A new theory of cytotoxic T-lymphocyte memory: implications for HIV treatment

Dominik Wodarz1*, Karen M. Page2, Ramy A. Arnaout3, Allan R. Thomsen4, Jeffrey D. Lifson5 and Martin A. Nowak1

1Institute for Advanced Study, Olden Lane, Princeton, NJ 08540, USA
2Centre for Mathematical Biology, Mathematical Institute, University of Oxford, 24–29 St Giles’, Oxford OX1 3LB, UK
3The Wellcome Trust Centre for the Epidemiology of Infectious Disease, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK
4Institute of Medical Microbiology and Immunology, The Panum Institute, 3C Blegdamsvej, DK-2200 N Copenhagen, Denmark
5Laboratory of Retroviral Pathogenesis, AIDS Vaccine Program, SAIC Frederick, National Cancer Institute—Frederick Cancer Research and Development Center, Building 335, Fifth Floor, Frederick, MD 21702, USA

We use simple mathematical models to examine the dynamics of primary and secondary cytotoxic T-lymphocyte (CTL) responses to viral infections. In particular, we are interested in conditions required to resolve the infection and to protect the host upon secondary challenge. While protection against reinfection is only effective in a restricted set of circumstances, we find that resolution of the primary infection requires persistence of CTL precursors (CTLp), as well as a fast rate of activation of the CTLp. Since these are commonly the defining characteristics of CTL memory, we propose that CTL memory may have evolved in order to clear the virus during primary challenge. We show experimental data from lymphocytic choriomeningitis virus infection in mice, supporting our theory on CTL memory. We adapt our models to HIV and find that immune impairment during the primary phase of the infection may result in the failure to establish CTL memory which in turn leads to viral persistence. Based on our models we suggest conceptual treatment regimes which ensure establishment of CTL memory. This would allow the immune response to control HIV in the long term in the absence of continued therapy.

Keywords: cytotoxic T-lymphocyte memory; HIV; primary infection; secondary infection; therapy; mathematical modelling

1. INTRODUCTION

A normal, efficient immune response to a virus comprises both innate components (natural killer cells, interferons, and other cytokines) and acquired components (B and T cells and their products). Immunological memory—defined as increased protection of the host against reinfection by a pathogen—is a central feature of the immune system. Protectiveness of B-cell memory, as well as transmission of protective antibodies from mother to offspring, has clearly been demonstrated (Thomsen & Marker 1988; Baldridge & Buchmeier 1992; Tew et al. 1992, 1997; Castelmur et al. 1993; Zinkernagel et al. 1996). However, the role of CDb+ cell responses to viral infections, as well as the protectiveness of cytotoxic T-lymphocyte (CTL) memory, are still under intense debate (Thomsen & Marker 1988; Lau et al. 1994; Matzinger 1994; Sprent & Tought 1994; Bachman et al. 1997; Doherty et al. 1996; Kundig et al. 1996a; Mullbacher & Flynn 1996; Doherty 1997, 1998; Selin & Welsh 1997; Zinkernagel et al. 1996; Dutton et al. 1998; Oehen & Brduscha-Riem 1998).

Experimental evidence suggests that CTL may be a major branch of the immune system fighting viral infections (Lin & Askonas 1988; Koenig et al. 1993; McMichael et al. 1995; Ehl et al. 2000). The T-cell receptor (TCR) recognizes viral antigen in conjunction with major histocompatibility complex (MHC) class I molecules and eventually causes the destruction of the infected cell. In the thymus, a large diversity of TCR specificities is created, which is narrowed down by positive and negative selection in order to ensure MHC restriction and avoid auto-reactivity. Naïve CTL precursors (CTLp), i.e. those which have never seen antigen, circulate around the body. Upon infection of the host, the antigen-specific CTLp start to proliferate and differentiate into effector cells. This is called the primary response to the pathogen. It may lead to two alternative outcomes, depending on viral and host characteristics: either the virus infection is cleared, or the virus establishes a persistent infection. If the virus is cleared, an elevated number of CTLp is observed long after virus clearance (Jamieson & Ahmed 1989; Oehen et al. 1992; Hsu et al. 1994; Lau et al. 1994; Mullbacher 1994). This ‘CTL memory’ is thought to protect the host more efficiently against secondary challenge with the virus.

*Author for correspondence (wodarz@ias.edu).
There has been considerable controversy about the exact nature and protectiveness of CTL memory. In particular the role of persisting antigen in maintaining CTL memory and ensuring an efficient secondary response is under intense debate (Jamieson & Ahmed 1989; Beverley 1990; Gray & Matzinger 1991; Oehen et al. 1992; Hou et al. 1994; Lau et al. 1994; Mullbacher et al. 1994; Bruno et al. 1995; Kundig et al. 1996a,b). Recent results indicate that the maintenance of CTLp is antigen independent and that the efficacy of the secondary response may or may not require persistent antigen, depending on the kinetics of effector cell production (Kundig et al. 1996a,b; Ehl et al. 1997). This has been shown with lymphocytic choriomeningitis virus (LCMV) and vesicular stomatitis virus (VSv) in mice. If the secondary infection is intravenous, then protection seems to be independent from the persistence of antigen (Kundig et al. 1996a,b). In this case, the virus population directly encounters the memory CTLp which leads to instant CTL activation and thus elimination of the infection before the appearance of clinical symptoms. On the other hand, protection against peripheral infection appears to be dependent on the persistence of antigen (Kundig et al. 1996a,b). This is because antigen persistence induces the expression of relevant markers on the surface of CTLp, such as LFA-1 and VLA-4 (Anderson et al. 1994; Zimmerman et al. 1996). This ensures constant recirculation through non-lymphoid tissues, which is required to recognize the invading virus before it replicates to high levels.

Many questions still remain open. Most importantly, it has not been established whether antigen persists after clearance of the infection, and if it does, how it persists and in what form. Since many viruses invade the host by peripheral infection, it is still unresolved whether the elevated number of CTLp, which remain present after virus clearance, confer protection against secondary challenge.

