Vaccination and the population structure of antigenically diverse pathogens that exchange genetic material

SUNETRA GUPTA, NEIL M. FERGUSON AND ROY M. ANDERSON

The Wellcome Trust Centre for the Epidemiology of Infectious Disease, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

SUMMARY

Populations of antigenically diverse pathogens undergoing genetic exchange may be categorized into strains on the basis of a set of principal protective antigens. The extent to which polyvalent vaccines based on these protective antigens can alter the population structure of the pathogen is determined by the degree of cross-protection between strains. In the case where there is no cross-protection, vaccinating against a particular strain will have no effect on the others. As cross-protection increases, the strains containing the antigenic variants included in the vaccine will be diminished in prevalence, and those that do not will increase in prevalence. The rise in prevalence of the latter will become more and more exaggerated as cross-protection increases. However, beyond a critical level of cross-protection, in the absence of vaccination, the steady state of the system is asymmetric in that a certain subset of strains (with non-overlapping repertoires of antigenic variants) will dominate over the others in terms of prevalence. Under these circumstances, a vaccine consisting of the most immunogenic combinations of antigenic variants can cause a dramatic increase in frequency of a subset of rare strains.

1. INTRODUCTION

The key protective antigens of many important pathogens in human communities are often highly variable, such as those of *Plasmodium falciparum*, *Neisseria meningitidis* and the human immunodeficiency virus (HIV). Two options are available in designing vaccines against such antigenically diverse organisms. Vaccines may be developed that artificially boost immune responses against determinants common to all strains. This strategy is complicated by the fact that these immune responses do not appear to have a substantial protective effect under natural conditions, unless repeated exposure occurs such that cross-protective immunity develops slowly over many years.

An alternative is to consider vaccine preparations involving the highly immunogenic variable antigens, but this has the disadvantage that the entire spectrum of antigenic types may not be covered. Such multivalent vaccines offering partial protection are been considered in a number of cases. One such example is *Neisseria meningitidis*, where subtype-specific sequences of the subcapsular outer membrane protein PorA play a dominant role in the induction of bactericidal antibodies. Por A has two principal antigenic regions VR1 and VR2 (Maiden *et al*. 1991). It has been established that the human immune system is capable of mounting a bactericidal immune response to VR1 and VR2 epitopes (Wiertz *et al*. 1991), which are consequently being seen as candidates for new vaccines against *N. meningitidis*. A genetically engineered construct comprising two such strains (van der Ley & Poolman 1992) is currently under trial.

Another example is *Streptococcus pneumoniae*, where large-scale efficacy trials are currently being conducted of vaccines consisting of conjugated pneumococcal polysaccharides of only a selection of the several serotypes predominantly causing disease in infants and young children (Watson *et al*. 1995). Organisms that exist as a set of strains would thus be subjected to selection against some, but not all, strains. In the case where there is virtually no inter-strain competition (due to lack of cross-immunity, for instance) and no genetic exchange (i.e. the clonal case), it is clear that such a strategy would be effective in eradicating or greatly reducing the abundance of a certain proportion of strains and would have almost no effect on the others, with the overall consequence of reducing infection, and (unless there was a strong conserved response protecting against disease) the total disease burden. In this paper, we consider the effects of utilizing such vaccines within a system where pathogens exchange genetic material and where different levels of cross-immunity operate between the constituent strains.

The rational design of vaccines involves identifying a critical set of protective antigenic targets among the various elements that are presented to the host's immune system. Recent advances in molecular methods have made it possible to pick out the precise coding sequences of defined epitopes within key protective antigens, using a number of direct and indirect

Proc. R. Soc. Lond. B (1997) **264**, 1435–1443
Printed in Great Britain

strategies. The choice of epitopes and antigens for inclusion in such subunit vaccines is often determined by quantitative assessments of immune responses to these determinants observed in immune individuals or in immunogenicity studies in naive individuals (Ada 1995). A previous study (Gupta *et al*. 1996) indicates that the pathogen population may be structured by the host's immune responses to key protective antigens into an equilibrium state where certain combinations of antigens or 'strains' dominate, in terms of prevalence, over the other 'strains', despite the occurrence of frequent recombination events. As a multivalent vaccine whose design is based solely on laboratory studies may include either the high-frequency ('dominant') or low-frequency ('subdominant') strains, we will categorize the results according to the pre-vaccine prevalence of the strains included in the vaccine. Our aim is to determine how the use of multivalent vaccines targeted at highly immunogenic antigens can alter the balance between the frequencies of the parasite strains.

2. A SIMPLE MODEL FOR VACCINATION WITHIN RECOMBINING PATHOGENS

We consider a pathogen population where a strain is defined by n loci, each with a number of alleles. Each combination of alleles at the different loci thus constitutes a strain. For example, in the case where there are two immunologically dominant loci, each with two alleles or variants, the four possible types of strains are *ay*, *ax*, *bx* and *by*, where *a* and *b* are alleles at one locus, and *x* and *y* are alleles at the second locus. Strains that do not share any alleles do not interfere with each others' transmission, because the immune responses directed against one strain, say *ax*, are ineffective against the other, say *by*, as neither anti-*a* nor anti-*x* responses will recognize either *b* or *y*. By contrast, for strains that do share alleles, cross-protection may range in effectiveness from none to complete, depending on the fate of a particular strain, say *ax*, within a host who has protective immune responses either against *a* or *x*, but not both (due to exposure to either *ay* or *bx*). We define a cross-protection parameter, γ, measuring the degree to which infection with a given strain limits the transmission of strains that share any of its alleles. If $\gamma = 0$ then the strains do not interact, whereas if $\gamma = 1$ then there is total cross-protection between the strains. We assume, without any loss of generality, that immunity to a given strain i does not modify the probability of infection by any other strain j, but only reduces the probability of transmission of that strain by a factor $1 - \gamma$. Hence the process of acquisition of immunity to a given strain is independent of the other strains, and the proportion immune to a given strain i, z_i, is simply given by

$$\frac{\mathrm{d}z_i}{\mathrm{d}t} = (1 - z_i)\Lambda_i - \mu z_i; \tag{1}$$

where $\Lambda_i = \lambda_i + v_i$ represents the per capita rate of acquistion of immunity for the fraction of the population, $1 - z_i$, who are susceptible to strain i. (See Appendix 1 for the relationship between the model described here and the full SEIR model for a pathogen with M strains.) Immunity to strain i can be acquired through two routes: by vaccination, at a rate v_i; or by exposure to the wild agent, at a rate λ_i, the force of infection of strain i. In either case, for simplicity we assume that immunity is lifelong. Therefore, individuals only stop being immune upon death, which occurs at a per capita rate of μ.

