Dynamics of immune escape in HIV infection

Domänik Wodarz and Martin A. Nowak
Institute for Advanced Study, Princeton, New Jersey, U.S.A.

Introduction

Human immunodeficiency virus (HIV), the pathogen causing AIDS, shows highly complex interactions with the host. It infects cells that are involved in the immune response and simultaneously induces the collapse of the immune system, thereby killing the patient (Fig. 1). After infection, in the primary phase, HIV replicates to high titers before virus load drops to relatively low levels. The initial fall in virus load is often accompanied by a rise of anti-viral immune responses [1-3], and marks the beginning of the asymptomatic period during which the patient remains healthy. The duration of the asymptomatic phase is highly variable, but typically lasts for about 10 years [4]. As the infection continues, the immune system slowly deteriorates until it finally collapses leading to the transition to full-blown AIDS [5]. The development of AIDS is often characterized by a sharp and sudden rise in virus load.

The factors contributing to the eventual progression to AIDS as well as the reason for the variability in the duration of the asymptomatic period are still not properly understood. Recent quantitative virological data have shown that the asymptomatic phase of HIV infection is not a period of latency, but a dynamic process involving continuous rounds of de novo replication and infection [6-12]. It was shown that the half life of productively infected cells is around 2 days and that free virus particles in the plasma have an even shorter half life of about 6 hours. This rapid viral turnover during the asymptomatic phase indicates a great potential for the virus to evolve in response to select pressure exerted by the immune system or by drug treatment [13-14].

Such selection dynamics involve nonlinear interactions of many different components and make the use of mathematical models necessary to provide a correct interpretation of empirical results as well as to generate new insights and hypotheses. We demonstrate this approach by discussing mathematical models describing the dynamics between HIV and the immune response. Specifically, we show that the evolution towards the created antigenic diversity may be the driving force underlying HIV disease progression and the reason for the eventual breakdown of the immune system upon development of AIDS. Such insights have important implications for designing efficient treatment regimens for HIV-infected patients.

Key words: Mathematical models, population dynamics, virus, HIV, AIDS, immune response, CTL, escape, evolution.

Summary

The dynamics between pathogens and the immune system involve complicated interactions of many different components and thus the use of mathematical models necessary to provide a correct interpretation of empirical results as well as to generate new insights and hypotheses. We demonstrate this approach by discussing mathematical models describing the dynamics between HIV and the immune response. Specifically, we show that the evolution towards the created antigenic diversity may be the driving force underlying HIV disease progression and the reason for the eventual breakdown of the immune system upon development of AIDS. Such insights have important implications for designing efficient treatment regimens for HIV-infected patients.

HIV and the immune system

The interactions between HIV and the immune system are complex. Viral proteins, called antigens, are displayed on the surface of HIV-infected cells as well as on antigens presenting cells (APCs). Specific immune mediators recognize a certain part of the antigen, called epitope. Central to the antiviral immune response are T helper cells.
cells. T helper cells recognize viral antigen on the APCs. This triggers them into proliferation and induces them to release immunomodulatory chemicals called cytokines. T helper cells facilitate the activation of two major branches of the immune system responsible for fighting viral pathogens.

1. They interact with B cells and activate them. Activated B cells release neutralizing antibodies into the plasma which recognize and attach to free virus particles. This may prevent the virus from entering susceptible target cells or may lead to the destruction of the virus.

2. The other branch of the immune system fighting viral infections is the CTL (cytotoxic T lymphocyte) response. CTLs recognize viral antigen on the surface of infected cells. Upon recognition, they are driven into proliferation and this proliferation requires the presence of the cytokines released by the T helper cells. On further encounter with viral antigen, CTLs release certain chemicals resulting in the lysis of infected cells. Besides their lytic action, CTLs are also thought to release chemical substances which can block viral replication.

These immune mediators are antigen specific. That is, if the epitope changes through mutation (antigenic variation), they may lose the ability to recognize and kill infected cells. This is called immune escape [18-20] and is shown schematically in Fig. 2. Several epitopes recognized by immune cells may easily change without compromising the replicative capacity of the virus. Other epi-
Wodarsz and Nowak: Dynamics of immune escape in HIV infection

![Diagram of immune escape in HIV infection](image)

**Fig. 2.** Antigen recognition by immune cells. Several cell types of the immune system responsible for fighting viral antigens carry receptors on their surface which specifically recognize viral antigens. An example are CTLs interacting with infected cells. If the virus mutates and the antigen consequently changes in sequence and configuration, a given receptor may lose its ability to recognize the newly evolved virus strain. This is called immune escape by the virus.

