copies per *Salmonella* cell. At such high levels, GFP_OVA is one of the most abundant proteins in *Salmonella* cells (Fig. 1*B*).

To compare the responses of naive and prestimulated T cells,

DO11.10 CD4 T cells were activated *in vitro* with ovalbumin peptide before adoptive transfer. These prestimulated CD4 T cells required ≈60,000 GFP_OVA copies per *Salmonella* cell for saturating *in vivo* reactivation (Fig. 4*A*, which is published as supporting information on the PNAS web site), suggesting that naive and prestimulated DO11.10 CD4 T cells have comparable *in vivo* antigen sensitivities in agreement with previous *in vitro* data (25). Exchange of the natural ovalbumin epitope for the superagonistic epitope variant OVA[E336A] (26) enhanced the antigen sensitivity of naive DO11.10 CD4 T cells (Fig. 4*B*), suggesting that TCR-pMHC affinity modulates *in vivo* CD4 T cell activation in agreement with previous studies (27). However, saturating DO11.10 blast formation still required high levels of ≈35,000 GFP_OVA[E336A] copies per *Salmonella* cell.

These data suggest that CD4 T cells preferentially recognize highly abundant *Salmonella*-encoded antigens. However, TCR transgenic animal models, including the DO11.10 adoptive transfer model, usually use nonphysiologically high precursor frequencies of monospecific T cells that can cause significant T cell competition for cognate peptide–MHC complexes, resulting in a low apparent antigen sensitivity (27). To evaluate this potentially distorting effect in our system, we varied the DO11.10 precursor frequency in infected mice (Fig. 5, which is published as supporting information on the PNAS web site). In case of significant intraclonal T cell competition, increasing the number of DO11.10 precursors should progressively inhibit DO11.10 T cell blast formation. However, within a DO11.10 CD4[+] T cell frequency range of 0.5–3.0 per 100 nontransgenic CD4[+] T cells, the proportion of blast-forming DO11.10 T cells remained rather constant, suggesting a minor role of T cell competition. This is supported by immunhistochemical data from Peyer's patches cryosections that provide little evidence for direct intraclonal DO11.10 T cell interactions during their induction (ref. 16 and data not shown). The observed requirement of rather high antigen levels for potent T cell responses is thus unlikely to be caused by experimental artifacts due to high precursor frequencies in the TCR transgenic model.

Identification of Highly *in Vivo* Expressed *Salmonella* Antigens. Based on the hypothesis of preferential T cell responses to abundant antigens, attractive vaccine antigen candidates could be selected

among highly *in vivo* expressed antigens, but such antigens are difficult to identify using currently available technologies.

One attractive approach for identifying strongly *in vivo* expressed *Salmonella* genes could be based on the use of the GFP as a quantitative reporter for gene expression (28). GFP quantification in infected hosts has been hampered by the preponderance of strongly autofluorescing tissue fragments (29), but our recently developed two-color flow cytometric method allows spectral distinction between GFP fluorescence and background autofluorescence (17), resulting in a rather low detection threshold for *Salmonella* GFP expression in infected mice ($\approx$8,000 instead of >100,000 copies per *Salmonella* cell) (19). We have also shown that for quantifying high promoter activities, degradable GFP variants such as GFP_OVA (Fig. 6, which is published as supporting information on the PNAS web site) are more suitable compared to commonly used stable GFP, because degradation can prevent excessive GFP accumulation impairing *Salmonella* virulence (19). In this study, we combined this improved GFP *in vivo* detection method with the differential fluorescence induction approach (28) to identify highly expressed *Salmonella* genes. Randomly sheared fragments of *Salmonella* genomic DNA were inserted upstream of a plasmid-encoded promoterless *gfp_ova* gene, and the resulting promoter-trap library was transformed into *S. enterica* serovar Typhimurium. We retained the original episomal strategy (28) for promoter screening instead of inserting *gfp* at random positions in the chromosome to prevent inactivation of essential virulence genes (30) and loss of the corresponding promoter fusions during pool infections.