The proportion of the population who are 'completely susceptible' to strain i are those who have not been infected either by i or any strain sharing alleles with i. There are then a second class of individuals who have a reduced probability of becoming infectious with strain i after exposure: namely, those who have already been exposed to a strain that shares alleles with i. It is therefore necessary to keep track of those individuals who have been exposed to strain i or any strains sharing alleles with i. If we denote the proportion of the population in this class as w_i, its dynamics are simply given by:

$$\frac{\mathrm{d}w_i}{\mathrm{d}t} = (1 - w_i)\sum_j \Omega_{ij}\Lambda_j - \mu w_i, \tag{2}$$

where $\Omega_{ij} = 1$ if strains i and j share alleles (including the case where $i = j$), and $\Omega_{ij} = 0$ otherwise.

The fraction of the population that are completely susceptible to strain i is then just $1 - w_i$, so the rate at which this class of individuals becomes infectious is $(1 - w_i)\lambda_i$. The fraction of the population in the second susceptibility class (those that have not been exposed to i but have been exposed to a strain sharing alleles with i) is $w_i - z_i$. The rate at which this class becomes infectious is therefore $(1 - \gamma)(w_i - z_i)\lambda_i$. Assuming both groups lose infectiousness at a per capita rate of σ, the dynamics of the proportion of the population infectious for strain i, y_i, are thus given by

$$\frac{\mathrm{d}y_i}{\mathrm{d}t} = [(1 - w_i) + (1 - \gamma)(w_i - z_i)]\lambda_i - (\sigma + \mu)y_i. \tag{3}$$

Figure 1 schematically represents the effect of cross-protection for strain *ax* in the two-locus two-allele case.

For this paper, we reduce the complexity of the set of equations by making the approximation that the probability that an individual is immune to one strain is independent of their immunity to any other strain (Gupta *et al*. 1996). Hence the proportion of the population susceptible to strain i, and to any strain sharing alleles with i, is approximated by the product of the proportions susceptible to the individual strains, $\prod_j (1 - \Omega_{ij}z_j)$. Hence w_i can be approximated by

$$w_i \simeq 1 - \prod_j (1 - \Omega_{ij}z_j), \tag{4}$$

meaning the time evolution of w_i no longer needs to be explictly modelled (i.e. equation (2) can be dropped). This results in a significant drop in computational complexity, especially for systems with many antigenically active loci, each with many alleles.

The output of the models will be represented in terms of the proportion of the population with natural immunity to strain i, z_i^n, which is given by

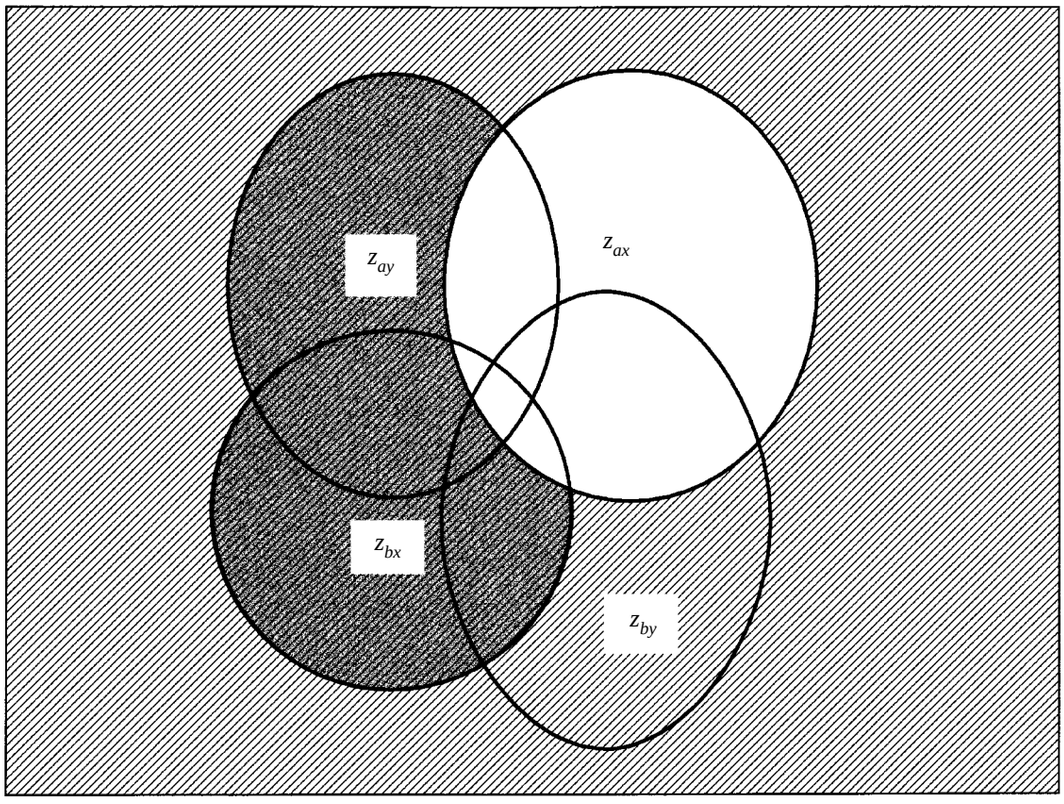

Figure 1. A schematic representation of the relationship between the proportions immune to strains sharing alleles in the two-locus two-allele case. We focus on the strain *ax*. Those immune to strain *ax* (z_{ax}) are shown in white. The area outside of z_{ax} is shown by the diagonal hatching. Within this area, the region marked in grey represents individuals who are not immune to *ax* but do have immunity to a strain that shares alleles with *ax* (i.e. *ay* or *bx*). Individuals in this category will transmit *ax* with probability $1-\gamma$ where γ is the degree of cross protection between strains sharing alleles. In the absence of cross-protection, the entire hatched region will be susceptible to *ax*.

$$\frac{\mathrm{d}z_i^n}{\mathrm{d}t} = (1 - z_i)\lambda_i - \mu z_i^n \qquad (5)$$

The dynamics of strain z_i^n are thus determined by the force of infection for strain i alone, whereas the total proportion immune, z_i, is determined by both the force of infection and the per capita rate of vaccination.

