Cluster formation for multi-strain infections with cross-immunity

Vincent Calvez[1], Andrei Korobeinikov[2], Philip Maini[2]

[1] Ecole Normale Supérieure, 45 rue d'Ulm, Paris 75005, France

vincent.calvez@ens.fr

[2] Centre for Mathematical Biology, Mathematical Institute,

University of Oxford, 24-29 St Giles', Oxford, OX1 3LB, UK

korobein@maths.ox.ac.uk, maini@maths.ox.ac.uk

July 2, 2004

Abstract

Many infectious diseases exist in several pathogenic variants, or strains, which interact via cross-immunity. It is observed that strains tend to self-organise into groups, or clusters. The aim of this paper is to investigate cluster formation. Computations demonstrate that clustering is independent of the model used, and is an intrinsic feature of the strain system itself. We observe that an ordered strain system, if it is sufficiently complex, admits several cluster structures of different types. Appearance of a particular cluster structure depends on levels of cross-immunity and, in some cases, on initial conditions. Clusters, once formed, are stable, and behave remarkably regularly (in contrast to the generally chaotic behaviour of the strains themselves). In general, clustering is a type of self-organization having many features in common with pattern formation.

1 Introduction

Many pathogens have several different antigenic variants, or strains, present in a host population simultaneously. The classic example is influenza [1, 12, 13, 7, 4], where there are several circulating subtypes, with many minor variants within each subtype. Other important examples are meningitis [9, 11], dengue [7] and malaria [8].

Because of similarities in, for example, their mechanisms of infection, strains may interact with each other [9]. Infection with one strain may partially protect

the host against infection with other strains. Cross-immunity is included in different ways in different models, but the general idea is the same: infection with one strain of the disease produces a lasting immune memory in the host which acts to protect against subsequent infection by other strains. That is, for two sufficiently close strains A and B, infection by strain A reduces the chance of a secondary infection by strain B. For instance, in the case of influenza, the surface protein hemagglutinin seems to be under strong positive selection because it is the target of the immune response, and therefore it presents high antigenic diversity in the virus population [1, 12, 13, 7]. This immune response may be enhanced because of a previous infection with a close variant.

There are different approaches to the cross-immunity problem [3, 6]. For instance, we can assume that a fraction, say γ_{BA}, of individuals infected with strain A gain complete immunity to strain B; alternatively, all the individuals infected with strain A may be assumed to acquire partial immunity against B (with a consequence that the force of secondary B-infection is reduced by a factor γ_{BA}). Another possible hypothesis is that the secondary infection is weaker and thus less transmissible by the infective host. These differences in the approaches to cross-immunity lead to variety of models providing sometimes controversial results. Under such circumstances it is reasonable to look for such features of the multi-strain system which are intrinsic to this system and are robust respectively a model choice.

A system of multiple strains interacting via host cross-immunity tends to self-organise into groups, or clusters. The tendency for strains to occur in clusters reflects the observed influenza dynamics [7, 13]. Cluster formation was observed and discussed by [9, 10]. It may be noteworthy that the phenomenon of clustering appears to be typical for many systems with internal order and may occur in such systems as multi-species predator-prey systems. For example, it was observed in neuronal networks [14, 15, 16, 17].

In this paper we consider formation of clusters in ordered multi-strain systems. We show that for complex systems several different types of cluster structure may arise. We also demonstrate that cluster structures are not specific to a particular model — on the contrary, they appear to be intrinsic to the given strain system. In general, cluster formation is a self-organisation phenomenon bearing many similarities to pattern formation. A remarkable feature of clusters is that they exhibit exceptional regularity even when dynamics of every strain is chaotic.

2 Model

Due to different approaches to cross-immunity, a variety of models of multi-strain infections has been developed. These models sometimes lead to different outcomes. It is important, therefore, to find such indicators which are characteristic to the system itself and robust respectively choice of a model.

We start from a comparatively simple model of a multi-strain infection suggested by [10]. The model is composed of only three compartments (and respectively three differential equations) for each strain. If $z_i(t)$ is the fraction of individuals who have been or are infected with the strain i (either they are infectious or not), $y_i(t)$ is the fraction of the infectious with the strain, and $w_i(t)$ is the fraction of individuals who have been infected (or are infected) by any strain sufficiently close to the strain i including i itself (that is $w_i = \cup_{j\sim i} z_j$), then the model equations are

$$\begin{aligned} \frac{d\,z_i}{dt} &= \beta_i y_i(1-z_i) - \mu z_i, \\ \frac{d\,w_i}{dt} &= \sum_{j\sim i} \beta_j y_j (1-w_i) - \mu w_i, \\ \frac{d\,y_i}{dt} &= \beta_i y_i \left[(1-w_i) + (1-\gamma)(w_i - z_i)\right] - (\mu+\sigma_i) y_i. \end{aligned} \tag{1}$$

For this model, cross-protection does not affect susceptibility but reduces transmissibility by a factor $1-\gamma$ (where the parameter γ measures the degree of cross-protection between two strains). Here $j \sim i$ means that the jth strain is related to the ith strain and can induce cross-protection (that is if $j \sim i$ then $\gamma_{ij} \neq 0$). The parameters $1/\mu$ and $1/\sigma$ are, respectively, host life expectancy and average period of infectiousness, β is transmission rate. We refer to this model as Gupta's model. This simple model has been analysed in [10] and provided important insights into pathogenes formation and strains genetic organisation.

To study the phenomenon of clustering we need to consider several levels of cross-protection. Whereas the original model implies only one level of cross-protection (γ if two strains are related, or zero if they are not), and neglects possible multiple infections by strains related to i. We relax these assumptions below to make the model more generally applicable, while striving to keep the model simple. We assume that the probability of cross-protection between strains i and j is γ_{ij} (that is, infection by the strain j reduces the probability that the host will be infected by the strain i by a factor γ_{ij}), and consider the barycenter of γ_{ij}, defined as

$$\Gamma_i = \left(\sum_{j\sim i, j\neq i} \gamma_{ij}\beta_j y_j \right) \Big/ \left(\sum_{j\sim i, j\neq i} \beta_j y_j \right). \tag{2}$$

We replace the coefficient γ in the system (1) with the barycenter Γ_i.