Here, we use mathematical models to analyse primary and secondary responses to viral infections. In particular, we investigate which factors determine whether the virus is cleared or persists during the primary response. We find that according to our model, virus clearance during the primary infection requires prolonged persistence of CTLp in the absence of significant levels of antigen, as well as an efficient rate of activation of the CTLp. Since these are commonly the defining characteristics of CTL memory, we hypothesize that the phenomenon referred to as ‘CTL memory’ has evolved in order to ensure virus elimination from the host during the primary infection. Protection against secondary challenge may be a carry-over effect resulting from the presence of elevated numbers of CTLp and may or may not be effective, depending on exact circumstances, such as route of infection, inoculum size, activation state or recirculation properties. We present experimental data from murine LCMV infection, supporting our hypothesis.

We extend our findings to analyse CTL memory in HIV-1 infection. A mathematical model describing the dynamics of CTL memory in the primary phase of the infection predicts that HIV may impair the generation of memory. It defines possible drug-treatment regimes that may reconstitute CTL memory in HIV-infected patients, resulting in immunological control of the virus in the absence of continued therapy.

2. CYTOTOXIC T-LYMPHOCYTE AND VIRUS DYNAMICS

In order to analyse the dynamics of antiviral CTL responses, we use the basic virus infection model (Nowak & Bangham 1996; DeBoer & Perelson 1998; Wodarz et al. 1999) taking into account uninfected (x) and infected (y) host cells. We assume that the CTL pool consists of two populations: the precursors (w) and the effectors (ζ). The model is explained schematically in figure 1 and is given by the following set of differential equations.

\[
\begin{align*}
\dot{x} &= \lambda - dx - \beta xy \\
\dot{y} &= \beta xy - ay - pzy \\
\dot{w} &= cyw(1 - q) - bw \\
\dot{z} &= cwz - hz \\
\end{align*}
\]

(1)

Target cells are produced at a rate \(\lambda\), die at a rate \(dx\), and become infected by virus at a rate \(\beta xy\). Infected cells die at a rate \(ay\) and are killed by CTL effector cells at a rate \(pzy\). Upon contact with antigen, CTLp proliferate at a rate \(cyw\) and differentiate into effector cells at a rate \(cwz\). CTL precursors die at a rate \(bw\), and effectors die at a rate \(hz\). Note that we have chosen a simplified and phenomenological model of CTL memory generation, since the exact mechanisms underlying the establishment of memory are still unknown. Our conclusions are therefore independent of the exact differentiation pathway of antiviral CTL. In addition, our results are independent of the mode of CTL action. Although our models describe a lytic CTL response, mathematical models describing non-lytic antiviral effector mechanisms mediated by CD8\(^+\) T cells have shown that these two pathways of CTL-mediated antiviral activity are qualitatively similar in the aspects relevant to the current analysis (Wodarz & Nowak 1998, 1999).

Before the host has been infected with the virus in question, the system is at the equilibrium, \(E_0\), given by \(x^{(0)} = \lambda/d, y^{(0)} = 0, w^{(0)} = 0, z^{(0)} = 0\). In practice, the number of virus-specific CTLp in the absence of the infection, \(w^{(0)}\), will be greater than zero. Thus, more precisely we should write \(w = \eta + c(1 - q)w - bw\) and \(w^{(0)} = \eta/h\). However, as long as \(\eta\) is small this represents only a small perturbation to the dynamics given by system (1).

Persistent infection of a naive host requires the basic reproductive ratio of the virus, \(R_0\), to be greater than unity. The basic reproductive ratio of the virus denotes the average number of infected cells produced by one infected cell at the beginning of the infection. For model (1) it is given by \(R_0 = \beta x/d\). Subsequent virus replication may either be limited by target cell availability, or by the CTL response. Target-cell-limited virus growth is described by equilibrium \(E_1\), given by

\[
\begin{align*}
\dot{x}^{(1)} &= a[\beta y^{(1)} = \lambda(a - d)/3, w^{(1)} = 0, z^{(1)} = 0. \\
\end{align*}
\]

If the immune system of the host is strong enough, the CTL population may expand in response to the virus

*Phil. Trans. R. Soc. Lond. B (2000)*
infection. The condition for CTL expansion is given by
\( \epsilon(1 - q)y^{(2)} > b \). If this condition is fulfilled, the system moves to equilibrium \( E_2 \), given by
\[
\begin{align*}
x^{(2)} &= \frac{\lambda x(1 - q)}{\alpha(1 - q) + b \beta}, \\
y^{(2)} &= \frac{b}{\epsilon(1 - q)}, \\
z^{(2)} &= \frac{\beta x^{(2)} - a}{p}.
\end{align*}
\]

We also note that for certain parameter values the system admits limit cycles.

### 3. PRIMARY INFECTION

Upon primary infection \textit{in vivo}, the virus population is either eliminated by the immune response, or the virus establishes a persistent infection controlled by the immune system. Since our model is deterministic, virus load cannot be reduced exactly to zero, but may be reduced to very low levels by a strong immune response. We can derive conditions for virus clearance versus persistence by assuming that the virus population goes extinct if \( y \) falls below a minimum value. The term 'virus clearance' \textit{in vivo} has to be defined in more detail. The host may recover from the infection either due to true clearance, i.e. extinction of the virus population, or due to reduction of virus load to undetectable levels. In the analysis of the model we will assume that virus is undetectable if \( y < y_{\text{min}} \) and extinct if \( y < y_{\text{ext}} \). Note that \( y_{\text{min}} \) is determined by the sensitivity of virus quantification, while \( y_{\text{ext}} \) corresponds to the situation where essentially less than one infectious unit is present in the host. Thus, \( y_{\text{min}} > y_{\text{ext}} \).

(a) **Elimination**

First, we consider the factors that determine whether the CTL response can drive the virus population extinct or not. The virus may be eliminated if virus load in the presence of the immune response is reduced to levels below the extinction limit, i.e. \( y^{(2)} < y_{\text{ext}} \). From the expression of the equilibrium virus load, \( y^{(2)} \), we can see that this may be achieved by a high rate of CTLp activation (high \( \epsilon \)) and a slow death rate of CTLp (low \( b \)). Note that elimination of the infection is independent of parameters involving CTL effector dynamics.