Mice were infected with the *Salmonella* promoter-trap library and 1 day later, *Salmonella* cells containing >100,000 copies of GFP_OVA were sorted from detergent-treated spleen homogenates by using high-speed two-color flow cytometry. The recovered clones were classified by using PCR–restriction fragment length polymorphism to minimize redundancy, and 110 nonredundant DNA-inserts were sequenced, revealing 58 different promoters (Table 1, which is published as supporting information on the PNAS web site). Twenty-two of these promoters were present in several independent overlapping inserts, suggesting that the identified set covers a large fraction of all *Salmonella* promoters with high *in vivo* activity. Moreover, all four promoters identified in a previous small screening using a different promoter-trap library (17) were reidentified in the present study, suggesting high technical consistency.

Previous data qualitatively support the *in vivo* activity of 19 identified promoters (Table 1). In particular, genes downstream of 17 promoters have been demonstrated to be required for full *Salmonella* virulence, suggesting that they are expressed at least at some stage of infection. The activity of eight promoters has been detected in infected mice by using qualitative reporter gene assays. However, our quantitative screening reidentified only two (P_{mgtA}, P_{phoP}) of >100 *in vivo* active promoters that had been identified with the qualitative *in vivo* expression technology (12). The recovery of a distinct promoter subset in this study would be compatible with our goal of selectively identifying some of the few exceptionally active promoters.

To validate the expected high *in vivo* activity of the identified promoters, the corresponding clones were individually characterized in single-clone low-dose infections. Two-color flow cytometry of spleen homogenates revealed transcriptional activities comparable to, or even exceeding, those of prototypic strong ribosomal promoters (Table 1, footnote ‡), which supports a selective recovery of highly active promoters. However, plasmid-encoded transcriptional fusions might not fully reproduce the activity of the corresponding native chromosomal promoters due to differences in copy number, topology, and gene context. To validate our strategy, we therefore chromosomally integrated *gfp* into nine identified loci. Because insertion of foreign genes might
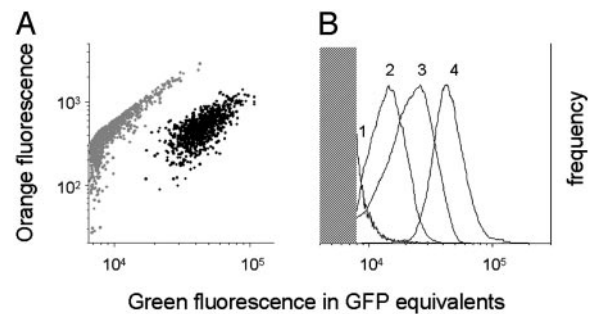

Fig. 2. *In vivo* GFP expression from single-copy transcriptional fusions to chromosomal *Salmonella* promoters. (*A*) Two-color flow cytometry of spleen homogenate of a mouse infected with SL1344 *sifB::gfp*. Gray dots represent autofluorescent host tissue fragments, and black dots represent GFP-containing *Salmonella* cells (17). (*B*) Comparison of GFP expression from various chromosomal *gfp* fusions in infected spleen (1, SL1344 *yjiS::gfp*; 2, SL1344 *mig14::gfp*; 3, SL1344 *nt01st5349::gfp*; and 4, SL1344 *virK::gfp*). The shaded area represents background autofluorescence preventing GFP detection. Only a small bright fraction of *Salmonella* strain 1 was detectable.

disturb the expression of the affected operons, we tested promoters that are likely to play only minor roles in *Salmonella* virulence. Indeed, none of the insertions caused significant *Salmonella* colonization defects in the first 4 days after low-dose infection (data not shown). Eight of nine single-copy constructs expressed GFP levels exceeding the detection threshold of $\approx$8,000 copies per *Salmonella* cell in infected mice (Fig. 2; and Table 2, which is published as supporting information on the PNAS web site), thus confirming the high *in vivo* activity of the respective promoters (30).

Immunization with Highly Expressed *Salmonella* Antigens Induces Protective Immunity. Based on the assumption that CD4 T cells preferentially recognize highly *in vivo* expressed antigens, *Salmonella* proteins expressed from promoters with high activity during infection could represent attractive antigens for vaccination against salmonellosis, which critically depends on CD4$^+$ T cells (31).

