# Evolutionary and immunological implications of contemporary HIV-1 variation

**Bette Korber*,†, Brian Gaschen*, Karina Yusim*,†, Rama Thakallapally*, Can Kesmir†,‡** and **Vincent Detours*,†**

**Division of Theoretical Biology and Biophysics, Los Alamos National Laboratory, Los Alamos, New Mexico, USA*
*†The Santa Fe Institute, Santa Fe, New Mexico, USA*
*‡Center for Biological Sequence Analysis, Department of Biotechnology, Technical University of Denmark, Denmark*

Evolutionary modelling studies indicate less than a century has passed since the most recent common ancestor of the HIV-1 pandemic strains and, in that time frame, an extraordinarily diverse viral population has developed. HIV-1 employs a multitude of schemes to generate variants: accumulation of base substitutions, insertions and deletions, addition and loss of glycosylation sites in the envelope protein, and recombination. A comparison between HIV and influenza virus illustrates the extraordinary scale of HIV variation, and underscores the importance of exploring innovative HIV vaccine strategies. Deeper understanding of the implications of variation for both antibody and T-cell responses may help in the effort to rationally design vaccines that stimulate broad cross-reactivity. The impact of HIV-1 variation on host immune response is reviewed in this context.

The natural variability of HIV-1 provides a framework for understanding the complex biology of the virus. As we strive to understand immune evasion and drug resistance, our fundamental knowledge of mutational processes and selection is expanded. Phylogenetic analysis of variable forms of HIV can provide a glimpse into the evolutionary and epidemiological history of the virus. Recent studies have exploited advances both in sequencing technology and in modelling methods. Interpretation of global data sets through the application of improved analysis strategies enables researchers to trace the roots of HIV in terms of primate lentiviral evolution, to estimate divergence rates and the timing of the most recent common ancestor of the epidemic strains, and to model demographic trends in the epidemic. Simultaneously, new experimental methods in immunology[1] have rapidly expanded our understanding of the host's response to the virus and the consequences of immune pressure. Scientists are now poised to conduct large scale population studies combining viral sequencing, host genetics, and immunology.

HIV-1's great diversity is seeded by the lack of a proof-reading mechanism in RNA viral polymerase reverse transcriptase, and the

*Correspondence to: Dr Bette Korber, Division of Theoretical Biology and Biophysics, Los Alamos National, Laboratory, Los Alamos, New Mexico 98545, USA*

 © The British Council 2001

consequential high error rate (0.2–2 mutations per genome per cycle)[2]; but that is not the full story. For example, HTLV-I, like HIV-1, infects $CD4^+$ T-cells and goes though a reverse transcription step, but is far less variable; this may be attributed to differences in the dynamics of the two viruses[3]. A high replication rate accompanied by rapid viral turn-over[4], as well as pressure for change coupled with tolerance of change must be part of the story. Beyond base-substitution, HIV is subject to recombination and relatively large insertions and deletions (indels), which rapidly generate radically divergent forms.

An HIV infection starts out with a homogeneous viral population[5,6]; over the course of a typical infection, viruses which have mutated to alter more than 10% of their DNA bases in the envelope gene arise[6–8]. The concept of a viral quasispecies was originally introduced to model within-host viral populations and assumes that the frequency of viral forms in the population will be dictated by their relative fitness; the quasispecies is the realization of the distribution of forms within the sequence space[9]. However, the term quasispecies has come to be used more loosely in the HIV literature to simply refer to the set of viruses found in an infected individual. Under circumstances of selective pressure, such as therapy[10] or immune pressure[11–13], the frequency of forms in the viral population can shift. An 'archive' of earlier forms of the virus is retained in proviral DNA and these forms on occasion can re-emerge[6,7,14], presumably under changing circumstances in the host or through recombination. Combination therapy is effective despite within-host variation because multiple mutations required for resistance to three or more drugs are difficult to achieve unless selected for sequentially[15]. The correlation between long asymptomatic periods and the breadth of the T-cell immune response[16], and the association of HLA-homozygosity with accelerated progression to AIDS[17], may be manifestations of this same phenomenon, as simultaneous evasion of multiple immune responses may be difficult for the virus to achieve.