However, the system shows more complicated dynamics. If the value of \( b \) is very low, it takes a long time for virus load to converge to equilibrium \( y^{(2)} \). After initial oscillations, virus load settles to a quasi-equilibrium, \( \tilde{y} \), which decays only at a very slow rate.
Figure 2. Quasi-equilibrium dynamics of system (1). (a) In computer simulations, virus load settles to a quasi-equilibrium (Appendix A), where virus load only very slowly converges to its true equilibrium. (b–c) For realistic parameter ranges, the dependency of the quasi-equilibrium and the true equilibrium on the rate of CTL activation, $c$, and CTLp longevity, $b$, is very similar. Baseline parameters were chosen as follows. $\lambda = 10; \delta = 0.1; \beta = 0.001; a = 0.5; \rho = 1; c = 0.1; b = 0.001; q = 0.1; h = 0.1$.

(figure 2a). Appendix A derives an expression of virus load at this quasi-equilibrium in dependence of the rate of CTL activation, $c$, and the longevity of CTL precursors, $b$. The rate of CTL activation, $c$, has identical effects on virus load at equilibrium $y^2$ and at the quasi-equilibrium $\tilde{y}$ (figure 2b). The same holds true for the effect of CTLp longevity ($b$) on $y^2$ and $\tilde{y}$ over realistic parameter ranges (figure 2c). For very low values of $b$, tending towards zero, we observe an increasing discrepancy between $y^2$ and $\tilde{y}$ (figure 2c), as $\tilde{y}$ converges only very slowly to $y^2$.

Figure 3. Properties of dynamic elimination. (a–b) A high rate of CTLp activation (high $c$) and persistence of CTLp in the absence of antigen (low $b$), promote dynamic elimination. Baseline parameters were chosen as follows. $\lambda = 10; \delta = 0.1; \beta = 0.001; a = 0.5; \rho = 1; c = 0.1; b = 0.001; q = 0.1; h = 0.1$.

(c) If initial oscillations are not strong enough to result in dynamic elimination of the virus, stable limit cycles may be observed. Baseline parameters were chosen as follows. $\lambda = 10; \delta = 0.1; \beta = 0.001; a = 0.5; \rho = 1; c = 0.1; b = 0.2; q = 0.1; h = 0.1$.

(b) **Dynamic elimination**

Up to now, we have considered elimination of the virus due to reduction of virus load in the presence of CTL at equilibrium. However, the model also offers an alternative mechanism of virus elimination. The predicted dynamics of the primary infection are characterized by vigorous oscillations. That is, the virus population first replicates up to a peak and is then reduced by the rising CTL response. This initial oscillation may drive virus load to very low levels, below the extinction limit. We call this process ‘dynamic elimination’. Using numerical
simulations, we can show that a long life span of CTL precursors (low $b$), as well as efficient activation of CTL precursors (high $c$), maximize the chance of dynamic elimination (figure 3). Thus, the parameters which are most directly associated with CTL memory enhance the chances of both dynamic elimination and equilibrium elimination (see §3(a)).

Dynamic elimination refers to the situation where the initial antiviral CTL response is so strong that the virus population becomes extinct following its peak during acute infection. Whenever the parameters of the model lead to equilibrium elimination (as described in §3(a)), they also make dynamic elimination likely. The reverse, however, need not be the case. The initial oscillation in virus load can be so vigorous (and lead to a reduction of virus load many orders of magnitude below its initial value), that dynamic elimination also occurs for parameter values that would otherwise admit a positive equilibrium. This implies that virus elimination or persistence can depend on the initial conditions of system (1) or in biological terms on the size of the virus inoculum and the exact route of infection.

Some caution is necessary here: we studied dynamic elimination in a very simple system. Spatial heterogeneity and other realistic factors that occur in vivo are likely to dampen the initial oscillation. This means that some parameter values that lead to dynamic elimination in system (1) may not do so in more realistic versions of the model. The worst case, however, is that in certain systems, dynamic elimination only occurs for the same conditions as equilibrium elimination. To summarize, figure 4a shows the dynamics of virus clearance and memory generation as predicted by our model.

(c) Persistence

If CTLp activation and persistence in the absence of antigen is less efficient, then our model predicts that the virus population cannot be driven extinct. In this case, we get persistent virus replication in the presence of high levels of CTLp and low levels of CTL effector cells. Whether the host clinically recovers from the infection may depend on the exact strength of the CTL response, relative to the viral replication rate.

If the CTL response is still sufficiently strong and lies above a certain threshold, then $y_{\text{ef}} < y^* < y_{\text{min}}$. That is, virus load is at undetectable levels. This will happen if the rate of CTLp activation in response to antigen $\phi$ is still relatively high and the life span of the CTLp ($1/b$) is still relatively long. Figure 4b shows the dynamics of the resolution of a primary infection in such a scenario. The outcome looks similar as in the previous case: virus load and CTL effector cells drop to low or undetectable levels while the population of CTLp remains high after the resolution of the infection. However, the mechanism underlying the maintenance of an elevated number of CTLp is different in this case. When the virus is eliminated, the CTLp population is maintained due to an antigen-independent mechanism such as a long life span. If the virus is not completely eliminated, the CTLp population is maintained by a combination of longevity of CTLp and constant background stimulation by the pathogen replicating at levels below the threshold of detection. In this case, a fraction of the CTLp will be cycling at a rate proportional to $c$.

If the efficacy of the CTL response is weaker and lies below a threshold, we obtain a persistent infection with virus load at detectable levels. This occurs if the rate of CTLp proliferation in response to antigen $\phi$ and the life span of CTLp ($1/b$) are significantly lower, so that $y^* > y_{\text{min}}$. If the virus load remains too high in the presence of the CTL response, clinical symptoms may develop.