To test this hypothesis, we selected 10 particularly attractive antigen candidates. We first reduced the number of candidate operons from 58 to 20 by requiring presence in different *S. enterica* serovars that are relevant for human enteric fever (Typi and Paratyphi A and B) but absence in closely related *E. coli* to avoid immune responses to gut commensals. Among this subset, we selected two proteins from the eight identified highly expressed operons of the PhoP/PhoQ virulence regulon (Mig-14, VirK), five proteins from the 10 identified highly expressed operons associated with the SPI-2 type III-secretion apparatus (the structural proteins SsaJ, expressed from the P_{ssaG} promoter; and SsaV, expressed from the P_{ssaM} promoter; the translocon component SseB, expressed from the P_{sseA} promoter; and the two secreted effector proteins SifA and SifB), an uncharacterized intracellularly induced gene (IicA) (32), as well as the putative aminoimidazol riboside permease (Stm4065) (33) and the periplasmic chorismate mutase AroQ (34). Five of these antigens (Mig-14, IicA, SsaJ, SseB, and SifB) could be obtained as recombinant proteins with sufficient yield and purity (data not shown). An initial small-scale experiment revealed that s.c. immunization of mice with either antigen resulted in specific serum antibody responses (Fig. 7*A*, which is published as supporting information on the PNAS web site). Immunization with either Mig-14, IicA, or SseB, but not SsaJ or SifB, resulted in a significantly lower bacterial load in spleen after a systemic *Salmonella* challenge infection compared to naive control animals (Fig. 7*B*). Because Mig-14 and SseB showed the strongest
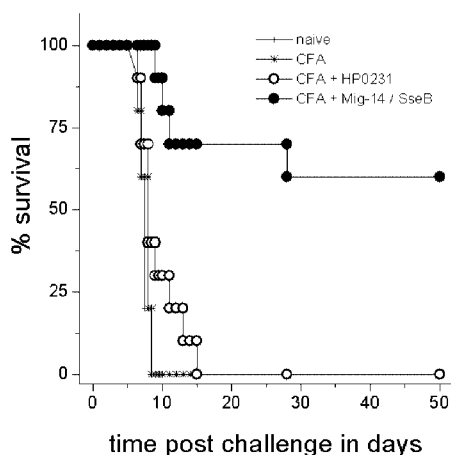
Fig. 3. Immunization with highly *in vivo* expressed *Salmonella* antigens protects mice against a lethal challenge infection with 500 LD$_{50}$ of virulent wild-type *Salmonella*. Survival data are shown for groups of 5 naive mice, 5 sham-immunized mice, 10 mice immunized with an irrelevant *Helicobacter* antigen (HP0231), and 10 mice immunized with a combination of Mig-14 and SseB.

protective effect in this pilot experiment, we validated a combination of these two antigens in a larger experiment for protection against an oral challenge infection with 2×10^{7} cfu virulent wild-type *Salmonella*. Control mice that had been mock-immunized or immunized with an irrelevant recombinantly expressed *Helicobacter pylori* antigen (HP0231) (23) all died within 6–15 days after challenge infection (Fig. 3). In contrast, 6 of 10 mice that had been immunized with the Mig-14/SseB mixture survived the challenge infection for at least 50 days. These results indicated that the *in vivo* abundantly expressed *Salmonella* antigens SseB and Mig-14 are able to induce protective immune responses. Interestingly, the protective efficacy of the mixture is comparable to that of killed whole-cell *Salmonella* vaccines (35), suggesting that SseB and Mig-14 are exceptionally efficacious antigens.

Discussion

The *in vivo* expression level of a microbial antigen could influence its recognition by the host immune system. However, previous evidence for this hypothesis has been scarce because of technical difficulties in quantifying microbial antigen expression in infected animals. In this study, we used a fluorescent model antigen that can be detected with high sensitivity to show that in a murine *Salmonella* infection model, a small minority of highly abundant *Salmonella* antigens appears to be preferentially recognized by both naive and prestimulated CD4 T cells. This finding is supported by the dominant CD4 T cell response to the natural *Salmonella* antigen flagellin (36) during early, but not late, stages of salmonellosis. Flagellin is one of the most abundant proteins in *Salmonella in vitro* cultures but becomes rapidly repressed during infection (6). In parallel with this repression, flagellin-specific CD4 T cell activation ceases (37) and immunization with flagellin induces only partial protection against low-dose *Salmonella* challenge infection (36). Furthermore, flagellin is not a dominant protective antigen in immunization with attenuated live *Salmonella* vaccines (38).