The impact of genetic exchange on the population structure of infectious disease agents may be examined within this framework by modifying the force of infection term, λ, to include the assumption that the progeny of parasites within hosts infectious for two or more strains will consist of defined fractions, α_{jk}, of the various combinations of the different strains, j and k, that may generate strain i through recombination (see Appendix 1). As the proportions that are infectious are very small ($\sigma \gg \mu$) the force of infection of strain i may be approximately represented as: $\lambda_i = \beta_i(y_i + \sum_{jk} \alpha_{jk} y_j y_k)$, where β_i is a combination of parameters affecting the transmission of strain i. The behaviour of the model is largely unaffected by the precise functional form of the recombination term. The basic reproductive number of strain i (the number of secondary infections generated by an infectious individual within a pool of susceptibles (Anderson & May 1991)), R_0^i, can be shown to be β_i/σ.

3. MODEL BEHAVIOUR WITHOUT VACCINATION

It can be demonstrated analytically that there are two types of equilibria for this system. These are illustrated in figure 2, which records the changes with time in the proportion immune to the four strains in the two-locus two-allele case, for different levels of cross protection (i.e. (*a*) $\gamma = 0$, (*b*) $\gamma = 0.75$, (*c*) $\gamma = 0.8$, (*d*) $\gamma = 1$). When there is no cross-protection, the proportion immune to a given strain, z_i ($= z_i^n$, as immunity can only be acquired naturally), will simply equilibrate independently of the other strains at $1 - 1/R_0^i$. This is shown in figure 2*a* for the strains *ax* ($R_0 = 5$), *ay* ($R_0 = 4.5$), *bx* ($R_0 = 4$), and *by* ($R_0 = 3.5$) respectively. As the degree of cross-protection increases, the equilibrium prevalence of each strain declines as a result of immunological interference by other strains; nonetheless they are maintained in the system at levels commensurate with their respective basic reproductive rates. However, such a symmetric solution is unstable at $\gamma = 1$, and as shown in figure 2*d*, it is replaced by an asymmetric solution, where one set of strains dominates over the other types. This situation occurs even when all strains have the same basic reproductive rate. The strains can be divided into two sets of non-overlapping or 'discordant' pairs *ax*,*by*

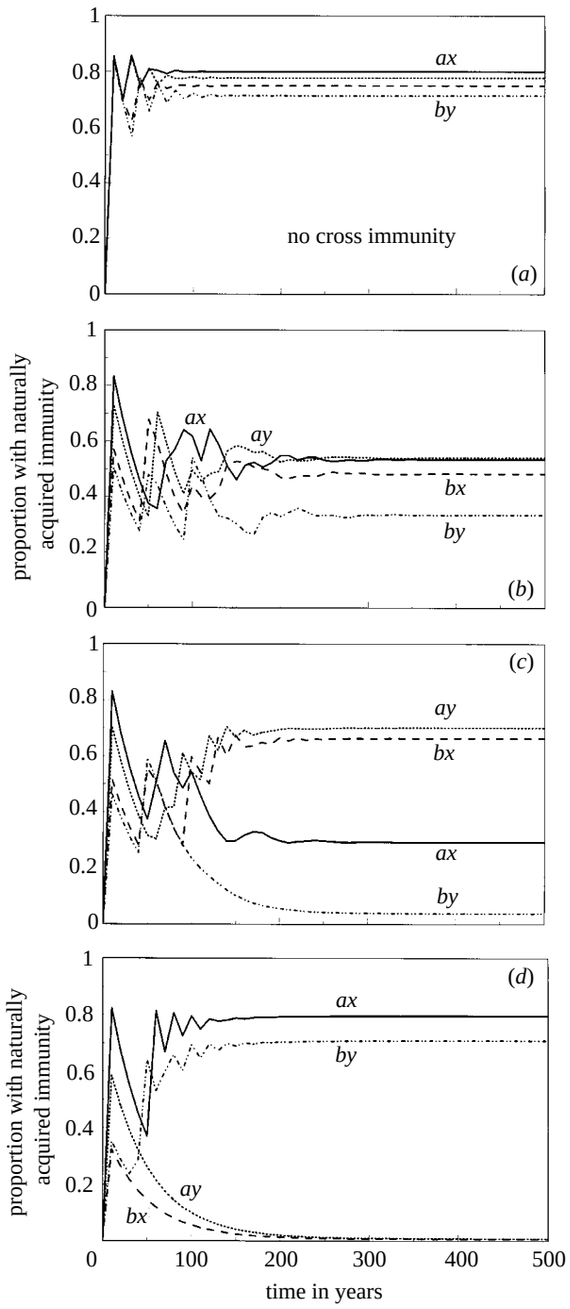

Figure 2. The changes with time in the proportions of the host population immune to the four strains strains *ax* ($R_0 = 5$), *ay* ($R_0 = 4.5$), *bx* ($R_0 = 4$) and *by* ($R_0 = 3.5$), for the following levels of cross protection: (*a*) $\gamma = 0$, (*b*) $\gamma = 0.75$, (*c*) $\gamma = 0.8$, (*d*) $\gamma = 1$.

and *ay*,*bx* that share no alleles with each other. It may be shown that provided the degree of cross-protection is high, one set of discordant strains will always strongly dominate, in terms of abundance, over the other discordant set. There thus exists a critical level of cross protection, γ_{crit}, in which the symmetric equilibrium becomes unstable and bifurcates to give upper and lower asymmetric branches (Gupta *et al.* 1996). The critical value of γ, calculated for the general two-locus two-allele case, using the approximation represented by equation (4), is:

$$\gamma_{\text{crit}} = \frac{1 - 1/R_0^j}{1 - 1/R_0^i R_0^j}$$

Here R_0^i and $R_0^{i'}$ are the basic reproductive rates of the dominant pair of strains and R_0^j is the lower value of the subordinate pair. Which pair will dominate beyond the critical level of cross-protection is determined by the combination of basic reproductive rates and may in fact change according the precise value of γ. This is demonstrated by the contrast between figures 2*c* and 2*d*. In the latter ($\gamma = 1$), *ax* and *by* dominate by virtue of *ax* having the highest R_0; by contrast, just beyond the critical level of cross-protection ($\gamma_{\text{crit}} = 0.75$), *ay* and *bx* dominate by virtue of having a higher average R_0 (figure 2*c*). In the following section we will examine the effects of vaccinating with various combinations of strains under these different equilibrium conditions.