4. SECONDARY INFECTION

In the previous sections we have established that elimination or efficient control of the virus in the primary infection may require a high rate of CTLp activation and persistence of these cells in the absence of antigen. Hence, one function of the generation of CTL memory may be to ensure recovery from the primary infection. The question now arises if and under what circumstances CTL memory is protective against secondary infection by the virus, the context in which memory is traditionally considered. CTL memory can be considered protective if it reduces the peak virus load upon secondary challenge compared to primary challenge, and thus reduces the degree of clinical symptoms experienced. We have to consider the two cases of virus clearance and reduction to undetectable levels during primary challenge separately.

(a) Protection after virus elimination

Because of the antigen-independent persistence of CTLp required to drive the virus population extinct during the course of the initial infection, a large number of CTLp may still be present during secondary challenge by the virus. However, CTL effector cells will be absent since they die at a relatively fast rate. Since effector cells are required to combat the infection, the initial growth rate of the virus at the beginning of the secondary challenge will be positive. Thus, protection against secondary challenge depends mainly on the amount of time required for the CTLp to migrate to the focus of infection and to differentiate into effector cells. In our model, this is captured in the parameter $c_q$, the rate of differentiation into effector cells. Figure 5 shows the effect of increased CTLp abundance on the size of the peak virus load on secondary challenge, assuming different rates of effector cell production ($c_q$). Increased CTLp levels are only protective if effector function is produced sufficiently fast (large $c_q$) once the pathogen has entered the host. Strikingly, if there is a longer time-delay in the production of effector function (small $c_q$), increasing the abundance of CTL precursors even by four orders of magnitude does not lead to a significant reduction of the peak virus load and thus of clinical symptoms (figure 5a).

While we have defined protection upon secondary challenge as reduction in peak virus load, and thus clinical symptoms, it is interesting to consider the effect of increased CTLp numbers on the ability of the CTL response to eliminate the secondary infection. If the rate of CTL activation and the longevity of CTLp are high enough for the equilibrium virus load to lie below the extinction threshold, increasing the level of CTLp does not have any effect on the ability of the CTL response to clear the infection. However, the initial number of CTLp does have an influence on the minimum virus load during...
Figure 4. Resolution of the primary infection and generation of CTL memory as predicted by the model. (a) The virus population is driven extinct. Parameters are as follows. $\lambda = 10$; $d = 0.1$; $\beta = 0.001$; $a = 0.5$; $\rho = 1$; $c = 0.1$; $b = 0.001$; $q = 0.1$; $k = 0.1$. (b) Virus load is reduced below the threshold of detection, but not to extinction. Parameters are as follows. $\lambda = 10$; $d = 0.1$; $\beta = 0.001$; $a = 0.5$; $\rho = 1$; $c = 0.063$; $b = 0.0015$; $q = 0.1$; $k = 0.1$. Virus persistence results in the continued presence of a low-level CTL effector activity, and memory is maintained by a combination of CTLp longevity and antigenic stimulation. If the virus is cleared, CTLp are maintained independent of antigen, and the effector response diminishes over time.

(b) **Protection during viral persistence**

If the virus persists after primary infection (above or below the detection limit), CTL memory will be maintained by a combination of antigen-independent persistence of CTLp and constant stimulation of the CTL population. Hence, in contrast to the previous case, there will not only be memory precursor cells, but also memory effector cells. The size of the precursor population is given by $w^{(2)}$ and the size of the effector population by $z^{(2)}$. Therefore, upon secondary challenge, the initial growth rate of the virus is zero. Consequently, upon secondary challenge, the virus cannot grow and the host is always protected against clinical symptoms.
of another acute infection. However, this protection may be compromised if re-infection occurs with a virus strain that has mutated to become competitively superior to the immunizing strain, e.g. due to a faster rate of replication. In this case, the invading virus variant will have a positive initial growth rate and can therefore replicate up to a peak. The level of the peak virus load increases with increasing competitive superiority of the mutant compared to the wild-type (figure 3c).

5. CYTOTOXIC T-LYMPHOCYTE MEMORY IN LYMPHOCYTIC CHIORIOMENINGITIS VIRUS INFECTION

Lymphocytic choriomeningitis virus (LCMV) infection of mice is an excellent experimental system to study CTL responses under controlled conditions in vivo (Lehman-Grube 1971; Thomsen & Marker 1988; Zinkernagel 1993), and mice seem to be protected against reinfection (Zinkernagel et al. 1996; Kundig et al. 1996a; Planz et al. 1997). Recent studies have demonstrated that the CTL response may not completely clear LCMV, but reduces virus load below detectable levels (Planz et al. 1997). This is in agreement with previous studies showing that continued immune surveillance is required to keep LCMV in check (Volkert & Lundstedt 1968). Our model suggests that in this case, memory is likely to be protective. This is because persisting antigen sustains a small population of CTL effector cells that can rapidly attack the virus at an early stage of the disease without the need for CTLp expansion and differentiation. In agreement with theory, Zimmerman et al. (1996) found that in this case a fraction of the CTLp is cycling while the remaining CTLp show a resting phenotype.

However, our models also suggest that the protective-ness of memory maintained by persistent antigen is compromised if the challenging virus has different properties compared to the original strain. The differences required to overcome memory need not be immune escape mutations, but may simply be differences in the replication kinetics of the virus.

On the other hand, if the primary infection has been cleared to extinction, protection upon secondary challenge is not as straightforward, since after some time all memory cells show the CTLp phenotype and the effector cell population has vanished due to the limited life span of these cells. In this case, an elevated abundance of CTLp is only protective if the time window between invasion of the pathogen and induction of antiviral CTL effector action from this pool is very short. If there is a significant time-delay, the persistence of an elevated number of CTLp may be less protective or not protective at all. By the time CTL effector function has been produced, virus load will have already risen to high levels. These predictions are confirmed by recent experimental data. While CTLp may persist independently of antigen in mice (Jamieson & Ahmed 1989; Hou et al. 1994; Lau et al. 1994; Mullbacher et al. 1994; Bruno et al. 1995; Doherty et al. 1996), protectiveness of this memory CTL population crucially depends on the kinetics of effector cell production (Ehl et al. 1997). The longer it takes for the effector function to be generated, the less protective the memory population becomes. Consequently, if immunized mice are injected intravenously with LCMV, CTL memory is protective, since the antigen directly encounters the CTLp pool (Kundig et al. 1996a,b). On the other hand, if LCMV is injected peripherally, and the memory CTL population is maintained in the absence of antigen, the CTL precursors do not instantly recognize the presence of the virus, and this time-delay enables the virus to out-replicate the CTL response. This renders the memory population unprotective (Kundig et al. 1996a,b; Ehl et al. 1997).