Substituting the barycenter Γ_i into (1) and using variables $V_i = 1 - z_i$, $X_i =$

$1-w_i$, $Y_i = \frac{\beta_i}{\mu} y_i$ and $\tau = \mu t$, we obtain the system

$$\begin{aligned} \frac{dV_i}{d\tau} &= 1 - \Big(1 + Y_i\Big)V_i, \\ \frac{dX_i}{d\tau} &= 1 - \Big(1 + \sum_{j\sim i} Y_j\Big)X_i, \\ \varepsilon_i \frac{dY_i}{d\tau} &= \Big((1-\Gamma_i)V_i + \Gamma_i X_i - r_i\Big)Y_i. \end{aligned} \tag{3}$$

Here $\varepsilon_i = \mu/\beta_i$ and $r_i = (\mu + \sigma_i)/\beta_i$..

Obviously $\Gamma_i \equiv \gamma$ for Gupta's model (when γ_{ij} is either γ, or zero). Furthermore, computations show that for this model the function $\Gamma_i(t)$ most of the time remains taking one of a few constant values, with rapid shifting between these values (see Fig. 4). This justifies the use of the function $\Gamma_i(t)$.

3 Structure of a strain set

Systems of strains were formed as a result of a genetic process, and they generally inherited some internal order associated with this process. Having this intrinsic order, strains may be organised in an ordered set, or a discrete strain space every point of which represents a strain. The idea of the strain space allows us to use the concept of "immunological distance". The immunological distance between two strains may be assumed to depend inversely on their mutual level of cross-protection.

The structure of the strain space depends on underlying immunological and genetic processes. For instance, [7] considered the simplest possible strain space: a linear strain space. In this case strains are arranged in a line, and they postulated $\gamma_{ij} = \exp(-(\frac{i-j}{d})^2)$, where d is a constant. A multi-dimensional strain space may be organised in the same way, with immunological distance defined, for example, as the sum of horizontal and vertical distances. [5, 6] considered a system of four strains arranged in a circle. In this case each strain is assumed to interact more strongly with its adjacent neighbours than with the strain opposite.

Studying the maintenance of strain structure in a recombining virus population, [9] have introduced a simple framework where strains are organised as follows: each strain is characterised by a combination of alleles at loci which are of immunological interest. Strains induce cross-immunity if they share at least one allele. For example, in the case of two loci and two possible alleles at each locus (say a or b for the first locus, and x or y for the second one respectively) there are four different strains: the original strains ax and by, and the recombinant strains ay and bx. To visualise such a strain structure we will use a multi-dimensional graph where a dimension corresponds to a locus, and vertices represent strains. Figure 1 illustrates the structure of the above mentioned four-strain system (two

loci and two possible alleles at each locus). Figure 2 shows the strain space of an eight-strain system organised on three loci with two alleles at each locus.

4 Results

Cross-immunity may structure a set of strains into groups, or clusters. These groups can behave at least in three ways: remain in homogeneous equilibrium when no structure is observed (Fig. 3 a), oscillate when the clusters alternate recurrently in succession (Fig. 3 b), or one group may dominate with the others driven below survival level (Fig. 3 c) [9, 10]. The phenomenon of clustering is conserved for all sufficiently large levels of cross-protection. Of course, when $\gamma \to 0$, the equations are decoupled, and the clustering disappears.

In the case of the four-strain system shown in Fig. 1 it is natural to expect the formation of two clusters of non-overlapping (or discordant) strains, namely ax groups with by, and ay groups with bx (in Fig. 1 we respectively mark the strains by squares and cycles). Indeed, such clustering has been observed [6, 5, 9, 10]. Figure 3 illustrates the strain dynamics: it is easy to see the formation of two clusters.

However, a multi-strain system with only one level of cross-protection, which is the same for all related strains, is hardly realistic. As the number of strains grows, and especially if there are several different levels of cross-protection, the self-organisation of the system may be more complicated. Furthermore, it may be different for different levels of cross-protection. For instance, for the eight-strain system shown in Fig. 2 at least two different types of clustering are possible. From now on we will use the terms *cluster structure* and *type of cluster structure*. The difference between these objects is that different cluster structures may be of the same type. Below we will show this using an example.

For a system of eight strains organised in three loci with two alleles each (Fig. 2) we assume two levels of cross-protection: namely γ_1 if the strains share one allele, or γ_2 if they share two alleles. Naturally, $\gamma_1 \leq \gamma_2$. For this system one can expect formation of a structure of four clusters with two discordant strains each [10]. Every cluster of such structure corresponds to one of the four main diagonals of the cube in Fig. 2. However, this type of cluster structure was observed only when γ_1 and γ_2 are sufficiently close. As the difference between γ_1 and γ_2 grows, a new type of cluster structure appears: now there are two clusters, α and β, with four strains each (α is composed of the strains (1,1,1), (1,2,2), (2,1,2) and (2,2,1), and β of (2,2,2), (2,1,1), (1,2,1) and (1,1,2)). In Fig. 2 the strains of these tetraedric clusters are marked respectively by cycles and squares. This second type of clustering can hardly be expected *a priori*. However, this cluster structure exists for a much wider range of γ_1 and γ_2 than the first type. Fig. 4 illustrates the dynamics of the second type of clustering for the eight-strain system: the forces of infection are shown for two values of ε.