4. MODEL BEHAVIOUR WITH VACCINATION

This section is organized into two parts, as the equilibrium population structure of the parasite changes dramatically at a critical value of γ, with two classes of dynamical behaviour (namely, symmetric and asymmetric solutions). In the first part, we discuss the consequences of introducing vaccination against a single strain in the situation where a symmetric equilibrium prevails due to cross-protection being lower than the critical level. In the second part, we consider the effects of vaccinating against the dominant and the subdominant types within an asymmetric equilibrium (i.e. where non-overlapping strain structure has been established due to strong cross-protection between parasite types sharing relevant immunogens).

(*a*) *Symmetric equilibrium:* $\gamma < \gamma_{\text{crit}}$

If the level of cross-protection is not high enough to precipitate an asymmetric equilibrium, all the parasite types will cocirculate at frequencies commensurate with their basic reproductive rates. We consider here the consequences of vaccinating against a selection of strains, between the limits of no cross-protection and the critical threshold where cross-protection is sufficient to destabilize the symmetric equilibrium. In the limit where there is no cross-protection ($\gamma = 0$), vaccinating against any combination of strains will simply result in the selective elimination of those types without any effect on the remaining types. The proportion naturally immune to the latter will thus remain at $1 - 1/R_0^i$, and the proportion naturally immune to the strains included in the vaccine will fall to zero (provided coverage and efficacy are adequate), as the proportion of the population immune to these strains as a result of vaccination rises. As cross-protection increases, the existence of a large proportion of the population immune by vaccination against a particular strain, say *ax*, will begin to have an effect on other parasite types sharing alleles with *ax* (i.e. *ay* and *bx*). At the same time the steady state prevalence of each strain (i.e. the prevalence at the time the vaccination programme is implemented) will decrease as γ increases due to competition between the strains. Thus

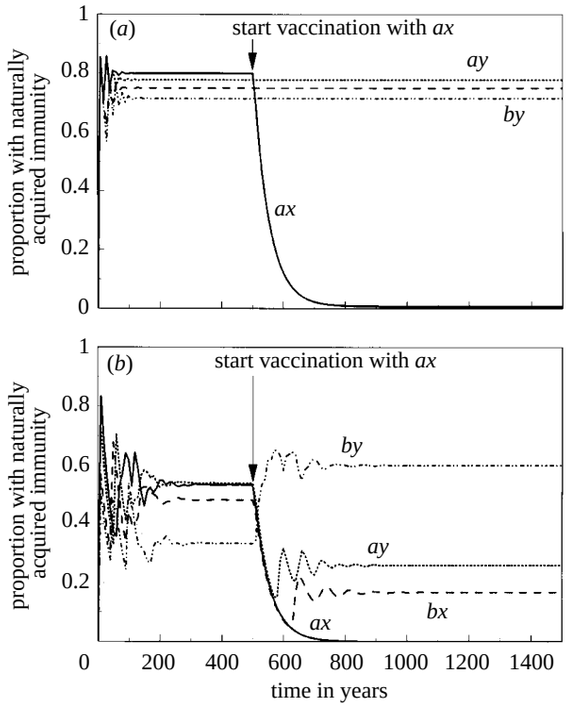

Figure 3. The effect of vaccinating against a single strain, *ax*, in the two-locus two-allele case described above, for (*a*) $\gamma = 0$ and (*b*) $\gamma = 0.75$.

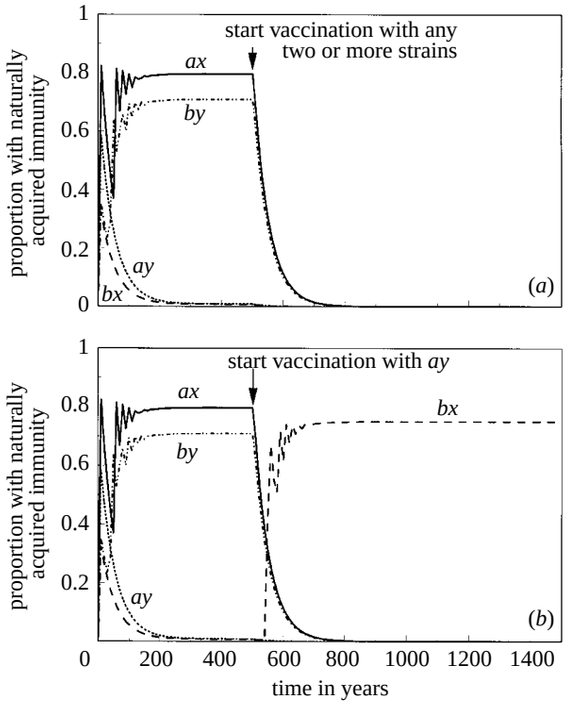

Figure 4. The effects of vaccinating (*a*) with any two or more strains, and (*b*) with a single subdominant strain, *ay*, in the case where cross-protection is complete (i.e. $\gamma = 1$).

as γ increases, vaccination against a single strain (e.g. *ax*) will have the effect of increasing competition against *ay* and *bx*, and reducing competition against the discordant strain *by*. The net effect will be that as γ increases, the relative change in frequency of the discordant strain will increase until the threshold where $\gamma = \gamma_{crit}$ is reached. Figure 3 illustrates this effect of vaccinating against a single strain, *ax*, in the two-locus two-allele case described above, contrasting (*a*) $\gamma = 0$ and (*b*) $\gamma = 0.75$.