Additional support for preferential cellular immune responses to abundant antigens comes from studies on CD8 T cell activation in other infection models. The dominant protective *Listeria monocytogenes* antigen p60 is poorly recognized by CD8 T cells even if there are as many as 8,000 p60 copies per infected host cell, and saturating CD8 T cell activation appears to require at least 35,000 p60 copies (39). Another study reported dose-dependent CD8 T cell responses to a recombinant vaccinia virus that did not saturate even at exceedingly high antigen levels (40).

The available experimental evidence thus appears compatible with a preferential cellular immune recognition of abundant antigens. As a consequence, selective testing of such antigens could substantially facilitate vaccine development against intracellular pathogens. However, identification of highly *in vivo* expressed antigens has been hampered by technical difficulties. We therefore developed a quantitative GFP-based screening that enabled us to identify 58 *Salmonella* operons with high expression levels in infected mice. Indirect evidence supports the *in vivo* expression of many of the identified operons, and single-copy transcriptional fusions to several chromosomal loci confirmed their high expression in infected mice, which demonstrates the utility of our quantitative screening strategy.

Immunization experiments revealed that three of five selected candidate antigens with high *in vivo* expression induced significant protective immunity against *Salmonella* challenge infection. The protective efficacy of a mixture of two of them was comparable to inactivated whole-cell vaccines, suggesting that the two highly *in vivo* expressed antigens are greatly superior to the vast majority of *Salmonella* antigens that are expressed at lower levels (35). These data thus support the utility of antigen selection based on *in vivo* expression level. On the other hand, the small-scale pilot experiment suggested that two of five tested antigens may induce poor protection despite their ability to induce antibody responses. A limited "hit rate" for identifying protective antigens has also been reported for the widely used "reverse vaccinology" approach for antibody-mediated immunity (usually 2–5% of the selected antigens are actually protective) (1).

The identification of the highly expressed but apparently nonprotective antigens SsaJ and SifB indicates that, in addition to expression level, other antigen parameters may influence protective efficacy. In particular, promoter activities as measured in this study do not always correlate with protein abundance because of posttranscriptional regulation and protein turnover (41). *In vivo* protein quantification is possible using epitope tagging (14), but this demanding technique is difficult to apply to global screening. As another factor, the content, processing, and presentation of MHC class II-restricted T cell epitopes are likely to have a major impact on protectivity. Presentation of T cell epitopes can be analyzed for individual antigens by using T cell *ex vivo* restimulation assays after infection or immunization, but this is not easily possible for thousands of antigens. T cell epitopes can be predicted by *in silico* analysis of primary sequences, but ambiguous results usually require experimental validation. Antigen localization within the *Salmonella* cell might also affect immune recognition, with accessible surface antigens being potentially more protective. However, the putative function of Mig-14 as a transcription factor (42) suggests that at least this partially protective antigen resides in the *Salmonella* cytoplasm. Surface localization can be directly determined by using proteome techniques (43), but this is difficult for *ex vivo* samples. Surface localization can also be predicted *in silico* based on sequence motifs, but experimental data for various bacteria indicate that many surface-exposed antigens actually lack such motifs (1, 43). Taken together, potentially relevant additional antigen properties are currently difficult to analyze on a genome-wide scale and, even for individual antigens, detailed characterization might be actually more time-consuming compared to immunization experiments that yield direct protectivity data. We therefore propose the use of expression level during infection as a powerful and accessible parameter to preselect a small set of promising antigen candidates that is subsequently validated in immunization experiments. The rapid identification of several protective *Salmonella*

antigens based solely on *in vivo* expression levels in this study supports this strategy. On the other hand, as experimental evidence and practical techniques for other relevant antigen parameters might become available, they could be combined with expression level to further improve antigen identification.