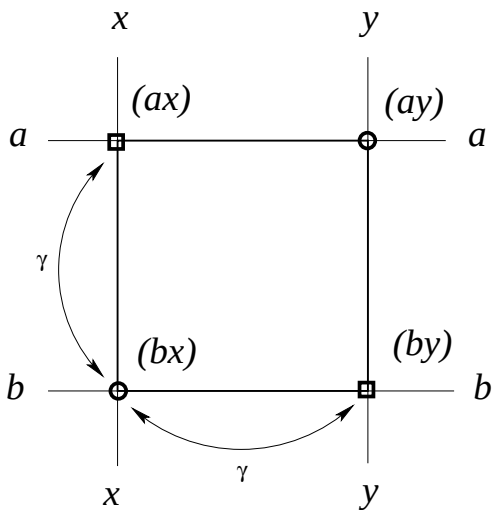


Figure 1: Strain space of a four-strain system: two loci and two possible alleles at each locus (see text for details).

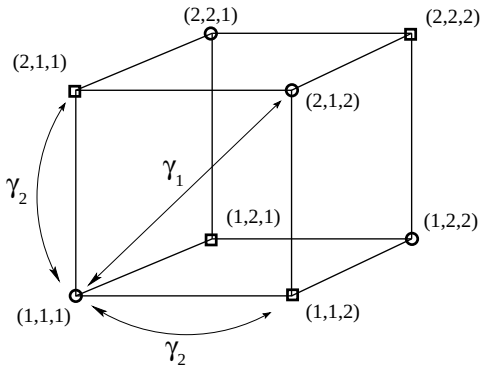


Figure 2: Strain space of an eight-strain system: three loci with two alleles possible at each locus. Here, for instance, $(2, 2, 1)$ means that the second, the second and the first alleles are respectively at the first, the second and the third locus.

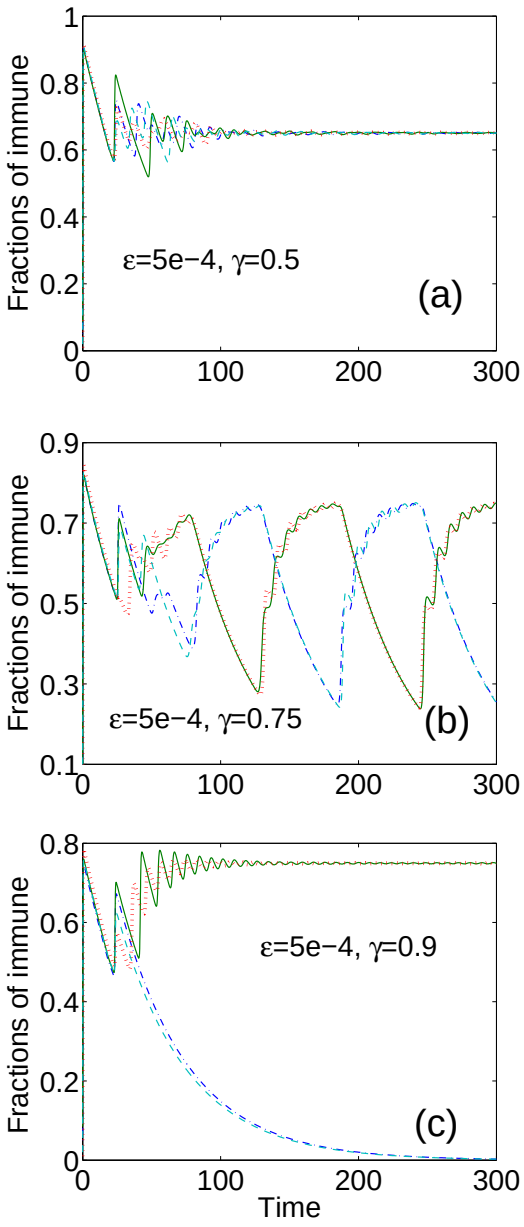


Figure 3: Dynamics of the four-strain system shown in Fig. 1. (b) and (c) illustrate formation of two clusters each consisting of two strains. In (a) the system is in homogeneous equilibrium, and no definite clustering can be observed.

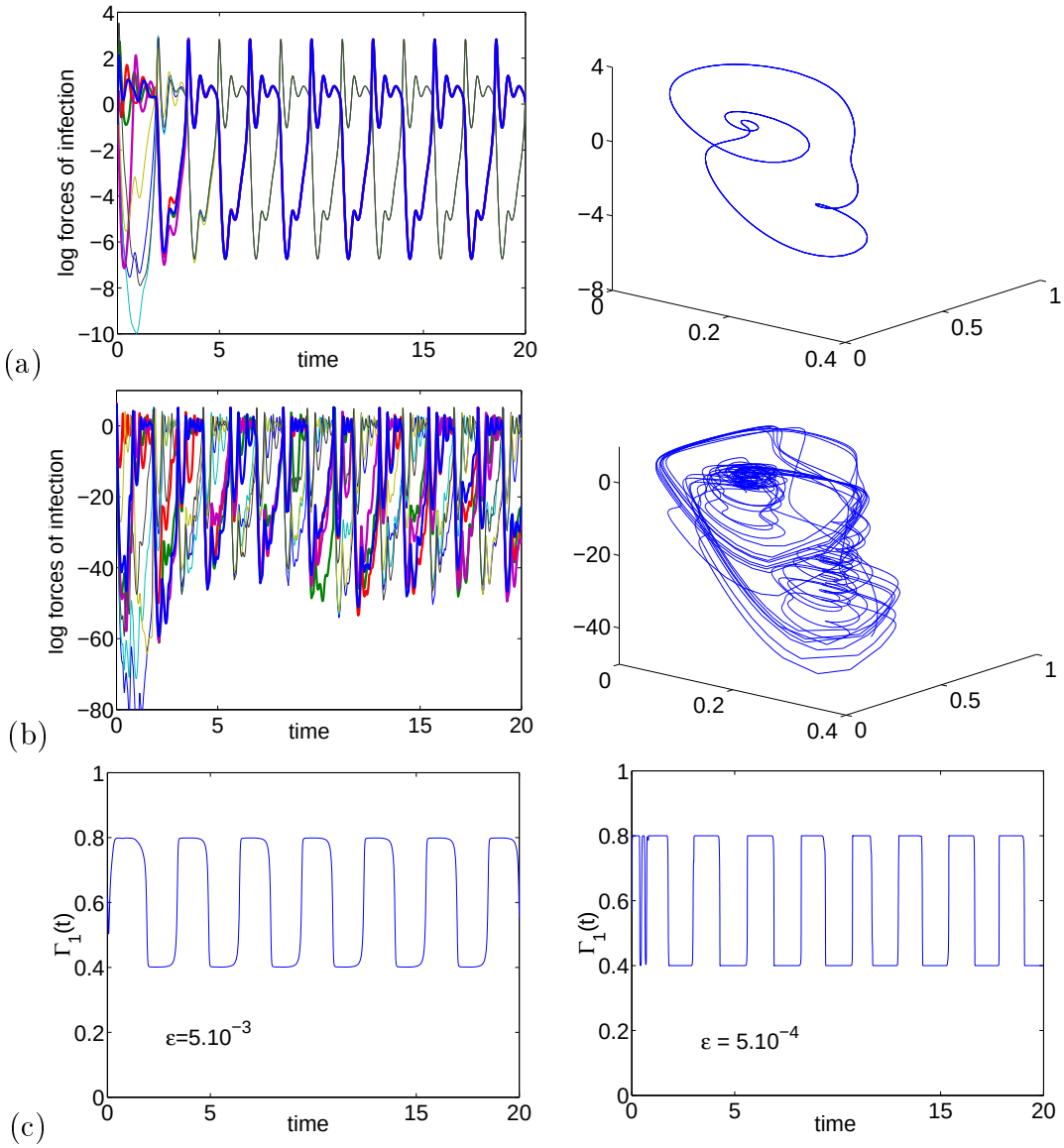
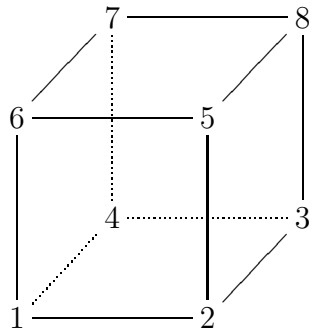