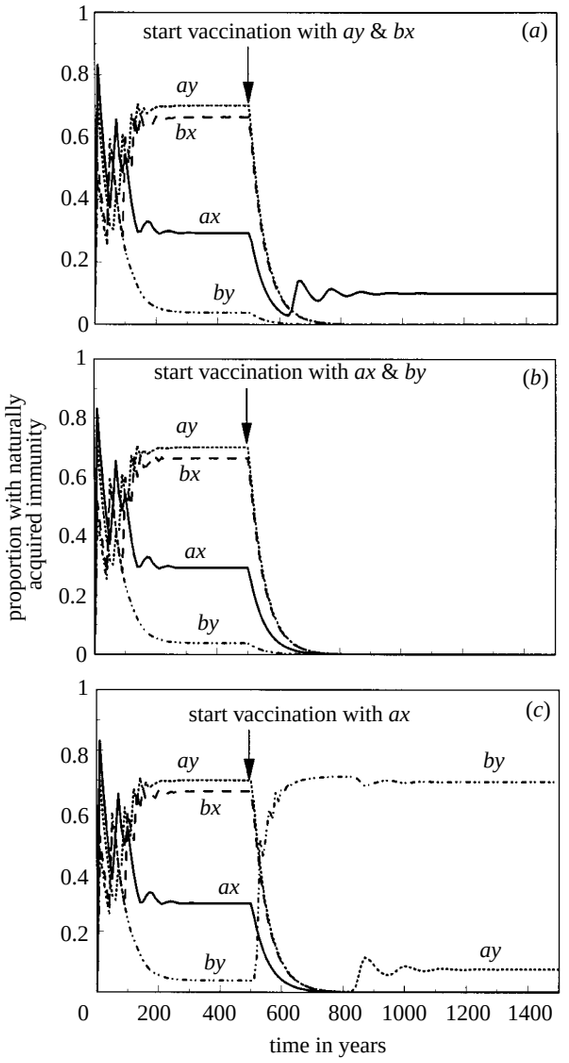

Figure 5. The effects of vaccinating (*a*) with the dominant strains, (*b*) with the subdominant strains, and (*c*) with a single subdominant strain, in the case where cross-protection is incomplete ($\gamma = 0.8$) but strong enough to yield an asymmetric equilibrium.

(*b*) *Asymmetric equilibrium:* $\gamma > \gamma_{crit}$

If the level of cross-protection is sufficient to permit one set of strains to dominate over the others, then we must separately consider the effects of vaccinating against the dominant and the subdominant types.

Figure 4 illustrates the general patterns that arise in the case where cross-protection is complete (i.e. with reference to figure 2*d*). Figure 5 demonstrates the patterns that emerge when cross-protection is incomplete in the special case where the strain with the highest

R_0 is subdominant (i.e. with reference to figure 2*c*). Numerical studies suggest that the following general conclusions apply to both cases:

1. Vaccinating against one or both of the dominant strains at a sufficient level (dictated by the magnitude of R_0) will result in the elimination of the dominant strains. However, when cross-immunity is complete, as shown in figure 4*a*, the subdominant strains (*ay* and *bx*) will also disappear because they will no longer be generated via recombination from the dominant types, and will be under equally strong immune selection from the vaccinated proportion of the host population. Under incomplete cross-immunity, certain subdominant strains may survive at low prevalences, as shown in figure 5*a*, particularly if these have higher basic reproductive rates than the dominant types.
2. Vaccinating against both subdominant types will eliminate these and also reduce the prevalence of the dominant types. In the limit of total cross immunity, as shown in figure 4*a*, *ax* and *by* will be eradicated. This is because increasing the proportion immune to *ay* and *bx* through vaccination will effectively also reduce the pool of susceptibles for *ax* and *by*. Where cross-protection is not complete, the dominant types may survive at depressed levels. However, in the particular case where a subdominant strain has a higher R_0 than the dominant strains, the latter may be eradicated even when cross-protection is incomplete, as shown in figure 5*b*.
3. Vaccinating against a single subdominant strain (e.g. *ay*) will cause the other subdominant (*bx*) to be selected for. In the case where cross-immunity is complete, as shown in figure 4*b*, *bx* will be the only surviving strain, as increasing the proportion immune to *ay* through vaccination will have fatal consequences for the dominant strains, *ax* and *by*. With partial cross-immunity, the dominant strains may survive as shown in figure 5*c*, where *ay* continues to survive at a low prevalence after vaccination with *ax*. In general, it is the dramatic increase in the prevalence of the discordant subdominant that predominantly characterizes this outcome.

5. DISCUSSION

Antigenic diversity within a parasite population can only be stably maintained under conditions where the immune responses to antigens conserved between the strains are relatively ineffective, as otherwise the repertoire would be reduced through competitive exclusion (May & Anderson 1983; Bremermann & Thieme 1989; Gupta *et al*. 1994). In preparing vaccines for use against organisms that exhibit extensive antigenic diversity, the question thus arises as to whether we should target the more immunogenic polymorphic determinants or the less immunogenic conserved determinants. While an ideal vaccine would be one that was capable of artificially boosting a conserved immune response that is only weakly protective under natural conditions, it may not be possible to create such a construct. The alternative is to arrange a cocktail of the dominant polymorphic antigens. It is unlikely, however, that such a cocktail would include the entire range of relevant epitopes.

In this paper, we have explored the epidemiological consequences of using vaccines based on polymorphic antigens that only partially cover the spectrum of antigenic types or strains circulating within a community, where the latter frequently exchange genetic material. We have shown previously (Gupta *et al*. 1996) that populations of infectious agents may be structured into independently transmitted strains by a dominant immune response against a polymorphic determinant, when there is a moderate-to-strong cross-protection between genotypes that share alleles at the loci that define a strain. When this is the case, one set of discordant strains (with non-overlapping combinations of alleles) will dominate over all other strains. If all strains appear to be more or less equally prevalent, then cross-protection between parasite types sharing alleles must be lower than the critical level required. In the case where there is no cross-protection, vaccinating against a particular strain will have no effect on the others. As cross-protection increases, the parasite types that share alleles with those included in the vaccine will be dimished in prevalence, while those that do not (i.e. the discordant strains) will increase in prevalence. The increase in prevalence of the discordant strains will become more and more exaggerated as cross-protection increases. Beyond a critical level of cross-protection, however, the discordant strains will only show a dramatic increase if they are initially subdominant, as discordant dominant strains will already be extremely frequent in the population. There are thus two circumstances under which a rare strain may increase suddenly in frequency after vaccination.