Phil. Trans. R. Soc. Lond. B (2000)
Figure 6. Data re-plotted from Christensen et al. (1994) and Thomsen et al. (1996). (a) Dynamics of LCMV infection in class-II-deficient mice compared to wild-type mice. Although initial control of the virus is similar in wild-type and mutant, blood virus titres resurge in class-II-deficient mice. (b) The primary CTL response in wild-type and mutant mice is similar. However, class-II-deficient mice show a lack of CTL memory response.

Given the findings that CTL memory only protects against secondary challenge in a restricted set of circumstances, but is both necessary and sufficient for the resolution of the primary infection, we propose that the memory phenotype—traditionally considered in the context of protection—has evolved primarily in order to resolve the virus infection. With this in mind, it would be interesting to reconsider and re-interpret experimental findings and to devise new experiments to test our hypothesis. We present data from LCMV infection of MHC class II knockout mice (figure 6) and CD40L−/− knockout mice (figure 7), supporting our theoretical predictions.

During the early phase of the infection, both wild-type and class-II-deficient mice reduce virus load to undetectable levels (figure 6). However, in contrast to wild-type mice, virus load re-emerges to high levels in class-II-deficient mice about a month after infection. Resurgence of virus load is associated with a lack of a significant memory CTL response in class-II-deficient mice (figure 6). Thus, the absence of CD4+ T-cell help may interfere with the generation and/or maintenance of CTL memory (Battegay et al. 1994; Von Herrath et al. 1996; Thomsen et al. 1996).

Similar results are obtained from CD40L−/− mice infected with LCMV. The primary response to LCMV-TRA2 in wild-type and mutant mice was similar (figure 7). However CD40L−/− mice lack CTL memory (figure 7). While virus load is suppressed to low or undetectable levels in wild-type mice, levels of LCMV remain high at 28 and 80 days post-infection in CD40L−/− mice (figure 7). A possible mechanism for the lack of CTL memory in CD40L−/− mice is that the interaction between T-helper cells, antigen-presenting cells and CD8+ T cells is impaired (Ridge et al. 1998; Schoenberger et al. 1998). Although these interactions may not be required for initial activation and differentiation of primary effector cells, they are essential for the generation of memory cells (Borrow et al. 1996, 1998; Thomsen et al. 1998).

Interpretation of these data strongly supports our theory that CTL memory may be necessary to successfully resolve the primary infection. The experimental results can be explained intuitively as follows. If CTL memory is compromised, a reduction of virus load is
accompanied by a decline of the CTL population to low levels. This decline of the CTL population enables the virus to regain a positive growth rate and therefore to settle at an equilibrium level that is well above the detection limit. In the wild-type mice, where memory is intact, reduction of virus load does not result in a decline of the CTLp to low levels. Instead, the CTLp settle at a stable memory level and therefore prevent the virus from attaining a positive growth rate. Consequently the virus stays at undetectable levels or goes extinct.

6. CYTOTOXIC T-LYMPHOCYTE MEMORY IN HIV INFECTION

These findings also have implications for understanding the mechanism underlying the survival of viruses persistently infecting their host despite the presence of an efficient CTL response. Examples of such viruses are HIV, human T-cell leukaemia virus (HTLV) and Epstein-Barr virus (EBV). All these viruses infect cells that are involved in the immune response and could therefore potentially interfere with the generation and/or maintenance of CTL memory, especially when the target cells are CD4+ T cells (Borrow et al. 1996, 1998; Von Herrath et al. 1996; Thomsen et al. 1996, 1998).

Particularly interesting is the case of CTL memory in HIV infection. Our model describing the dynamics of CTL memory can easily be adapted to HIV infection by assuming that memory generation depends on CD4+ T-cell help, and that infection of CD4+ T cells results in impaired T-cell help. This is expressed in the following set of differential equations (Wodarz et al. 1998):

\[\begin{align*}
\frac{dx}{dt} &= \lambda - dx - \beta xy \\
\frac{dy}{dt} &= \beta xy - ay - pyz \\
\frac{dz}{dt} &= cxyw - cqw - bw \\
\frac{dw}{dt} &= cxyw - hw
\end{align*}\]

(2)

The model is derived from system (1) and assumes that the target cells for the virus are CD4+ T cells. It includes the additional feature that expansion of the CTLp.
population is proportional to both antigen \( y \) and the number of uninfected CD4\(^+\) T cells \( x \) capable of delivering T-cell help. We also assume that differentiation into effector function is independent of CD4\(^+\) T-cell help (Doherty 1993). We will briefly summarize the properties of these equations (for more details, see Wodarz et al. 1998). If the basic reproductive ratio of the virus, \( R_0 \), is greater than unity, the system may converge to one out of two equilibria. Either CTL memory is successfully established, or it fails to become established. The establishment of memory depends on host and viral parameters as well as on initial conditions. Among the host parameters, a high rate of CTLp activation (high \( e \)) and longevity of CTLp (small \( b \)) promote the establishment of CTL memory. On the other hand, a high rate of viral replication (large \( \beta \)) interferes with the generation of memory because it reduces the number of functional T-helper cells to low numbers and thus significantly compromises the rate of CTLp expansion. If the viral replication rate is below a certain threshold value \( \beta < \beta_1 \), CTL memory may be established. On the other hand, if the replication rate of the virus is above another threshold, \( \beta > \beta_2 \), CTL memory can never be successfully established. However, if viral replication lies between these two thresholds, the outcome depends on the initial conditions. A high initial virus load, a low initial CD4\(^+\) T-cell count, as well as the naive state of the host, interfere with the generation of CTL memory. This behaviour is summarized in figure 8.