While immune escape clearly can drive positive selection in specific HIV epitopes[11,13,18,19], neutral mutations and genetic drift can also contribute to the overall diversity of the virus[20,21]. One strategy used to try to understand the relative importance of positive selection through immune evasion in HIV evolution is to quantify the force of positive selection within a given gene through analyzing the ratio of synonymous to non-synonymous substitutions[22,23]. Interpretation of average values is complicated because positive selection can be limited to narrow immunogenic domains within proteins, and average values of these ratios for full genes can be misleading. Many forces concurrently influence the evolutionary pattern *in vivo*, such as counterbalancing influences of retention of protein structure and function. Hence, there has been a recent interest in estimating these parameters for each

site[21,24–26]. It will be important to establish how mutations which arise in the context of immune evasion within a single host can influence HIV diversity in populations. Such information will help define what is required at a molecular level for stimulating broadly cross-reactive immune responses.

HIV-1 subtypes are distinct lineages of the virus that are defined on the basis of genetic distances and phylogenetic clustering and are labelled A–K (Fig. 1f). It remains unclear whether these genetically defined subtypes will provide a useful way to consider global or regional variation for vaccine design. Neutralizing antibody serotypes do not correlate well with genetically defined subtypes[27,28]; the antigenic domains that define reactive sites are generally cross-reactive with a subset of viruses, and both cross-reactive and non-reactive viral forms can be found within the same subtype and within different subtypes. Cytotoxic T-cell (CTL) epitopes show a spectrum of cross-reactive responses to epitopes derived from proteins of different subtypes than the stimulation strain[29–34].

In this paper, we will consider the impact of variation in several ways. The first section concerns what can be learned from viral genetic diversity and phylogenetics about the history of the epidemic and rate of evolution. The second section is a summary of the extent of HIV variation as it is known today. The third briefly reviews the influence of variation on design and selection of the annual influenza vaccine (Fig. 2), and contrasts influenza and contemporary HIV-1 variation (Fig. 1). The influenza vaccine is often suggested to be a model system for HIV; however, the variability issues facing those involved with influenza vaccine design, while scientifically challenging and complex, are fundamentally distinct from the issues that need to be faced for an HIV vaccine. Finally, in the last section, the relationship between subtypes and known CTL epitopes is discussed, and the potential impact of variation on epitope processing is modelled.

## The origins of contemporary HIV-1 diversity

### *Primate lentiviruses*

Lentiviruses have been found in many African primate species[35], and the phylogenetic relationships between the viruses have been used to study the primate origins of HIVs[36]. HIV-2, the distant cousin of HIV-1 that also causes AIDS in humans, is most closely related to simian immunodeficiency viruses (SIVs) found in sooty mangabeys[37,38], while HIV-1 is most closely related to viruses that have been isolated from chimpanzees[36,39,40]. HIV-1 has three distantly related groups: (i) the main group (M), the group of viruses that cause the global pandemic; (ii) O,

a form that is found with a low prevalence in west central Africa and has also been found in Europe[41,42]; and (iii) N, a very rare form found in Africa[43]. There are only a small number of viral sequences available from chimpanzees, and they are very diverse[36]; more information is needed before the precise relationships between chimpanzee viruses and HIV-1 M, N, and O groups can be elucidated.

### *Modelling the history of the HIV epidemic through phylogenetics*

The earliest human sample found to contain HIV-1 was taken in 1959, and found among stored samples from the city now known as Kinshasa in the Democratic Republic of the Congo[44]. Phylogenetic analysis was used to validate the sample; it behaved like an older sequence would be expected to behave in terms of its position in the tree, branching closer to the root than modern sequences[44]. So the epidemic had its origins sometime prior to 1959. The lack of hard data prior to 1959 makes this a nebulous period, but important in the human history of HIV. Assuming the evolutionary behaviour of the virus is consistent, one can extrapolate into this pre-1959 period by estimating the rate of diversification using modern isolates with known dates of sampling. This allows a projection based on genetic distances and time, that can be used to estimate the most recent common ancestor of a given viral lineage[14,45,46]. Current estimates point to an origin of the HIV-1 M group near 1930, with error bars that span roughly a decade or so, depending on the method used[14,45]. These methods depend upon models of evolution that have inherent assumptions that are imperfectly met by real data[47,48]. Despite this, the method used in Korber *et al*[14] was validated through accurately estimating the timing of two control time points: the 1959 sample, and the beginning of the epidemic in Thailand. Salemi *et al*[45] used a strategy that excluded the most variable positions, and used a hepatitis sequence set as their control.