Figure 4: (*left*) Dynamics of forces of infection for $\gamma_1 = 0.4$, $\gamma_2 = 0.8$ and either $\varepsilon = 5.10^{-3}$ (a) or $\varepsilon = 5.10^{-4}$ (b). We have choosen one tetraedric cluster, and have bolded all strains inside. (*right*) Trajectories of the first strain in the $(v_i, x_i, \ln y_i)$ phase plane. Note that, specially for small ε, while the dynamics of each single strain is chaotic, the cluster as a whole behaves remarkably regularly. For a better understanding, we prefer to show forces of infection in a log-scale. In this example, the coefficients of correlation $R_{i,j}$ are 1 if strains are in the same cluster and -0.6 otherwise (a); are between 0.75 and 1 if strains are in the same cluster and between -0.45 and -0.19 otherwise (b). In (c) we show the new function $\Gamma_1(t)$ which shifts very quickly between the two different levels of cross-protection.

For a multi-strain system, the dynamics of a single strain is sometimes chaotic[1] [10]. However, under the same conditions which cause chaotic strain dynamics, clusters usually behave in a surprisingly regular fashion. This regularity is hardly to be expected *a priori*, and it motivates the study of clusters.

To qualitatively describe the cluster formation, we introduce the *clustering matrix* M. We set m_{ij} equal to 1 if strains i and j belong to the same cluster, and $m_{ij} = 0$ otherwise (the matrix is symmetric). For instance, if the vertices of the cube Fig. 2 are ordered as following



then the correspondant matrices for the first and second types of clustering are

$$M_I = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad M_{II} = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

respectively. To quantitatively estimate clustering for an ordered multi-strains system, we are to calculate for each couple of strains (i, j) the correlation coefficient [2]

$$R_{i,j} = \frac{\left\langle y_i(t) y_j(t) \right\rangle_T - \left\langle y_i(t) \right\rangle_T \left\langle y_j(t) \right\rangle_T}{\sqrt{\left(\left\langle y_i(t)^2 \right\rangle_T - \left\langle y_i(t) \right\rangle_T^2 \right) \left(\left\langle y_j(t)^2 \right\rangle_T - \left\langle y_j(t) \right\rangle_T^2 \right)}}.$$

Here $\left\langle y(t) \right\rangle_T$ is the mean average

$$\left\langle y(t) \right\rangle_T = \frac{1}{T} \int_{\mathcal{I}} y(\tau) d\tau.$$

[1]Here and through this paper, by the term "chaos" we imply deterministic chaos.

The time interval $\mathcal{I}$ should be sufficiently long and exclude the transient regime. The coefficient $R_{i,j}$ is a measure of synchronisation of the time series for the forces of infection; $R_{i,j} = 1$ when complete synchronization occurs. It is thereby a straightforward procedure to relate the coefficients $R_{i,j}$ to the coefficients of clustering m_{ij}, for instance applying a threshold function[2] $\mathcal{H}$ to the matrix R.

An alternative way is to calculate a similarity matrix with coefficients

$$S^2_{i,j} = \frac{\left\langle \big(y_i(t) - y_j(t)\big)^2 \right\rangle_T}{\sqrt{\left\langle y_i(t)^2 \right\rangle_T \left\langle y_j(t)^2 \right\rangle_T}}.$$

To systematically estimate clustering, we use the comparison between the effective correlation matrice $\mathcal{H}(R)$ and the reference clustering matrices we have defined, namely M_I and M_{II}. We have plotted in Fig. 5 the euclidian distance between correlation matrices and the second type reference matrix: $\big\|\mathcal{H}(R) - M_{II}\big\|_2$, for different values of γ_1 and γ_2. The shift between the two types of clustering occurs when $\gamma_1 \approx \gamma_2$.