Since HIV replicates up to high viral loads during the primary phase of the infection, our model suggests that this may result in T-helper cell impairment sufficient to prevent the establishment of CTL memory. Since, according to our argument, the CTL memory phenotype is required to clear the infection, HIV may replicate persistently. Figure 9a simulates the primary phase of HIV infection. The virus population initially grows to a peak and then declines to settle at a stable equilibrium level. The CTL memory population expands up to a peak and then declines towards extinction. Since the memory CTL response has gone extinct, the virus may only be controlled by non-memory CTL responses which can persist during continued antigentic stimulation, but which decline to extinction if this stimulus diminishes. According to models, such immune responses cannot eliminate or efficiently control the infection in the long term (Nowak et al. 1991, 1995; Nowak & Bangham 1996; Wodarz et al. 1998). Quantification of HIV-specific CD8\(^+\) T cells with the tetramer method (Altman et al. 1996) during anti-retroviral therapy revealed a sharp decline of virus-specific CTL after removal of the antigenic stimulus (Gray et al. 1999; Ortiz et al. 1999). This indicates that persistent viral replication is required to maintain the CTL response in these patients. Moreover, in long-term non-progressors controlling viraemia in the absence of therapy, extremely low viral loads are associated with a strong CTL response (Harrer et al. 1996a,b) and vigorous HIV-specific CD4\(^+\) T-cell proliferative responses (Rosenberg et al. 1997). Interestingly, HIV-exposed but uninfected individuals tend to show specific CTL responses detectable 34 months after the last virus exposure (Bernard et al. 1999).

(a) Treatment during primary HIV infection

The model can be used to analyse anti-retroviral drug therapy during the primary phase of HIV infection. This can be done by replacing the infection term by \( s \beta x \), where \( s \) denotes the efficacy of the drug inhibiting viral replication. Its values range from \( s=1 \) (no effect of drug) to \( s=0 \) (100% efficient drug). The model predicts that a certain drug therapy regime during the primary phase of HIV infection may result in the establishment of CTL memory. This in turn allows the immune system to eliminate or control the infection without continued administration of antiviral drugs. Such a drug treatment regime is illustrated in figure 9b. The virus first needs to replicate up to a certain level in order to induce stimulation and

Figure 8. Schematic summary of the dynamics of CTL memory generation in HIV infection (equation (2)). Increasing the replication rate of the virus increases the amount of immune impairment and thus interferes with memory generation. For intermediate replication rates, establishment of memory depends on the initial conditions. For details, see text.

Phil. Trans. R. Soc. Lond. B (2000)
expansion of the CTLp population. As virus load rises, treatment has to be initiated. If treatment is effective, it will reduce the amount of viral replication to zero or low levels. This results in reduced amounts of immune impairment which in turn allows the generation of CTL memory. Once significant CTL memory has been generated, drug treatment can in principle be withdrawn. According to model predictions, the CTL memory response cannot go extinct anymore because the dynamics of the CTLp population depends on the initial number of CTLp upon resumption of efficient viral replication, e.g., when drug treatment is discontinued. Treatment has allowed the number of CTLp to rise to levels high enough to prevent exhaustion of memory. Because of the presence of CTL memory, the immune system will be able to control the infection in the long term, suppressing viral replication to low levels.

The duration of treatment required for the successful establishment of CTL memory depends on the replication rate of the virus, the rate of CTL activation, and the time when treatment is initiated. The lower the rate of CTL activation, and the higher the replication rate of the virus, the longer the duration of treatment required for the establishment of memory. The latter that treatment is initiated, the higher the virus load may rise, and the longer the duration of therapy required to overcome the T-helper cell impairment by the virus. If treatment is initiated after a certain time threshold, CTL memory cannot be generated anymore. This time threshold in turn depends on the rate of viral replication and the efficacy of the CTL response.

Figure 9. Simulation of the primary phase of HIV infection. (a) Due to T-helper cell infection, CTL memory fails to be generated and a persistent infection is established. (b) Drug therapy (shaded region) during the early phases of HIV infection may allow the establishment of CTL memory and thus virus control in the absence of continued therapy. Parameters were chosen as follows. \( \lambda = 1; d = 0.1; \beta = 0.5; a = 0.2; p = 1; c = 0.1; b = 0.001; q = 0.5; h = 0.1; s = 0.0042 \).

Figure 10. Drug treatment (shaded region) during the asymptomatic period of the infection. (a) Effective treatment reduces virus load to low or undetectable levels. However, withdrawing the drugs results in the re-emergence of virus load to pre-treatment levels. (b) CTL memory may be reconstituted by a therapy regime consisting of two phases. In the first phase, drug treatment reduces virus load to low levels. This treatment should continue at least until virus load has reached a new equilibrium and may be continued for longer in order to allow a certain degree of regeneration of the CD4\(^+\) cell count. The second phase includes a drug holiday followed by another period of drug therapy. This allows reconstitution of CTL memory and control of viraemia in the absence of treatment. For details, see text. Parameters were chosen as follows. \( \lambda = 1; d = 0.1; \beta = 0.5; a = 0.2; p = 1; c = 0.1; b = 0.001; q = 0.5; h = 0.1; s = 0.0042 \).

(b) Treatment during the asymptomatic phase of HIV infection

If memory generation fails during the primary phase of the infection, the virus population will settle to an equilibrium, and the level of virus load may be determined by a combination of target cell availability and non-memory immune responses. This marks the beginning of the asymptomatic period which, in the absence of therapy, will eventually culminate in the development of AIDS. Although drug therapy may suppress virus load to very low levels during the asymptomatic phase and therefore ensures a longer life span of the patient (Arnaout et al. 1999), therapy has to continue for life. However, severe side-effects induced by the drugs, as well as problems regarding drug resistance, render this strategy doubtful.

Here, we propose a conceptual treatment regime that should theoretically allow reconstitution of CTL memory, resulting in virus control in the absence of continuous therapy (figure 10).