These analyses provide information about the rate of evolution, and also the rate of expansion of the epidemic. Dating the most recent ancestral sequence of the pandemic strains does not, however, decisively show when the first M group ancestral virus infected a human. There could have been an ancestral virus in chimpanzee with multiple subsequent cross-over events between species, or a period of time during which the virus was carried in humans prior to the expansion of the M group. If the evolution of the HIV-1 M group took place in humans, then the time of the most recent common ancestor of the M group would be an upper bound on the cross-species transmission event that ultimately gave rise to the epidemic.

Coalescent theory provides a way to extend the boundaries of what can be learned through phylogenetic studies, and model the demographic

history of an epidemic through the phylogenetic branching pattern in a tree[49–52]. But to move beyond the phylogenetic tree into modelling estimations of past epidemic growth, additional assumptions are required. Current coalescent methods are based on the assumptions that: (i) the variance in the reproductive rate of the virus is constant in time (the reproductive rate is related to the number of new infections that arise from a single infection); (ii) the available sequences are sampled randomly; and (iii) the population is panmictic, *i.e.* the virus is circulating within a population that is not subdividing into isolated lineages and that there is no immigration of new forms into the population. Allowing for these assumptions, again imperfectly met by real data (although clearly some data sets are in better accordance with the assumptions than others), one can model the expansion of the effective population size through the history of the epidemic lineages represented in a tree. Coalescent theory has been applied to the study of HIV at many levels: within a person[50], comparisons of subtypes[49,51], and to the evolution of the M group virus in the Democratic Republic of the Congo (DRC)[52]. The DRC set is the most diverse set of HIV-1 M group sequences currently available from any nation, and is considered to be from a region of Africa which harboured the virus for much of its history in humans[53]. Analysis of this set of central African viruses suggested a period of slow expansion early in the epidemic, with more rapid expansion in recent decades[52], consistent with the timing estimates that suggest the virus was in the human population for many decades prior to AIDS being detected and defined[14,45].

## The extent of contemporary HIV-1 diversity

### *HIV subtypes*

HIV-1 genetic subtypes have a certain arbitrariness in terms of their definition, and biologists can (and do) reasonably argue for different nomenclature schemes. Nonetheless, a standard nomenclature system is important, as it provides a common language for referring to related lineages and captures a fundamental feature of the virus: the gene and protein sequences within a subtype are more closely related to one another than to the HIV genes and proteins from other subtypes. The associations and groupings of the subtypes can be statistically validated through phylogenetic analysis (Fig. 1f)[54–56]. For the sake of uniformity in the field and to facilitate communication, a team of experts recently assembled to update the nomenclature system[57,58]. The genetically distinct subtypes are labelled A–K (with no subtype E or I). Circulating recombinant forms (CRFs) describe viral genomes that contain clearly