Breaking the "natural" constraint $\gamma_1 \leq \gamma_2$; or even introducing a new coefficient γ_0 between opposite strains may lead to new types of clustering. For instance, the extremal case $\gamma_2 \ll \gamma_1$ generates four clusters of two neighboured strains, which are arranged in the same direction. Contrary to previous cases, this organization is not unique, but contain three possible structures, which are in the same orbit under the action of cube's rotations (this is the reason why we distinguish between the *type of cluster structure* and the *cluster structure*):

$$\begin{pmatrix} 1&1&0&0&0&0&0&0\\ 1&1&0&0&0&0&0&0\\ 0&0&1&1&0&0&0&0\\ 0&0&1&1&0&0&0&0\\ 0&0&0&0&1&1&0&0\\ 0&0&0&0&1&1&0&0\\ 0&0&0&0&0&0&1&1\\ 0&0&0&0&0&0&1&1 \end{pmatrix}, \begin{pmatrix} 1&0&0&1&0&0&0&0\\ 0&1&1&0&0&0&0&0\\ 0&1&1&0&0&0&0&0\\ 1&0&0&1&0&0&0&0\\ 0&0&0&0&1&0&0&1\\ 0&0&0&0&0&1&1&0\\ 0&0&0&0&0&1&1&0\\ 0&0&0&0&1&0&0&1 \end{pmatrix} \text{ and } \begin{pmatrix} 1&0&0&0&0&1&0&0\\ 0&1&0&0&1&0&0&0\\ 0&0&1&0&0&0&0&1\\ 0&0&0&1&0&0&1&0\\ 0&1&0&0&1&0&0&0\\ 1&0&0&0&0&1&0&0\\ 0&0&0&1&0&0&1&0\\ 0&0&1&0&0&0&0&1 \end{pmatrix}.$$

We have to stress that the phenomenon of clustering is robust. We have modified in time the rates of infection β_i of (3) by gaussian-distributed perturbations of different amplitudes. Results (see Fig. 6) show resistance to perturbation. Robustness can also be noticed in a different situation: simulations show that decreasing ε increases chaotic dynamics. However, clustering is still distinguishable (see Fig. 4).

The tendency of the strains to self-organise into clusters, and the remarkable regularity of the dynamics of these clusters, contrasting to the chaotic behaviour

[2]Within simulations, I have used a function of the form $\mathcal{H}(r) = \dfrac{\exp(\alpha(r-a))}{1+\exp(\alpha(r-a))}$.

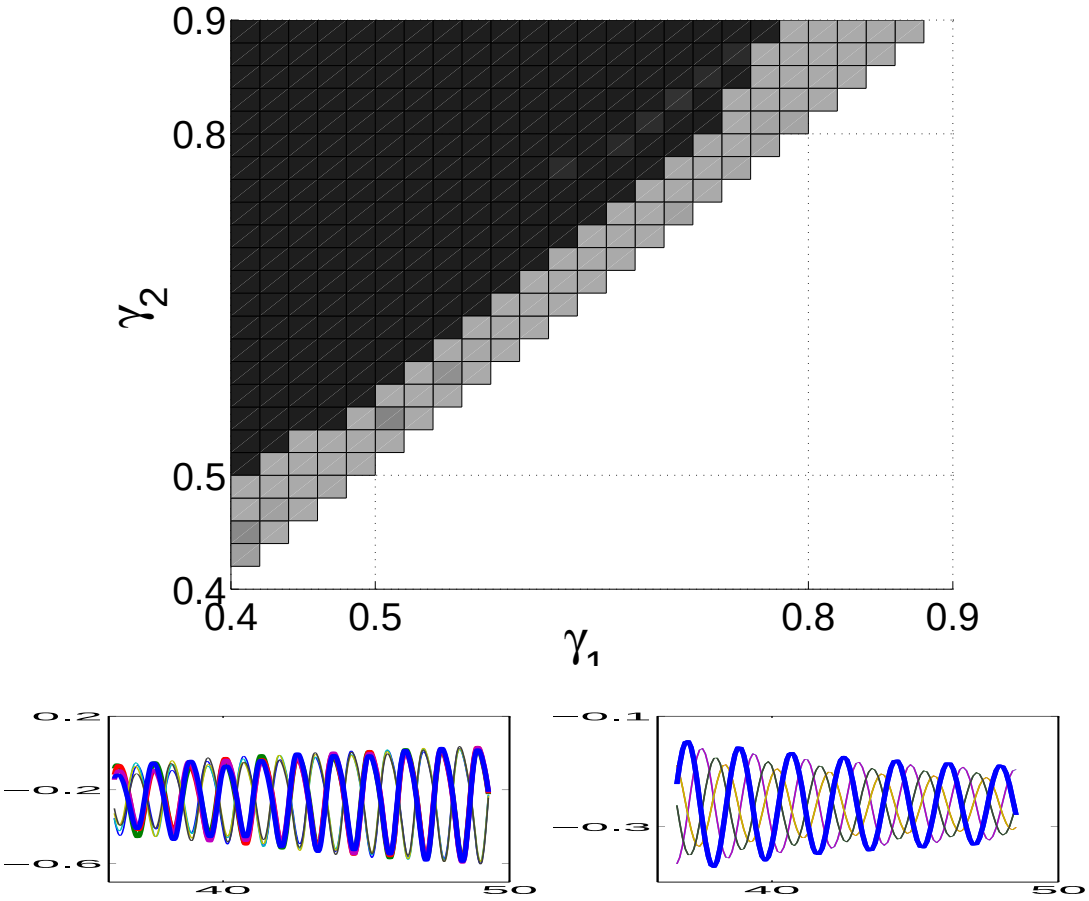


Figure 5: (*top*) Distance of the matrix of correlation to the second reference matrix M_{II} as a function of γ_1 and γ_2. Black area corresponds to 0, that is matrices coincide, and grey area corresponds to a renormalized distance equal to 1: this is where first type of clustering occurs. We have no data about the white area in this diagram because it concerns the case $\gamma_1 > \gamma_2$. In fact it corresponds to a third type of clustering, mentionned in the text. (*bottom*) The transition between the two types of clustering is very fast: the two figures shows this transition at very low amplitudes for $\gamma_2 = 0.8$: (*left*) $\gamma_1 = 0.72$; (*right*) $\gamma_2 = 0.74$.