This approach includes two phases of treatment. The initial treatment starts in the asymptomatic period and should reduce virus load to low levels. This initial treatment should be sustained at least until virus load has settled to its new equilibrium or until it has dropped to

\[ v(t) - v(t) = f(t) \]

\[ \frac{dS}{dt} = \lambda S - dS - aS = 0 \]

\[ \frac{dC}{dt} = \beta S - \delta C - aC = 0 \]

\[ \frac{dL}{dt} = \delta C - bL - aL = 0 \]

\[ \frac{dM}{dt} = \gamma L - cM - aM = 0 \]

\[ \frac{dP}{dt} = \phi M - pP - aP = 0 \]

\[ \frac{dV}{dt} = \theta P - qV - aV = 0 \]

\[ \frac{dI}{dt} = \psi V - hI - aI = 0 \]

\[ \frac{dF}{dt} = \omega I - sF - aF = 0 \]
undetectable levels. A longer duration of treatment would be preferable to maximize the CD4+ T-cell count increase and immune system recovery during this phase. The next phase involves that the patient should go on a drug holiday. The drug holiday will allow a rise in virus load that essentially mimics the primary phase of HIV infection. The positive growth rate of the virus population provides another stimulus for the expansion of a CTL memory response. As virus load increases, drug therapy should be resumed in order to reduce T-helper cell impairment. This should allow the generation of CTL memory, similar to the situation in the primary infection described above. Once CTL memory has successfully been established, it may be possible to withdraw drug treatment for good. Depending on the efficacy of the memory response, the virus may be controlled by the immune system in the absence of therapy. Factors determining the duration of the second phase of therapy are the same as those that determine the duration of treatment during primary infection. Thus, a lower rate of CTL activation, a higher rate of viral replication, and later initiation of treatment while virus load rises require a longer duration of the secondary phase of therapy in order to reconstitute CTL memory. Repeated cycles of therapy and drug holidays may maximize the chances of success. In order to minimize the emergence of resistant mutants, discontinuation of therapy should involve abrupt simultaneous cessation of all drugs.

However, this treatment regime may become problematic if the patient has already progressed further in the disease process and has a highly depressed CD4+ cell count. In such patients, T-cell help will be weak even under therapy, and this prevents the reconstitution of CTL memory. Hence, our models predict that initiating this treatment regime early in the course of the infection may maximize the chances of achieving sustained immunological control of the virus in the absence of life-long drug treatment.

Our predictions are in very close agreement with clinical data from HIV-infected patients (Lisziewicz et al. 1999; Lori et al. 1999; Markowitz et al. 1999; Ortiz et al. 1999). They are also supported by experiments on drug treatment during the primary phase of SIV infection in macaques, as well as by re-challenge studies during the asymptomatic period (Lifson et al. 2000).

7. CONCLUSIONS

We have used mathematical models to analyse the dynamics of primary and secondary challenges in viral infections. We determined the conditions required for CTL-mediated clearance of the primary infection, and for CTL-mediated protection against secondary challenge. The fundamental result of our studies is that CTL memory is required for the resolution of the primary infection. On the other hand, the degree of protection against secondary challenge may depend on the exact circumstances, such as the route of infection, inoculum size and viral phenotype. This suggests that CTL memory may have evolved in order to resolve the primary infection. Our predictions about the dynamics of secondary challenge by the virus are in good agreement with findings in the literature and may explain conflicting experimental observations as well as the controversy concerning the nature and protectiveness of CTL memory.

This novel perspective on the dynamics of CTL responses offers new insights into mechanisms underlying viral persistence in a variety of infections, including LCMV, HTLV and HIV. In particular, the implications for HIV infection are of special importance. The models offer a new understanding about the relationship between immune impairment, memory generation and viral persistence. Based on this understanding, the models suggest possible treatment regimes that may allow the successful establishment of CTL memory during the primary phase, or the reconstitution of CTL memory during the asymptomatic period of HIV infection. Such treatment regimes could shift the balance between HIV and the immune response in favour of the immune system. Our theory, together with further experimental studies, may allow us to devise a clinically feasible treatment schedule for HIV-infected patients that may result either in long-term control of the virus, or in elimination of HIV from the host without the need of life-long antiviral therapy.

We thank Charles Bangham, Stephan Ehl and Paul Klenerman for helpful comments and discussions. This work was supported by the Institute for Advanced Study (D.W. and M.A.N.), the Wellcome Trust (K.M.P.), and the Marshall Aid Commemoration Commission (R.A.A.).

APPENDIX A. THE QUASI-EQUILIBRIUM IN SYSTEM (1)

In this Appendix, we derive analytical insights into the long-term behaviour of system (1) for small values of $b$ (that is long-lived CTL, $\dot{w}$).

(a) Two-species model with no decay of the immune cells ($b=0$)

Let us first consider the limiting case, $b = 0$. We have

$$\begin{cases} \dot{x} = \lambda - dx - \beta xy \\ \dot{y} = y(\beta x - a - pc) \\ \dot{w} = \epsilon (1 - q)yw \\ \dot{z} = eqyw - hz \end{cases}$$

(A1)

In this case, there is no interior equilibrium and the infection will eventually be eliminated.

We study a simplified system by assuming that $x$ is at its equilibrium value, $x = \lambda/d$, and $z$ is at a steady state given by $z = 0$, which implies $z = eqyw/h$. These assumptions lead to

$$\begin{cases} \dot{y} = y(A - Byw) \\ \dot{w} = Cyw \end{cases}$$

(A2)

where $A = \beta \lambda/d - a$, $B = pcq/h$ and $C = \epsilon (1 - q)$, in terms of the full model parameters. We call the initial values of $y$ and $w$, $y_0$ and $w_0$, respectively.