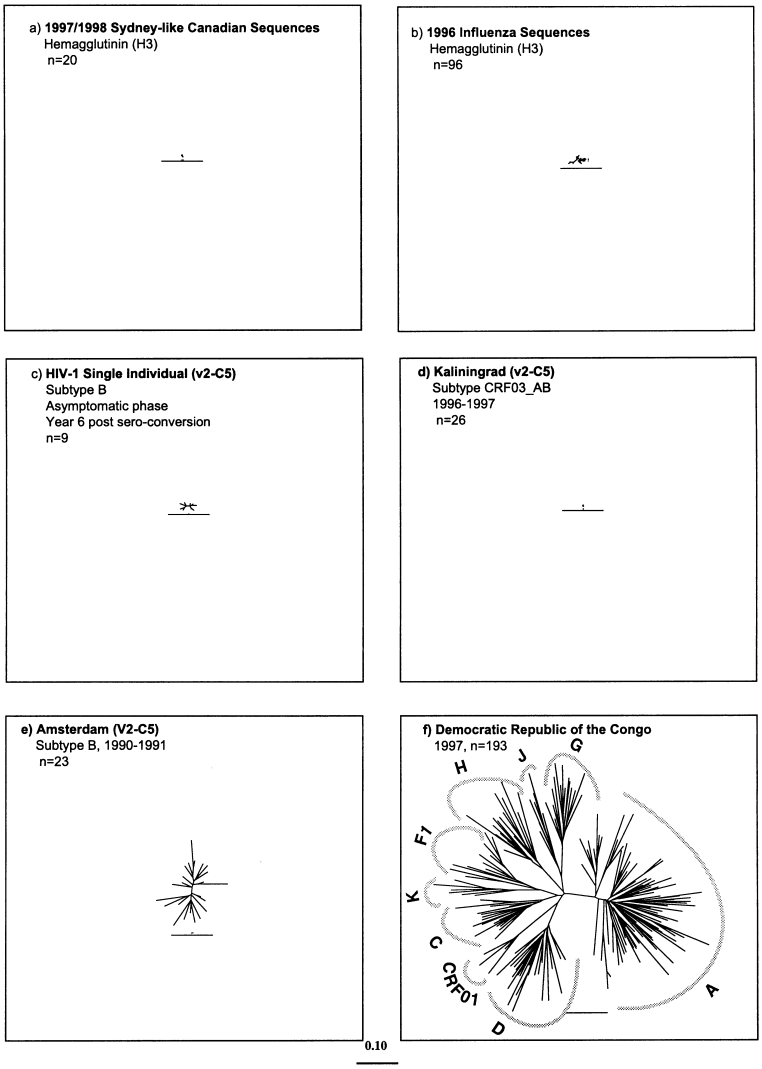


**Fig. 1** A comparison of evolutionary distances of HIV-1 envelope sequences encoding V2-C5 and influenza HA1 domain of the HA gene through phylogenetic analysis. All panels show maximum likelihood trees using a REV model allowing for rate variation at different sites, following the strategy described in Korber *et al*[14]. The scale bar is the same in a–f for comparisons. (a) Tree based on 20 HA1 domain sequences of A/Sydney-like viruses circulating in Canada during the first half of the 1997–1998 flu season (Osiowy CK, unpublished observations. Accession numbers: AF087700-AF087708 and AF096306-AF096316). (b) Tree based on 96 HA1 domain sequences of human influenza H3N2 viruses. The tree contains all sequences from the Influenza Sequence Database, Los Alamos National Laboratory[131], with an isolation year of 1996. (c) Tree based on 9 V2-C5 sequences from a single asymptomatic individual collected at one time point 73 months post-seroconversion – this was a subtype B infection, and is typical of intrapatient diversity[7]. (d) Tree based on HIV-1 subtype CRF03_AB V2-C5 sequences from 26 individuals from Kaliningrad, representing a unique situation where a recombinant form of the virus spread explosively through a population of i.v. drug users, and all viruses were very closely related to a single common ancestor[65]. These samples were collected during 1997–1998, within a year of the introduction of the strain into the population. (e) Tree based on HIV-1 V2-C5 env sequences from a subtype B epidemic, sampled from 23 individuals residing in Amsterdam in 1990–1991[117]. (f) Tree based on HIV-1 V2-C5 sequences sampled in 1997 from 193 individuals residing in the Democratic Republic of the Congo (DRC), a remarkably diverse set[53].

delineated sections derived from different subtypes, that share a common ancestor, and that are the basis of multiple infections. CRFs are thus epidemic strains, which, like subtypes, are of global importance. There are currently 11 defined CRFs, and more are in the pipeline. Some of the CRFs infect great numbers of people: CRF01(AE) in Asia (originally identified in Thailand and called subtype E[59,60]), CRF02(AG), with a prototype isolate IbNg, found throughout western and central Africa[61–63], a newly defined subtype B and C CRF common in China[64], and a subtype A and B recombinant form found in Russia[65]. Given the exquisite specificity of immune receptors, the genetic relatedness of subtypes and CRFs viruses will certainly have some immunological consequences, but the implications for vaccine design are unclear.