**Fig. 3.** Asymmetry in the dynamics between HIV and the immune response. A given immune cell may only be able to fight the virus strain $v$ if specific for $v$. However, the HIV population may impair the immune response regardless of its specificity.

The immune responses described above are typical for most viral infections and are usually sufficient to control or eradicate a virus from the host. However, HIV is different from most other viral infections in that it fights back against the immune system. The virus may infect T helper cells thereby rendering them non-functional or destroying them. Since the loss of T helper cell function also compromises the efficacy of the CTL and antibody responses, HIV is said to cause immune impairment. However, whereas the immune response may need to be specific for a given viral strain in order to fight it, any HIV strain may infect T helper cells regardless of their specificity and consequently weakens the overall immune system. This inherent asymmetry in the dynamics between HIV and the immune system is illustrated in Fig. 3.

**Variability in HIV infection**

For infection to be possible, HIV has to integrate into the genome of its target cells. This process requires the action of the enzyme reverse transcriptase, copying the viral RNA genome into a DNA sequence which is required for integration. However, reverse transcription is an error-prone enzyme leading on average to one mistake per genome per replication cycle [21-22]. This is, there is a relatively high chance that changes are incorporated into the HIV genome during its reproduction. This makes it possible that various characteristics of HIV may be rapidly changing [23-25]. In fact, the error rate of the reverse transcriptase of HIV seems to maximize the chance of producing antigenic escape mutants [26]. Such escape mutants may be controlled by other immune cells, specific for the newly formed virus mutant. The continuous production of escape mutants is thought to be an important viral strategy for maintaining a persistent infection [27-31]. These dynamics have been formulated in simple mathematical models. The core of such models is the above mentioned asymmetry between HIV and the immune system, i.e. that immune cells need to be specific to kill the virus, but that the virus may kill immune cells regardless of their specificity [13, 32-34].

The model takes into account three variables: the different virus strains, $v$, and the strain-specific immune responses, $x$, and a cross-reactive immune response, $z$. Describing the overall virus population by $v = x + z$, it is given by the following set of differential equations:

\[
\begin{align*}
\frac{dv}{dt} &= v(z - px - qv) \\
\frac{dx}{dt} &= x(z - bx - wx) \\
\frac{dz}{dt} &= b(z - b) - wz
\end{align*}
\]

The virus strains replicate on average at a rate $r$ and are killed by the strain-specific and cross-reactive immune responses at a rate $px$ and $qv$, respectively. Both types of immune responses become activated by the virus at rates $cv$ and $kv$, and decay in a rank $bx$ and $wx$, respectively. Impairment of both immune responses by the infected virus population is captured in the terms $-wz$ and $-bx$.

In addition, the model also includes a stochastic element because it allows the emergence of new mutants. The probability that a new mutant emerges in the time interval $[t, t+dt]$ is given by $Pv$, where $P$ is the mutation rate. The simplest assumption is that $P$ is constant, but it is more realistic to assume that $P$ is proportional to the...
Fig. 4. Evolution of many different virus strains during the time of infection. Initially, the strains grow at high levels which may cause the clinical symptoms observed during primary HIV infection. The subsequently emerging escape mutants are suppressed at a faster rate, because of the action of cross-reactive immune responses. Different virus strains grow to different levels according to their growth rates. The accumulation of viral diversity breaches the diversity threshold after approximately 7 years in this simulation. In the final phase, the fastest growing strains dominate the virus population. The y-axis indicates the relative concentration of different virus mutants.

1. viral load (v), since the number of mutation events is proportional to the number of replication events.

2. The equations were analysed by a combination of analytic and numerical studies. Three parameters ranges can be distinguished.

3. The rate of viral replication (r) and the degree of immune impairment (a) may outline the combined effect of strain-specific and cross-reactive immune responses. This happens if \( r_a > \lambda \). In this case, the immune response goes extinct and the virus may grow uncontrolled. Hence, no antigenic variation will be observed and the fastest replicating strain will dominate the virus population. This parameter region corresponds to immediate progression to AIDS without the presence of an asymptomatic period.

4. If \( r_a < \lambda \), the cross-reactive immune response (\( \lambda \)) is sufficiently large compared to virus replication (r) and the degree of immune impairment (a), leading to stable control of virus population by the immune system.

5. The levels of virus load under the control of the immune response depend on the rate of viral replication, the amount of immune impairment induced by the virus, the efficacy of the immune response, as well as on the number of antigenic variants. Although an increase in the number of antigenic variants leads to an increase in virus load, uncontrolled virus replication is not observed in this parameter region. This may correspond to chronic HIV infection without development of disease. Although some HIV-infected patients are long-term non-progressors, it is not clear whether they will develop symptoms at a later stage.