Numerical simulations show that system (A2) is an excellent approximation of system (A1) for the initial peak of $y$, and the subsequent long-term dynamics. Initially, $y$ grows exponentially until it reaches some maximum value and then abruptly starts to decrease.
(i) **Peak dynamics**

Initially, $\omega$ and $\gamma$ are very small and so the exponential growth rate for the infected cells is $A$. We have approximately

\[ y(t) = y_0 e^{\omega t}. \]

The maximum is reached when

\[ y(t)w(t) = \frac{A}{B}. \]  

(A4)

Solving

\[ \dot{w} = C_{w_0} e^{\omega t} w, \]  

we get

\[ w(t) = w_0 \exp \left( \frac{C_{w_0}}{A} (e^{\omega t} - 1) \right). \]  

(A6)

Substituting for $w$ and $y$ in equation (A4) yields

\[ w_0 y_0 \exp \left( A t_m + \frac{C_{w_0}}{A} (e^{\omega t_m} - 1) \right) = \frac{A}{B}, \]  

where $t_m$ is the time at which the maximum in $y$ is reached. Equation (7) can be written as

\[ x + Ke^t = L, \]  

(A8)

where $x = At_m$, $K = C_{y_0}/A$, and $L = \ln (A/Bw_0 y_0) + C_{y_0}/A$. The solution can be expressed as

\[ t_m = \frac{L}{A} - \frac{1}{A} \Omega(K e^t), \]  

(A9)

where $\Omega$ is the omega (or LambertW) function, which satisfies $\Omega(x e^{\Omega(x)}) = x$. Substituting into equation (3), we get the maximum value of $y$, which is given by

\[ y_m = \frac{y_0}{K} \Omega(K e^{t_m}). \]  

(A10)

(ii) **Long-term dynamics**

In the second phase of the dynamics, in which $\gamma$ is decreasing, we find that $\xi \equiv yw$ is approximately constant. Considering the equation for $\xi$, we find that this implies

\[ A - Byw + Cy = 0. \]  

(A11)

Substituting into equation (2), we obtain

\[ \dot{y} = -Cy. \]  

(A12)

This has solution

\[ y(t) = \frac{1}{C(t - t_0) + \frac{1}{y(t_0)}}, \]  

(A13)

where $t_0$ is some fixed time. If we assume that this approximate solution is valid from the point at which $y$ reaches its maximum, then we can substitute in $t_0 = t_m$ and $y(t_0) = y_m$. This yields

\[ y(t) = \frac{1}{C(t - \tau)}, \]  

(A14)

where

\[ \tau = \frac{1}{A} \left( \frac{L}{\Omega(K e^{t_m})} - \frac{1}{\Omega(K e^t)} \right). \]

(b) **Two-species model with decay of immune cells**

(\( b \neq \theta \))

With decay of the immune cells the system (A1) becomes

\[ \begin{align*}
\dot{x} &= \lambda - dx - \beta xy \\
\dot{y} &= y(\beta x - a - p \zeta) \\
\dot{w} &= c(1 - q) yw - hw \\
\dot{z} &= cq yw - hz
\end{align*} \]

(A15)

and hence (A2) becomes

\[ \begin{align*}
\dot{y} &= y(A - Byw) \\
\dot{w} &= C_{w_0} e^{\omega t} - bw.
\end{align*} \]  

(A16)

Again, the dynamics have two phases: first, $y$ grows exponentially with rate $A$, and then it decreases with $yw$ in quasi-equilibrium.

(c) **Peak dynamics**

Once again, in the initial phase, we have

\[ y(t) = y_0 e^{\omega t}, \]  

(A17)

and we solve

\[ \dot{w} = C_{w_0} e^{\omega t} - bw, \]  

(A18)

to give

\[ w(t) = w_0 \exp \left( \frac{C_{w_0}}{A} (e^{\omega t} - 1) \right). \]  

(A19)

The first equation of system (A16) implies $yw = A/B$ at the maximum, so substituting for $\gamma$ and $w$, as before, we get an equation in the time, $t_m$, at which the maximum is reached:

\[ w_0 y_0 \exp \left( (A - b)t_m + \frac{C_{w_0}}{A} (e^{\omega t_m} - 1) \right) = \frac{A}{B}. \]  

(A20)

Equivalently, we have

\[ xs + Ke^t = L, \]  

(A21)

where $x = At_m$, $K$ and $L$ are as in the previous section and $r = (A - b)/A$. This is just

\[ x + Ke^t = L', \]  

(A22)

where $K' = K/r$ and $L' = L/r$ and so has solution

\[ t_m = \frac{L'}{A} - \frac{1}{A} \Omega(K' e^{t_m}), \]  

(A23)

and

\[ y_m = \frac{y_0}{K'} \Omega(K' e^{t_m}). \]  

(A24)
Long-term dynamics

In the second phase of the dynamics, in which \( \gamma \) is decreasing, \( \xi = \varphi \) is roughly constant, which implies

\[
Cy - b + A - Byw = 0,
\]

(A25)

and hence

\[
y = y(b - Cy).
\]

(A26)

This has solution

\[
\frac{1}{y} = \frac{C}{b} \left(1 - \frac{1}{y(b)}\right) \left(1 - \frac{e^{(a-a_{0})t}}{a_{0}}\right) + \frac{1}{y(0)} \left(1 - e^{(b-a_{0})t}\right),
\]

(A27)

where \( y(0) \) is some fixed time. If we assume this solution is valid from the point at which \( y \) reaches its maximum, then substituting \( t_{0} = t_{m} \) and \( y(t_{0}) = y_{m} \), we get

\[
\frac{1}{y} = \frac{C}{b} \left[1 - \frac{e^{(a-a_{0})t}}{a_{0}}\right] + \frac{1}{y_{0}} e^{(b-a_{0})t}.
\]

(A28)

REFERENCES


Markowitz, M. (and 10 others) 1999 The long-term response to combination therapy commenced early in HIV infection: 3 year follow-up. What’s happened since Vancouver? In Sixth Conference on Retroviruses and Opportunistic Infections. (Abstract)

Matzinger, P. 1994 Immunology: Memories are made of this? [see news; comment]. Nature 369, 605–606.


Von Herrath, M. G., Yokoyama, M., Dockter, J., Oldstone, M. B. & Whitton, J. L. 1996 CD4-deficient mice have reduced levels of memory cytoytic T lymphocytes after immunization and show diminished resistance to subsequent virus challenge. J. Virol. 70, 1072–1079.