The first is where cross-protection is strong enough to cause competition between all the parasite types but not strong enough to generate an asymmetric equilibrium. Under these circumstances, vaccinating against a particular set of antigenic types will greatly reduce the selection pressure against its discordants, and cause them to increase in frequency. The second scenario pertains to the situation where cross-protection is high enough to cause the parasite population to separate into dominant and subdominant types at equilibrium (i.e. manifesting as strain structure). In this case, vaccinating against a subset of subdominant strains will give an advantage to its discordant subdominants, which will then exhibit a dramatic increase in frequency. This second case precisely underscores the importance of conducting appropriate population studies before proceeding with a vaccine that has been proven to be effective in laboratory studies and phase I trials against only a subset of the antigenic types circulating in a defined population. It is conceivable that the epitopes of subdominant combinations will have been identified *in vitro* as they would be highly immunogenic (the subdominant combinations may possibly be more immunogenic than the dominant combinations) and may be singularly effective in a vaccine preparation. As indicated in figures 4*b* and 5*c*, the use of such a vaccine may cause a precipitous increase in the prevalence of a strain that was previously only present at low levels due to immune selection.

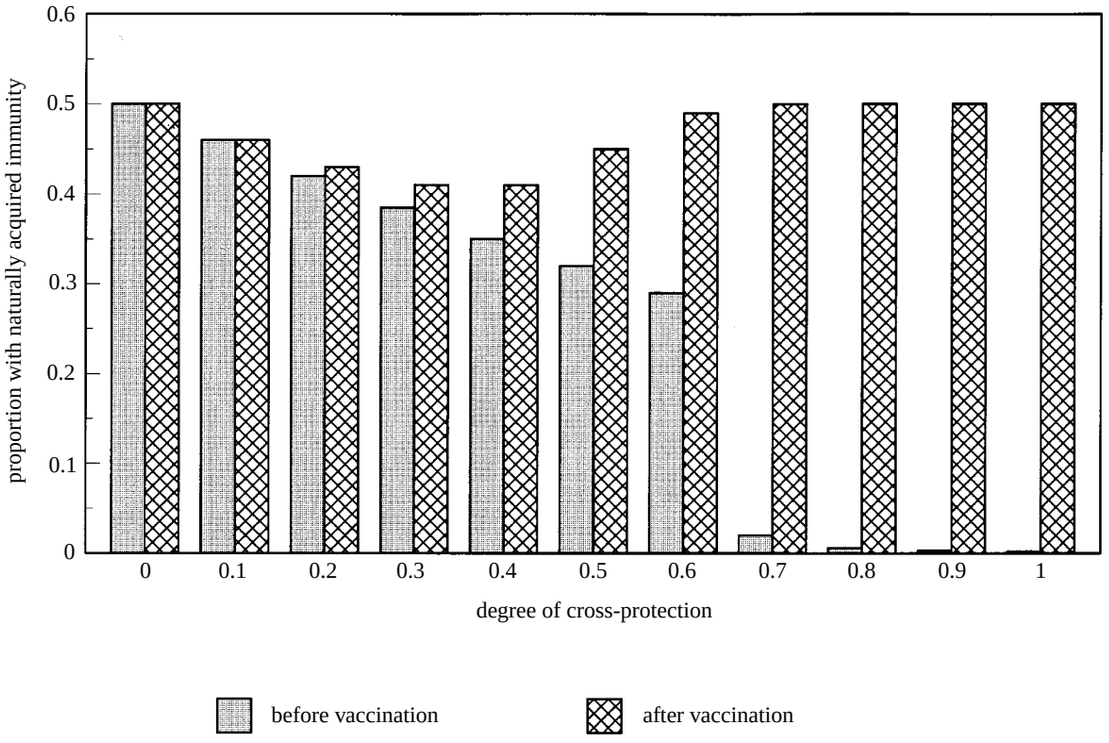

Figure 6. The expected change in the frequency of a strain as a result of vaccination with its discordant strain (in the two-locus two-allele case, where all strains have an R_0 of 2) as a function of the degree of cross-protection, γ.

Assessing the current population structure of the parasites circulating within a community is therefore essential to the design of a control programme involving partial cocktails of strain-specific vaccines. However, where cross-protection is too low to precipitate a clearly defined strain structure, these judgements cannot be based on population surveys alone, as many parasite types may be cocirculating at similar frequencies within the population. In this case, it will be important to establish the precise degree of cross-protection operating between types that share antigens or antigenic variants. Figure 6 shows the expected change in the frequency of a strain as a result of vaccination with its discordant strain (in the two-locus two-allele case, where all strains have an R_0 of 2) as a function of the degree of cross-protection, γ. When $\gamma = 0$, there is no change in frequency; as γ increases the equilibrium prevalence in the absence of vaccination declines and the relative change with vaccination increases. When γ exceeds its critical value, the pre-vaccination prevalence of a subdominant strain will decline to very low levels as shown in the figure; thus vaccination will cause a very large increase in prevalence.

Altering the community ecology of the system in this manner has the danger of greatly increasing the burden of disease should the selected types be even slightly more virulent than the types displaced by vaccination. Even if this is not the case, the rise in frequency of strains against which there are no available vaccines will have long-term consequences for control. The problem is analogous to the evolution of vaccine-resistant strains or escape mutants (McLean 1995; Lipsitch 1997), with the added complexity of genetic interaction between the resistant strains and the types included in the vaccine. On the other hand, vaccinating with a range of subdominants may be more useful than might be thought by simply considering the frequencies of these strains within a parasite population. This may have the effect of reducing the frequencies of the dominant strains considerably. In the limit of total cross-protection, the latter will be eliminated and complete eradication may be effected with sufficient vaccine coverage.

In conclusion, by integrating components of epidemiology, population genetics and immunology in a single theoretical template, this study emphasizes that the design of vaccines should incorporate a number of elements in addition to the immunogenicity and safety of combinations of antigens, namely the details of the parasite population structure and perhaps even more critically, the extent of cross-protection between parasite types. It is important to distinguish here between cross-protection as a consequence of sharing relevant antigens or antigenic variants, and cross-reactivity between different antigenic types, as the latter may be quantified by laboratory studies, but the former is less likely to be determined by experimentation. Longitudinal cohort based studies recording the sequence of transmissible infections within individuals may provide information on cross-protection. Epidemiological studies documenting the history of exposure to different serotypes can also assist in quantifying the degree of cross-protection. While such studies may be laborious and time-consuming, our analyses indicate

that they are essential to the design of a control programme involving multivalent vaccines.