The antigens identified in this study might be suitable for the development of improved vaccines against typhoid fever. Whole-cell killed *Salmonella* vaccines efficiently protect humans against typhoid fever but are not currently recommended because of severe adverse reactions. Our data suggest that SseB/Mig-14 might represent an attractive alternative human subunit vaccine that could also resolve major drawbacks of other currently available vaccines, such as poor efficacy in small children, a major target population for typhoid fever vaccination (44), and lack of protection against the increasingly prevalent *S. enterica* serovars Paratyphi A and B (45). In particular, protein antigens (such as SseB and Mig-14) are superior immunogens for small children compared to polysaccharides such as the capsule anti-gen Vi present in the only licensed subunit typhoid fever vaccine. Moreover, both SseB and Mig-14 are highly conserved among various *Salmonella* serovars (20) and thus might induce broad protection against all four different *Salmonella* serovars causing enteric fever in humans (Typhi and Paratyphi A, B, and C), which is not possible using currently licensed vaccines.

Conclusion

This study suggests that *Salmonella* antigens with high expression levels during infection are preferentially recognized by CD4 T cells and can induce protective immunity. Antigen selection based on *in vivo* abundance using a quantitative screening method thus facilitates the development of vaccines against typhoid fever and possibly other infectious diseases.

1. Adu-Bobie, J., Capecchi, B., Serruto, D., Rappuoli, R. & Pizza, M. (2003) *Vaccine* **21,** 605–610.
2. Zinkernagel, R. M., Ehl, S., Aichele, P., Oehen, S., Kundig, T. & Hengartner, H. (1997) *Immunol. Rev.* **156,** 199–209.
3. Pedersen, S., Bloch, P. L., Reeh, S. & Neidhardt, F. C. (1978) *Cell* **14,** 179–190.
4. Ghaemmaghami, S., Huh, W. K., Bower, K., Howson, R. W., Belle, A., Dephoure, N., O'Shea, E. K. & Weissman, J. S. (2003) *Nature* **425,** 737–741.
5. Irvine, D. J., Purbhoo, M. A., Krogsgaard, M. & Davis, M. M. (2002) *Nature* **419,** 845–849.
6. Eriksson, S., Lucchini, S., Thompson, A., Rhen, M. & Hinton, J. C. (2003) *Mol. Microbiol.* **47,** 103–118.
7. Burns-Keliher, L., Nickerson, C. A., Morrow, B. J. & Curtiss, R. (1998) *Infect. Immun.* **66,** 856–861.
8. Eriksson, S., Bjorkman, J., Borg, S., Syk, A., Pettersson, S., Andersson, D. I. & Rhen, M. (2000) *Cell Microbiol.* **2,** 239–250.
9. Watson, P. R., Paulin, S. M., Jones, P. W. & Wallis, T. S. (2000) *Microbiology* **146,** 1639–1649.
10. Clark, M. A., Reed, K. A., Lodge, J., Stephen, J., Hirst, B. H. & Jepson, M. A. (1996) *Infect. Immun.* **64,** 4363–4368.
11. Kok, M., Buhlmann, E. & Pechere, J. C. (2001) *Microbiology* **147,** 727–733.
12. Heithoff, D. M., Conner, C. P., Hanna, P. C., Julio, S. M., Hentschel, U. & Mahan, M. J. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 934–939.
13. Valdivia, R. H. & Falkow, S. (1997) *Science* **277,** 2007–2011.
14. Uzzau, S., Figueroa-Bossi, N., Rubino, S. & Bossi, L. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 15264–15269.