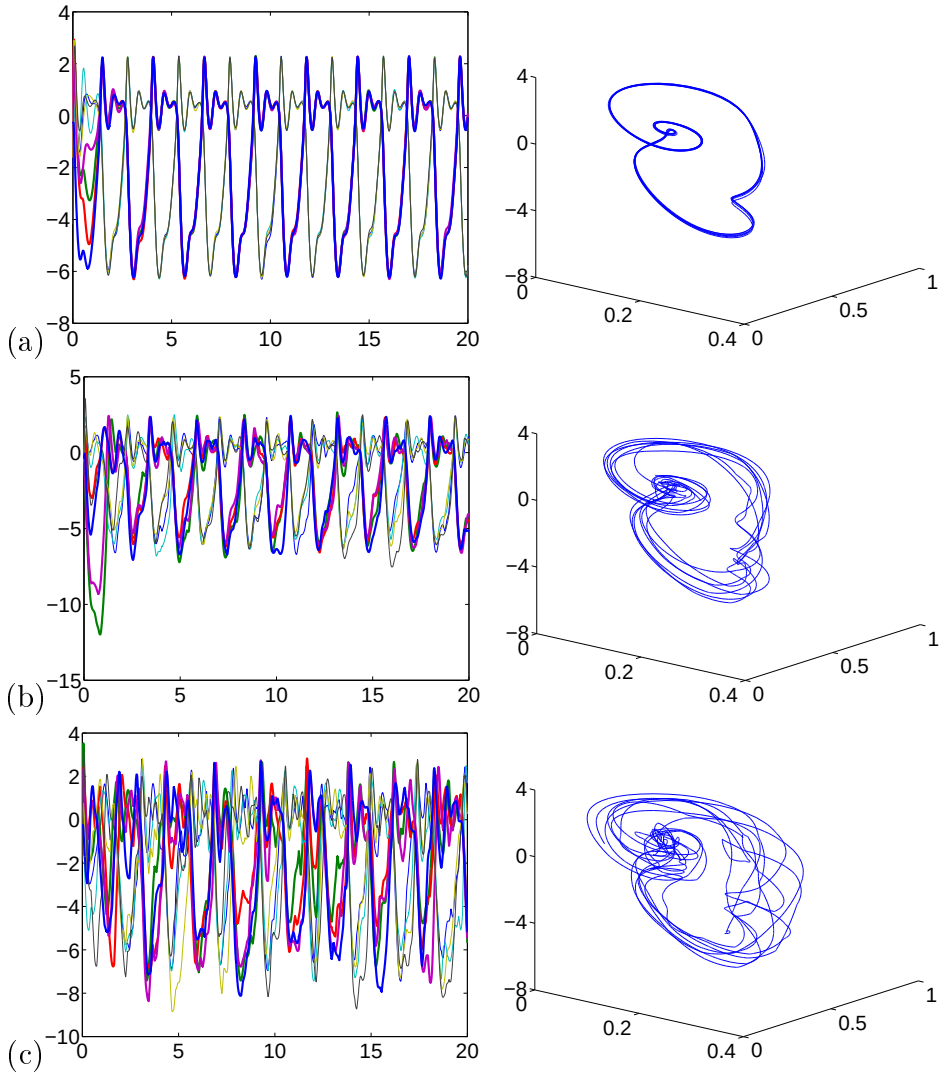


Figure 6: Perturbation of original model through the rates of infection with increasing amplitudes. All parameters are $1/\varepsilon = 200 + \Delta(t)$, $r = 0.25$, $\gamma_1 = 0.5$, $\gamma_2 = 0.8$; with Δ of amplitude 1 (a), 10 (b) and 20 (c). Again all strains of one tetraedric cluster are bolded.

of a single strain, remains as the number of strains grows. With an increasing number of strains, the number of cluster structures possible for the system grows as well, and new types of cluster structures appear. The sixteen-strain system, such that for each strain there are four loci with two alleles possible at each locus, may be visualised as a four-dimensional cube. We assume three levels of cross-protection for this system: γ_1 if the strains share one allele, γ_2 if the strains share two alleles and γ_3 for the strains sharing three alleles (naturally, $\gamma_1 \leq \gamma_2 \leq \gamma_3$). At least six cluster structures of three different types are possible for this system. Particularly, if γ_3, γ_2 and γ_1 are approximately equal, a structure of eight clusters with two discordant strains each appears (each cluster corresponds to a main diagonal of the four-dimensional cube; strains of a cluster are the ends of the diagonal). If γ_3 is sufficiently large compared with γ_1 and γ_2, then the system self-organises into two clusters of eight strains each. In this case the strains of a cluster share either no allele at all, or two alleles; there is no cluster with strains sharing one allele in this case. If both γ_3 and γ_2 are large compared with γ_1, then a new stable type of cluster structure appears. In this case eight clusters with two strains each form. A structure of this type differs from the above mentioned structure of the first type (eight clusters with two discordant strains each) as follows: in this case the strains of each cluster share one allele which is at the same locus for every cluster of the structure. That is for this cluster structure, the strains belong to the diagonal of the three-dimensional sides of the 4-D cube; whereas for the cluster structure of the first type the strains are those on the main diagonal of the four-dimensional cube. Since there are four loci for this system, four different structures of this type are possible.

We are also able to generate four clusters of four strains by choosing γ_2 larger than γ_1 and γ_3.

Self-organization of strains into clusters is not a particular feature of the model considered. Computations show that, for an ordered strain system given, the same cluster structures arise for other models, even if the dynamics of these clusters differ. It appears that a cluster structure is intrinsic to an ordered strain system.