More extensive sampling in regions of sub-Saharan Africa with great viral diversity has resulted in ever greater indications of the potential complexity of HIV diversity. In regions where multiple subtypes are co-circulating with a high prevalence, intersubtype recombination is common[61] and recombination between recombinants has also been observed. The large number of novel recombinants found suggests that multiple infections of HIV are not uncommon. Several new isolated examples of strains that are not clearly related to any defined subtype, and that are not obviously recombinants, have also been found. The subtypes themselves are growing ever more diverse with time. The combination of these factors results in a massive pressure on the current nomenclature system, and probably will eventually result in a breakdown of our current HIV classification scheme.

### *Measures of diversity*

There are many ways to compare and contrast the extent of diversity found in HIV-1, and the individual proteins show different levels and patterns of variation. One very basic measure is a tally of sites that are preserved throughout all of the M group, of sites that tolerate insertions and deletions, and of sites that are variable because of nucleotide substitutions. In an alignment of the 132 nearly full-length M group sequences publicly available in the HIV database alignment[66], spanning a region from the start of *gag* to the end of *env*, there are a total of 7667 bases in HXB2. HXB2 is a sequence derived from the first HIV-1 isolate[67] and is a standard reference strain. Roughly 13% (992) of the positions in HXB2 have deletions in one or more of the other 131 HIV-M group sequences, leaving 6675 positions in HXB2 that do not contain any deletions in the alignment. Of these 6675 alignable positions, 2301 bases (30% of the total 7667 bases) are invariant, and 4374 (57%) of the positions have substitutions.