An example of chronic infection without development of disease are African green monkeys infected with SIVmac, the natural strain for this host [35]. The lack of disease progression in this case may be due to stronger cross-reactive immune responses as well as slower replication rates of the virus.

3. If \( \lambda > r_a > \lambda \), the combination of cross-reactive and strain-specific immune responses may control the virus, but the strain-specific response alone is unable to do so. In this case, the dynamics depend on the amount of viral diversity. If antigenic diversity is low, the virus is controlled, corresponding to the asymptomatic phase of the infection. Increasing viral diversity results in an increase in viral load. The immune system keeps the virus population in check as long as viral diversity lies below a diversity threshold. Crossing the diversity threshold leads to a collapse of the immune response accompanied by uncontrolled virus growth with the fastest replicating strain evolving to dominate the virus population. These dynamics are illustrated in Fig. 4. Viral diversity can be captured in the Simpson's index giving the probability that two virus strains sampled at random belong to the same strain. It is defined as \( D = (v_i / v) \). It is a number between zero and one, with lower values of \( D \) indicating higher diversity. The diversity threshold can thus be defined as \( D < (r_a - \lambda a) / (r \lambda) \).

This behaviour may correspond to the typical course of disease progression observed in HIV-infected patients. As the beginning when viral diversity is low, the patient enters the asymptomatic phase in which the virus is controlled by the immune response. As diversity increases in response to selection pressure exerted by the immune system, virus load slowly increases. When the diversity threshold is crossed, the patient progresses to AIDS which eventually results in death. The amount of diversity required for the transition to AIDS depends on the efficiency
These models were formulated specifically to look at the dynamics of CTL (Fig. 5). HIV has many epitopes which CTLs may specifically recognize. Thus, there are several CTL responses, each specific for one of the HIV epitopes, fighting the same virus population. This is analogous to several species of predators (CTLs) competing for the same prey (virus) population (Fig. 6). The implications of these assumptions for the dynamics between HIV and the CTL response have been examined by mathematical models [37–39]. Consider the existence of two CTL epitopes, A and B. Since the virus can mutate to change the sequence of the epitope, we assume that there are \( i = 1, ..., n \) sequences in epitope A and \( j = 1, ..., m \) sequences in epitope B. With this in mind, a model with three types of variables may be constructed. They include: virus variants with sequence \( i \) in epitope A and sequence \( j \) in epitope B (\( x_{ij} \)); CTLs directed against sequence \( i \) in epitope A (\( y_i \)); and CTLs directed against sequence \( j \) in epitope B (\( y_j \)). The model is given by the following set of differential equations:

\[
\begin{align*}
\frac{dx_{ij}}{dt} &= v_i (x_{ij} - v_{ij} - g_{ij}) \\
\frac{dx_{ij}}{dt} &= x_{ij} + g_{ij} \frac{y_i}{y_i + \lambda} \\
\frac{dy_i}{dt} &= \gamma_i (y_i - y_i) \\
\frac{dy_j}{dt} &= \gamma_j (y_j - y_j)
\end{align*}
\]

where \( v_i = v_{ij} \) and \( v_j = v_{ij} \).

The virus variants \( x_{ij} \) replicate at a rate \( v_i \) and are killed by the CTLs at the rate \( v_i y_i \) and \( g_{ij} \). The CTLs are stimulated by their respective epitope \( i \). CTLs may either become activated from a precursor population at a rate \( v_i \frac{y_i}{y_i + \lambda} \) or the activated cells may proliferate in response to antigen at a rate \( v_i \frac{y_i}{y_i + \lambda} \). The parameter \( \gamma_i \) is thought to be relatively small, since proliferation rather than activation is the dominant response of CTLs upon antigenic stimulation. Finally, CTLs die at a rate \( \gamma_i \).

Two important parameters are \( \lambda \) and \( \gamma_i \), describing the immunogenicity of sequence \( i \) in epitope A and sequence \( j \) in epitope B.

Nowak et al. [37] analyzed two scenarios: either the virus population is antigenically homogeneous, i.e., all sequences in the respective epitopes are the same, or the virus population is antigenically heterogeneous.

For an antigenically homogeneous virus population, the model predicts that only the CTL response directed against the epitope with the strongest immunogenicity will survive. This is because the CTL response directed against the epitope with the highest immunogenicity is the most superior competitor and will reduce virus load to low levels which are not sufficient to maintain the alternative CTL responses. This is analogous to the competitive exclusion principle in ecology [40]. These dynamics may explain the concept of 'immunodominance', which denotes the observation that in a patient the CTL response is predominantly directed against one or only a few epitopes, although many epitopes may potentially be recognized by CTLs [41–45].