We thank Brian Greenwood and Adrian Hill for their helpful comments, and the Wellcome Trust for financial support.

APPENDIX 1

We consider a general host–pathogen system, where the pathogen has M strains, and cross-protection acts to reduce the infectiousness of an individual infected by a strain, i, if that individual has already been exposed to a 'related' strain, j. More specifically, let Ω_{ij} represent the 'relatedness' matrix, where $\Omega_{ij}=1$ if strains i and j are related, and $\Omega_{ij}=0$ otherwise (we assume that $\Omega_{ij}=\Omega_{ji}$ and that $\Omega_{ii}=0$). The action of cross-protection will be modelled by assuming that an infected individual who has already been exposed to a related strain only has a probability of $1-\gamma$ of becoming infectious (though such an individual will always become immune to that strain). However, we assume all individuals in an infectious state for a particular strain will be equally infectious for that strain.

We will use a generalized notation to denote individuals with a particular immunological history. Let $q^{i_1 i_2 \ldots i_n}_{j_1 j_2 \ldots j_m}$ be the fraction of the population that has been exposed to strains $i_1,i_2,\ldots,i_n,j_1,j_2,\ldots,j_m$ (in that order) and are simultaneously infectious for strains $j_1,j_2,\ldots,j_m$. Here $n+m\leqslant M$. q is then the proportion of the population that have been exposed to no strains.

The deterministic dynamics of this system are then described by the following equation set. For $m>0$, $n>0$ and $m+n<M$, and all unique values of $i_1,i_2,\ldots,i_n,j_1,j_2,\ldots,j_m$ between 1 and M:

$$\begin{aligned}\frac{\mathrm{d}q^{i_1 i_2 \ldots i_n}_{j_1 j_2 \ldots j_m}}{\mathrm{d}t} &= (1-\gamma\Gamma^{j_m}_{i_1 i_2 \ldots i_n j_1 j_2 \ldots j_{m-1}})\, q^{i_1 i_2 \ldots i_n}_{j_1 j_2 \ldots j_{m-1}}\lambda_{j_m} \\ &\quad + \gamma\Gamma^{i_n}_{i_1 i_2 \ldots i_{n-1} j_1 j_2 \ldots j_m} q^{i_1 i_2 \ldots i_{n-1}}_{j_1 j_2 \ldots j_m}\lambda_{i_n} + \sigma q^{i_1 i_2 \ldots i_{n-1}}_{j_1 j_2 \ldots j_m i_n} \\ &\quad - \left(\sum_k \epsilon_{i_1 i_2 \ldots i_n j_1 j_2 \ldots j_m k}\lambda_k\right) q^{i_1 i_2 \ldots i_n}_{j_1 j_2 \ldots j_m} \\ &\quad - (\mu+\sigma) q^{i_1 i_2 \ldots i_n}_{j_1 j_2 \ldots j_m}. \end{aligned} \tag{6}$$

For $m=0$ and $n>0$,

$$\begin{aligned}\frac{\mathrm{d}q^{i_1 i_2 \ldots i_n}}{\mathrm{d}t} &= \gamma\Gamma^{i_n}_{i_1 i_2 \ldots i_{n-1}} q^{i_1 i_2 \ldots i_{n-1}}\lambda_{i_n} + \sigma q^{i_1 i_2 \ldots i_{n-1}}_{i_n} \\ &\quad - \left(\sum_k \epsilon_{i_1 i_2 \ldots i_n jk}\lambda_k\right) q^{i_1 i_2 \ldots i_n} - \mu q^{i_1 i_2 \ldots i_n}. \end{aligned} \tag{7}$$

For $m>0$ and $n=0$,

$$\begin{aligned}\frac{\mathrm{d}q_{j_1 j \ldots j_m}}{\mathrm{d}t} &= (1-\gamma\Gamma^{j_m}_{j_1 j_2 \ldots j_{m-1}}) q_{j_1 j \ldots j_{m-1}}\lambda_{j_m} \\ &\quad - \left(\sum_k \epsilon_{j_1 j_2 \ldots j_m k}\lambda_k\right) q_{j_1 j_2 \ldots j_m} - (\mu+\sigma) q_{j_1 j_2 \ldots j_m}. \end{aligned} \tag{8}$$

For $m=0$ and $n=0$,

$$\frac{\mathrm{d}q}{\mathrm{d}t} = \mu - q\sum_k \lambda_k - \mu q. \tag{9}$$

Here $\epsilon_{k_1 k_2 \ldots k_N}$ is 1 if all of $k_1,k_2,\ldots,k_N$ are different, and 0 otherwise (hence $\epsilon_{i_1 i_2 \ldots i_n j_1 j_2 \ldots j_m k}=0$ for all k if $m+n=M$). μ is the mortality rate of the population, and $1/\sigma$ is the mean infectious period. $\Gamma^k_{l_1 l_2 \ldots l_m}$ is defined to be 1 if any of $\{l_1,l_2,\ldots,l_m\}$ is 'related' to k, and 0 otherwise

$$\Gamma^k_{l_1 l_2 \ldots l_m} = 1 - \prod_{r=1}^{m}(1-\Omega_{l_r k}). \tag{10}$$

λ_k is the force of infection for strain k, given by

$$\lambda_k = \beta \sum_{n=0}^{M}\sum_{m=0}^{M-n}\sum_{p=0}^{M-n-m}\sum_{i_1,\ldots,i_n}\sum_{j_1,\ldots,j_m}\sum_{l_1,\ldots,l_p}\frac{q^{i_1 \ldots i_n}_{j_1 \ldots j_m k l_1 \ldots l_p}}{1+(n+p)\nu}, \tag{11}$$

where ν $(0\leqslant\nu\leqslant 1)$ determines the level of infectiousness of multiply infected hosts. If $\nu=0$, a host infectious for multiple strains is as infectious for each strain as one infectious for a single strain. If $\nu=1$, the total infectiousness of a host is independent of the number of concurrent infections.