For comparison purposes, we considered the models suggested by [7], and [6]. The former is a comparatively simple SIR model purposed to investigate the role of cross-immunity in antigenic drift with a large number of strains. The model equations are

$$\begin{aligned} \frac{d\,S_i}{dt} &= \mu - S_i \sum_{j=1}^{n} \gamma_{ij}\beta_j I_j - \mu S_i, \\ \frac{d\,I_i}{dt} &= \beta_i I_i S_i - r_i I_i, \end{aligned} \tag{4}$$

where S_i and I_i are respectively the fractions of susceptibles and infectives for the ith strain, and γ_{ii} is postulated to be equal to one.

For a n strains system, [6] model is composed of $n + 2^n$ classes. The immune hosts are indexed by the subsets of $\{1, \ldots, n\}$: for $J \subset \{1, \ldots, n\}$, Z_J denotes the individuals who are immune to exactly all strains included in J (thus they are susceptibles to ${}^cJ$). The system's equations are :

$$\begin{aligned} \dot{I}_i &= \beta_i I_i \sum_{J:i\notin J} Z_J - (\mu + \sigma_i) I_i \\ \dot{Z}_J &= \sum_{i,K} C(K, J, i)\beta_i I_i Z_K - \sum_{i\notin J} \beta_i I_i Z_J - \mu Z_J + \mu\delta_{J,\emptyset} \end{aligned} \tag{5}$$

The term $C(K, J, i)$ represent the effect of cross-immunity. In fact, it is the rate of transfert from compartment K to compartment J after infection by strain i.

Despite the huge difference in model complexity, both models demonstrate similarities in cluster formation. The behaviour of both these systems is somewhat simpler than that of Gupta's model. Particularly, no alternation of clusters was observed for these models: depending on the system parameters, the phase trajectories of the system converge towards one of the system equilibria with damped oscillations. Nevertheless, the same cluster structures were formed for both of these models. These cluster structures coincide with those for the modified Gupta's model (3), and the values of the cross-protection parameters at which the system shifts from one type of structure to another vary insignificantly from one model to the other. For instance, for the eight-strain system Fig. 2, the type of cluster structure formed depends on the comparative values of γ_{ij}. As in the case of the system (3), a structure of four clusters with two discordant strains each appears when γ_1 and γ_2 are comparatively close and, as the difference between γ_2 and γ_1 grows, a shift to the structure of the second type (two clusters with four strains each) occurs. However, in contrast to the model (3), no regular oscillation of the clusters was observed: for both types of cluster structures solutions of the models tend to an equilibrium state.

5 Conclusion

Strains of a multi-strain infection tend to self-organise into groups, or clusters. For a complex strain system several different types of cluster structures are possible and may arise. Which cluster structure occurs in reality depends mostly on levels of cross-protection and, in some cases, on initial conditions. It is important to note the distinction between the terms "cluster structure" and "type of cluster structure", as several structures of the same type are possible for complex strain systems. Cluster structures which are possible for a strain system do not depend on the particular model used. In fact, the structures are fairly robust to different models. It appears that cluster structures of a particular strain system depend on the structure of the strain space and on levels of cross-protection.

It is not clear why some cluster structures arise while others do not. While clustering in the four-strain system is transparent enough, it is already not so clear why in the eight-strain system the cluster structure of second type (two clusters of four strains) appears. More complex systems, such as the sixteen-strain system, raise even more questions. For instance, it is not clear why no structure of four clusters with four strains each is possible (at least it was not observed) for such a system.

It is a challenge to provide an exhaustive list of type of clustering which may occur for a given set of strain.

Clustering does not occur only in cross-immunity systems. In fact, it seems to be a more general interaction in coupled dynamical systems. We will present there two simple dynamical systems which, at once they are suitably coupled, present self-organisation in clusters (see Fig. 7 and also [2]).

The Van der Pol oscillator.

$$\frac{d^2x_i}{dt^2} - \varepsilon\big(1 - x_i^2\big)\frac{dx_i}{dt} + \omega^2 x_i + \sum_{j\sim i} \gamma_{ij}\frac{dx_j}{dt} = 0$$

The Rössler oscillator.

$$\begin{aligned}
\frac{dx_i}{dt} &= -\omega y_i - z_i - \textstyle\sum_{j\sim i}\gamma_{ij}x_j \\
\frac{dy_i}{dt} &= \omega x_i + \alpha y_i \\
\frac{dz_i}{dt} &= 0.2 + (x_i - 10)z_i
\end{aligned}$$

Remark. Note that in the case of the Rössler oscillator for parameters we have chosen, the dynamic "at rest" (*i.e.* without coupling) is equilibrium. This situation is similar to cross-immunity which makes oscillate a system *a priori* at equilibrium.

One possible interpretation of the phenomenon is that cross-immunity, suppressing some strains, forms negative feedback between the corresponding vertices of the graph (such as in Fig. 1 and 2). This, in turn, induces a positive feedback on other vertices.

Consider the second type of clustering of the eight-strain case as an example: if we assume that cross-immunity enhances negative loops between closest neighbours only, it generates as a consequence positive loops between strains inside each diagonal of the faces, that is inside the two complementary tetraedra α and β we have defined previously. These two groups are related such that they provide positive interactions inside and negative interactions outside against the whole other cluster. Thus we can expect they synchronize. The same idea may explain the apparition of two clusters of eight strains in the sixteen-strain case.