15. Hoiseth, S. K. & Stocker, B. A. (1981) *Nature* **291,** 238–239.
16. Bumann, D. (2001) *Infect. Immun.* **69,** 4618–4626.
17. Bumann, D. (2002) *Mol. Microbiol.* **43,** 1269–1283.
18. Bumann, D. (2001) *Infect. Immun.* **69,** 7493–7500.
19. Wendland, M. & Bumann, D. (2002) *FEBS Lett.* **521,** 105–108.
20. McClelland, M., Sanderson, K. E., Spieth, J., Clifton, S. W., Latreille, P., Courtney, L., Porwollik, S., Ali, J., Dante, M., Du, F., *et al.* (2001) *Nature* **413,** 852–856.
21. Cormack, B. P., Valdivia, R. H. & Falkow, S. (1996) *Gene* **173,** 33–38.
22. Datsenko, K. A. & Wanner, B. L. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 6640–6645.
23. Sabarth, N., Hurwitz, R., Meyer, T. F. & Bumann, D. (2002) *Infect. Immun.* **70,** 6499–6503.
24. Pape, K. A., Kearney, E. R., Khoruts, A., Mondino, A., Merica, R., Chen, Z. M., Ingulli, E., White, J., Johnson, J. G. & Jenkins, M. K. (1997) *Immunol. Rev.* **156,** 67–78.
25. Kimachi, K., Croft, M. & Grey, H. M. (1997) *Eur. J. Immunol.* **27,** 3310–3317.
26. Janssen, E. M., van Oosterhout, A. J., van Rensen, A. J., van Eden, W., Nijkamp, F. P. & Wauben, M. H. (2000) *J. Immunol.* **164,** 580–588.
27. Kedl, R. M., Kappler, J. W. & Marrack, P. (2003) *Curr. Opin. Immunol.* **15,** 120–127.
28. Valdivia, R. H. & Falkow, S. (1996) *Mol. Microbiol.* **22,** 367–378.
29. Lee, S. H. & Camilli, A. (2000) *Curr. Opin. Microbiol.* **3,** 97–101.
30. Hautefort, I., Proenca, M. J. & Hinton, J. C. (2003) *Appl. Environ. Microbiol.* **69,** 7480–7491.
31. Hess, J., Ladel, C., Miko, D. & Kaufmann, S. H. (1996) *J. Immunol.* **156,** 3321–3326.
32. Pfeifer, C. G., Marcus, S. L., Steele-Mortimer, O., Knodler, L. A. & Finlay, B. B. (1999) *Infect. Immun.* **67,** 5690–5698.
33. Dougherty, M. & Downs, D. M. (2003) *J. Bacteriol.* **185,** 332–339.
34. Calhoun, D. H., Bonner, C. A., Gu, W., Xie, G. & Jensen, R. A. (2001) *Genome Biol.* **2,** 1–16.
35. Harrison, J. A., Villarreal-Ramos, B., Mastroeni, P., Demarco, d. H. & Hormaeche, C. E. (1997) *Immunology* **90,** 618–625.
36. McSorley, S. J., Cookson, B. T. & Jenkins, M. K. 2000) *J. Immunol.* **164,** 986–993.
37. McSorley, S. J., Asch, S., Costalonga, M., Reinhardt, R. L. & Jenkins, M. K. (2002) *Immunity* **16,** 365–377.
38. Kodama, C. & Matsui, H. (2004) *Infect. Immun.* **72,** 2449–2451.
39. Vijh, S., Pilip, I. M. & Pamer, E. G. (1998) *J. Immunol.* **160,** 3971–3977.
40. Wherry, E. J., Puorro, K. A., Porgador, A. & Eisenlohr, L. C. (1999) *J. Immunol.* **163,** 3735–3745.
41. Lee, P. S., Shaw, L. B., Choe, L. H., Mehra, A., Hatzimanikatis, V. & Lee, K. H. (2003) *Biotechnol. Bioeng.* **84,** 834–841.
42. Valdivia, R. H., Cirillo, D. M., Lee, A. K., Bouley, D. M. & Falkow, S. (2000) *Infect. Immun.* **68,** 7126–7131.
43. Sabarth, N., Lamer, S., Zimny-Arndt, U., Jungblut, P. R., Meyer, T. F. & Bumann, D. (2002) *J. Biol. Chem.* **277,** 27896–27902.
44. Kossaczka, Z., Lin, F. Y., Ho, V. A., Thuy, N. T., Van Bay, P., Thanh, T. C., Khiem, H. B., Trach, D. D., Karpas, A., Hunt, S., *et al.* (1999) *Infect. Immun.* **67,** 5806–5810.
45. Arya, S. C. & Sharma, K. B. (1995) *Vaccine* **13,** 1727–1728.