For more than two strains, this system is cumbersome to study in the above form. If detailed information of immunological history dynamics are not required, it is possible to perform an exact reduction of the system to a much simpler form. Let z_k denote the fraction of the population exposed to strain k at some time in their immunological history. Then z_k is given by

$$z_k = \sum_{n=0}^{M}\sum_{m=0}^{M-n}\sum_{p=0}^{M-n-m}\sum_{i_1,\ldots,i_n}\sum_{j_1,\ldots,j_m}\sum_{l_1,\ldots,l_p}\left(q^{i_1 \ldots i_n k l_1 \ldots l_p}_{j_1 \ldots j_n} + q^{i_1 \ldots i_n}_{j_1 \ldots j_n k l_1 \ldots l_p}\right). \tag{12}$$

Similarly, the fraction of the population exposed to strain k or any strain related to k, w_k, is given by

$$w_k = \sum_{n=0}^{M}\sum_{m=0}^{M-n}\sum_{i_1,\ldots,i_n}\sum_{j_1,\ldots,j_m}\Gamma^k_{i_1 i_2 \ldots i_n j_1 j_2 \ldots j_m} q^{i_1 \ldots i_n}_{j_1 \ldots j_m}. \tag{13}$$

Defining $y_k=\lambda_k/\beta$ as the mean relative infectiousness of the population for strain k, one can show using equations (6), (7), (8), (9) and (11) that $\{z_k,w_k,y_k\}$ form a closed dynamical system (in the $\nu=0$ case), being described by:

$$\frac{\mathrm{d}z_k}{\mathrm{d}t} = (1-z_k)\lambda_k - \mu z_k, \tag{14}$$

$$\frac{\mathrm{d}w_k}{\mathrm{d}t} = (1-w_k)\sum_j \Omega_{jk}\lambda_j - \mu w_k, \tag{15}$$

$$\frac{\mathrm{d}y_k}{\mathrm{d}t} = (1-w_k+(1-\gamma)(w_k-z_k))\lambda_k - (\mu+\sigma)y_k. \tag{16}$$

We now consider the effects of genetic exchange between pathogens in a single host. Let α^i_{jk} be the probability that recombination between strains j and k will result in progeny of strain i. Then the force of infection for strain k is given by

$$\lambda_k = \beta \sum_{n=0}^{M}\sum_{m=1}^{M-n}\sum_{i_1,\ldots,i_n}\sum_{j_1,\ldots,j_m}\sum_{u=1}^{m}\alpha^k_{j_u j_v}\frac{q^{i_1 \ldots i_n}_{j_1 \ldots j_m}}{1+(m-1)\nu}. \tag{17}$$

Note that if $\alpha_{ij}^{k} = 1$ if $i = j = k$, and 0 otherwise (i.e. asexual reproduction), this reduces to equation (11).

Although it is possible to specify reduced state-variable sets that accurately track, say, two concurrent infections, this is highly computationally intensive. We therefore make the approximation that the proportion of the population infected with any two strains is given by the product of the proportions singly infected with each strain, multiplied by a factor $1 - \gamma$ if the two strains are related. The triply infectious are estimated similarly, with factors of $1 - \gamma$ for every unique pair of strains which share alleles. Note that as the proportion of the population concurrently infected with m strains can be shown to be $O((\mu/\sigma)^m)$, the overall effect of concurrent infections is negligible for $\sigma \gg \mu$ (i.e. for a disease where the infectious period is much less than the host's lifespan). This results in relative insensitivity in model dynamics to the particular approximation adopted to represent recombination.

REFERENCES

Ada, G. L. 1995 The development of new vaccines. In *Vaccination and world health* (ed. F. T. Cutts & P. G. Smith). Chichester: John Wiley & Sons.

Anderson, R. M. & May, R. M. 1991 *Infectious diseases of humans: dynamics and control*. Oxford University Press, Oxford.

Bremermann, H. & Thieme, H. R. 1989 A competitive-exclusion principle for pathogen virulence. *J. Math. Biol.* **27**, 179–190.

Gupta, S., Swinton, J. & Anderson, R. M. 1994 Theoretical studies of the effects of genetic heterogeneity in the parasite population on the transmission dynamics of malaria. *Proc. R. Soc. Lond* B **256**, 231–238.

Gupta, S., Maiden, M., Feavers, I., Nee, S., May, R. M. & Anderson, R. M. 1996 The maintenance of strain structure in populations of recombining infectious agents. *Nature Med.* **2**, 437–442.

Lipsitch, M. 1997 Vaccination against colonizing bacteria with multiple serotypes. *Proc. Natn. Acad. Sci. USA* **94**, 6571–6576.

Maiden, M. C. J., Suker, J., McKenna, A. J., Bygraves, J. A. & Feavers, I. M. 1991 Comparison of the class 1 outer membrane proteins of eight serological reference strains of *Neisseria meningitidis*. *Mol. Microbiol.* **5**, 727–736.

May, R. M. & Anderson, R. M. 1983 Epidemiology and genetics in the coevolution of parasites and hosts. *Proc. R. Soc. Lond.* B **219**, 281–313.

McLean, A. R. 1995 Vaccination, evolution and changes in the efficacy of vaccines: a theoretical framework *Proc. R. Soc. Lond.* B **261**, 389–93.

van der Ley, P. & Poolman, J. T. 1992 Construction of a multivalent meningococcal vaccine strain based on the class 1 outer membrane protein. *Infect. Immun.* **60**, 3156–3161.

Watson, D. A., Musher, D. M. & Verhoef, J. 1995 Pneumococcal virulence factors and host immune responses to them. *Eur. J. Clin. Microbiol. Infect. Dis.* **14**, 479–490.

Wiertz, E. J., van Gaans van den Brink, J. A., Schreuder, G. M. T. H. *et al.* 1991 T cell recognition of Neisseria meningitidis class 1 outer membrane proteins. Identification of T cell epitopes with selected synthetic peptides and determination of HLA restriction elements. *J. Immunol.* **147**, 2012–2018.

Received 22 April 1997; accepted 10 June 1997

As this paper exceeds the maximum length normally considered for publication in *Proceedings* B, the authors have agreed to make a contribution towards production costs.