The phenomenon of self-organization of elements of an ordered system into clusters does not only occur in epidemiology: for instance, similar examples are

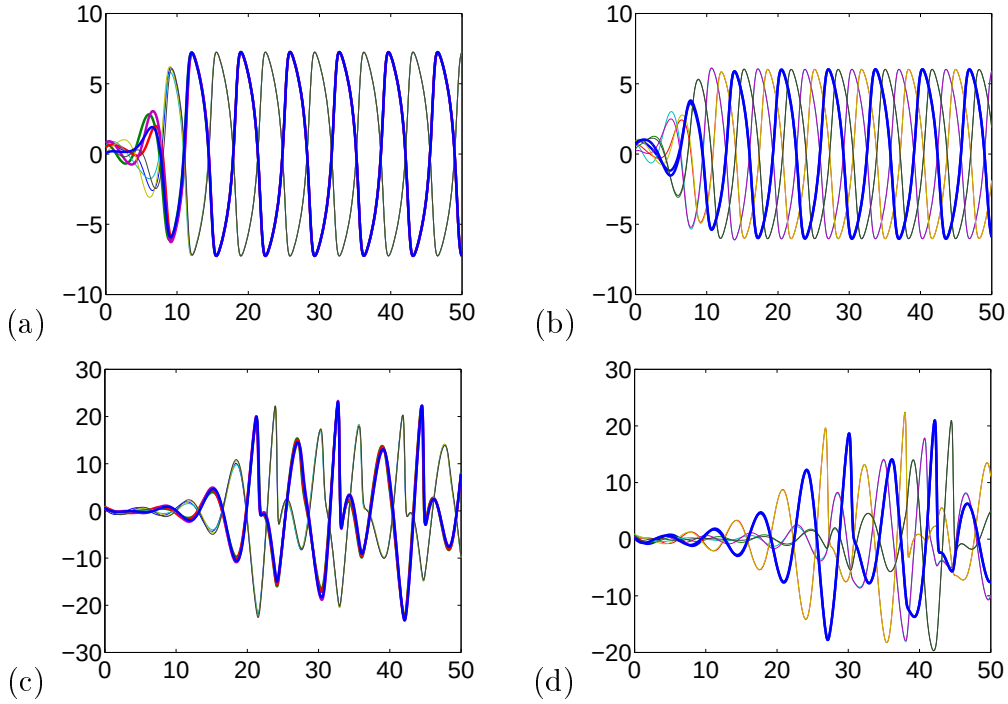


Figure 7: (*top*) Clustering for the Van der Pol-based system : $\omega = 1$, $\varepsilon = 0.1$ and (a) $\gamma_1 = 0$, $\gamma_2 = 0.4$: two clusters of tetraedric strains ; (b) $\gamma_1 = 0.4$, $\gamma_2 = 0.4$: four clusters of opposite strains. (*bottom*) Clustering for the Rössler-based system : $\omega = 1$, $\alpha = -0.1$ and (c) $\gamma_1 = 0$, $\gamma_2 = 0.2$: two clusters of tetraedric strains ; (c) $\gamma_1 = 0.2$, $\gamma_2 = 0.2$: four clusters of opposite strains. Once more time we have choosen one cluster and bolded strains inside (either two or four).

observed in neural networks, and we believe that it may occur in other applications. Clustering is a type of self-organisation similar to pattern formation.

The most remarkable feature of the clusters is that they behave remarkably regularly (at least for ordered strain sets), in contrast to the generally chaotic behaviour of isolated strains. Furthermore, a cluster structure, once formed, appears to be exceptionally stable. This stability implies that in many cases we can (and even should) consider the dynamics of a few clusters, instead of the dynamics of multiple separate strains, reducing in this way the system size.

The authors thank Dr S. Gupta and Mr M. Recker for helpful discussions and for providing a preprint.
AK is supported by the Foundation for Research, Science and Technology of New Zealand, through Project contract UOXX0101.
PKM was supported in part by National Science Foundation under Grant No PHY99-07949. PKM would like to thank KITP, UC Santa Barbara for their kind hospitality.

References

[1] Andreasen, V., Lin, J., Levin, S.A., 1997. The dynamics of cocirculating influenza strains conferring partial cross-immunity. J. Math. Biol. 35, 825–842.

[2] Anishchenko, V.S., Astakhov, V.V., Neiman, A.B., Vadivasova, T.E., Schimansky-Geier, L., 2002. Nonlinear dynamics of chaotic and stochastic systems, Springer, Berlin.

[3] Calvez, V., Korobeinikov, A.A., Maini, P.K., 2003. Models of multi strain infections. (preprint).

[4] Cliff, A.D., Haggett, P., Ord, J.K., 1986. Spatial aspects of Influenza epidemics, Pion, London.

[5] Dawes, J.H.P., Gog, J.R., 2002. The onset of oscillatory dynamics in models of multiple disease strains. J. Math. Biol. 45, 471–510.

[6] Gog, J.R., Swinton, J., 2002. A status-based approach to multiple strain dynamics. J. Math. Biol. 44, 169–184.

[7] Gog, J.R., Grenfell, B.T., 2002. Dynamics and selection of many-strain pathogens. Proc. Natl. Acad. Sci. 99, 17209–17214.

[8] Gupta, S., Trenholme, K., Anderson, R.M., Day, K.P., 1994. Antigenic diversity and the transmission dynamics of plasmodium falciparum. Science 263, 961–963.

[9] Gupta, S., Maiden, M.C.J., Feavers, I.M., 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.

[10] Gupta, S., Ferguson, N., Anderson, R.M., Chaos, 1998. Persistence, and evolution of strain structure in antigenically diverse infectious agents. Science 280, 912–915.

[11] Gupta, S., Anderson, R.M., 1999. Population structure of pathogens: the role of immune selection. Parasitol. Today 15, 497–501.

[12] Lin, J., Andreasen, V., Levin, S.A., 1999. Dynamics of influenza A drift : the linear three-strain model. Math. Biosciences 162, 33–51.

[13] Plotkin, J.B., Dushoff, J., Levin, S.A., 2002. Hemagglutinin sequence clusters and the antigenic evolution of influenza A virus. Proc. Natl. Acad. Sci. 99, 6263–6268.

[14] Rubin, J., Terman, D., 2000. Geometric analysis of population rythms in synaptically coupled neuronal networks. Neural Comput. 12, 597–645.

[15] Rubin, J., Terman, D., 2000. Analysis of clustered firing patterns in synaptically coupled networks of oscillators. J. Math. Biol. 41, 513–545.

[16] Terman, D., Lee, E., 1997. Partial synchronization in a network of neural oscillators. SIAM J. Appl. Math. 57, 252–293.

[17] Terman, D., Kopell, N., Bose, A., 1998. Dynamics of two mutually coupled slow inhibitory neurons. Physica D 117, 241–275.