On the other hand, if the virus population is antigenically heterogeneous, the outcome depends on whether the antigenic variants also differ in their replication kinetics. If the antigenic variants do not differ in their replication rates, the model again predicts the occurrence of im-
Antigenic variation in multiple epitopes leads to complicated dynamics which have been termed ‘antigenic oscillations’ [37] (Fig. 7), consisting of distinct peaks in viral abundance, often dominated by a single genotype, which arise when the CTL response against a given variant declines to low levels due to temporary lack of stimulation. This leads to concomitant fluctuations in immunodominance. In contrast to the occurrence of peaks in viral abundance upon the emergence of new variants (antigenic drift), antigenic oscillations are the result of the non-linear dynamics between the existing heterogeneous virus population and CTL responses against multiple epitopes and do not require the emergence of new mutants.

Considering the emergence of a new mutant in a given epitope, the model provides valuable insights, especially for understanding disease progression in HIV-infected patients. Suppose the existence of a homogeneous virus population and an immunodominant CTL response...
Antigenic variation may shift immunodominance. The emergence of an escape variant in an immunodominant epitope (A) leads to one of four possible outcomes. Details are real against one of two epitopes (A, Fig. 8). The emergence of an escape mutant in epitope A may lead to one of four possible outcomes, depending on the replication rates and the immunogenicities of the mutant relative to the wild-type (Fig. 8). Denote the replication rate of wild-type and mutant as $r_w$ and $r_m$, respectively, and denote the immunogenicity of wild-type and mutant epitope A as $c_w$ and $c_m$, respectively. The immunogenicity of epitope B is described by $k$. The four outcomes are as follows: (1) A new specific response against epitope A is induced by the mutant without affecting epitope B. This represents diversification in epitope A and will be observed if $1/c_w + 1/c_m > 1/k$ and $r_w < r_m$. (ii) No new response against epitope A is induced by the mutant, but the response against epitope B is enhanced. This is a partial shift in immunodominance and will occur if $1/c_w + 1/c_m < 1/k$ and $r_w > r_m$. (iii) The mutant may induce a new response against epitope A which outcompetes the response against the wild-type, consequently causing a partial shift in immunodominance. The condition for this outcome is given by $1/c_w + 1/c_m > 1/k$ and $r_w < r_m$. (iv) Finally, if $1/c_w > 1/k$ and $r_w < r_m$, the mutant virus outcompetes the wild-type, resulting in a complete shift in immunodominance. Thereby, the selective advantage of the escape mutant becomes negligible and it may not reach fixation even if there is no CTL response against the variant peptide.

In terms of HIV infection, this analysis extends and reinforces the idea that viral evolution towards increased antigenic diversity may drive progression of the disease. A stable response against an invariant epitope will lead to efficient control of the virus and slow progression. Viral diversity may lead to unstable dynamics, and the evolution of new antigenic variants may result in a shift in immunodominance to weaker epitopes leading to less efficient control of HIV and an increase in virus load. These results demonstrate that "multiple escape theory" has been confirmed experimentally by recent papers [1, 19, 20].

**Conclusion**

This review has shown how mathematical models may help us understand the way in which viral evolution may contribute to the progression of the asymptomatic period of the infection to the development of AIDS. In particular, evolution towards increased viral diversity may induce the destruction of the immune system. These findings also have implications for vaccination strategies to delay or prevent the onset of AIDS: Nowak and McLean [46] showed that the average number of escape mutants produced by a virus strain may be suppressed below one for the vaccine to be successful. This goal will only be achieved if the immune response against a sufficiently large number of strains is boosted, no matter how immunogenic the vaccine is. Therefore, the chances of successful vaccination will be maximized if a cross-reactive immune response is boosted [13, 46]. In addition, because of the competition occurring between CTLs against different epitopes, it would be advisable to boost the immune response against a single conserved epitope, even if this is not the dominant one, since this would affect more stable and effective control of the virus population [37, 39].

**References**

6. Wei H-P, Ghosh SE, Taylor ME, Johnson VA, Emini EA,
Wodarz and Nowak. Dynamics of immune escape in HIV infection


antigenic site after intramolecular selection of T-cell determinants. J Immunol 151: 1852-1858

Correspondence: Martin A. Nowak, Institute for Advanced Study, Olden Lane, Princeton, NJ 08540, U.S.A., e-mail: nowakian@so