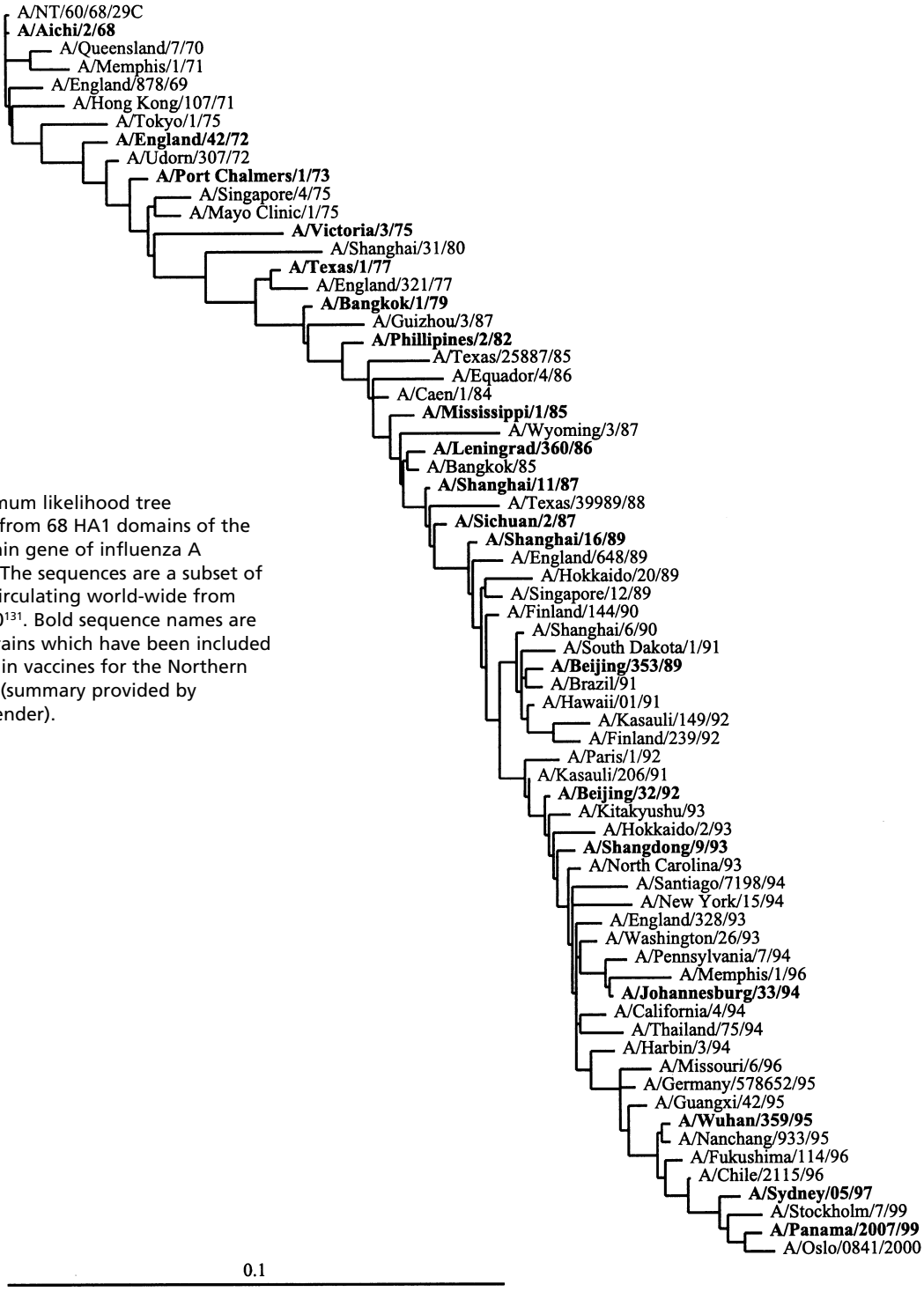


**Fig. 2** Maximum likelihood tree constructed from 68 HA1 domains of the hemagglutinin gene of influenza A subtype H3. The sequences are a subset of the viruses circulating world-wide from 1968 to 2000[131]. Bold sequence names are the H3N2 strains which have been included sequentially in vaccines for the Northern Hemisphere (summary provided by Catherine Bender).

Another way to consider variation is distinctive lengths and glycosylation patterns found in envelope proteins. Among the current database collection (December, 2000) of 566 full-length gp120 protein sequences, protein lengths vary from 484 to 543 amino acids because of the insertions and deletions found in hypervariable regions. The hypervariable regions are typically excluded from phylogenetic analysis, as insertions and deletions do not follow the evolutionary base substitution models used to determine phylogenetic relationships. Thus the branch lengths in the *env*-based trees are a profound underestimate of the true diversity of the protein, because they capture only one aspect of the variation. Many of the hypervariable sites and insertions and deletions involve glycosylation sequons (the amino acids NX[S or T], sites where N-linked glycosylation occurs). The number of N-linked glycosylation sites in gp120 ranges from 18–33, with a median value of 25. Glycosylation can reduce accessibility to neutralizing antibody epitopes[68].

Considering relative HIV variation by protein can also be informative. Using the full-length genome alignment, so a fair comparison between genes can be made, one can model the relative rate of variation between all sites in an alignment using maximum likelihood phylogenies, and then build more accurate phylogenies by including site-specific rate variation[69]. A gamma distribution was used to assign one of seven evolutionary rates to each site in the nearly full-length HIV-1 M group alignment (the 6675 positions with no insertions or deletions described above). The sites that were evolving most slowly were assigned to category 1, and the fastest were assigned to category 7; in this model, category 7 was estimated to be evolving 20 times faster than category 1. For simplicity, just four HIV-1 genes are shown in Figure 3, representative of the spectrum of evolutionary rates found in HIV-1: gp120, the most rapidly evolving protein; tat; and the gag genes encoding p17, and p24. Gag p24, like pol, is highly conserved. We also used the program SNAP[24] to measure the non-synonymous substitution rate for genes included in the full-length genome alignment. In this case, we were not trying to calculate the ratio of non-synonymous to synonymous substitution within genes, but the relative rates of non-synonymous substitution between genes in the 132 full-length sequences. Tat in particular has overlapping coding regions in different frames, and its synonymous substitution rates are low and strongly biased by this. But, as the full-length genome alignment was used to examine the different coding regions, the non-synonymous rates for the different genes can be compared directly. The highest rate of non-synonymous substitutions (dn) was found in Tat, lowest in gp24 (dn values: Tat, 0.170 ± 0.022; gp120, 0.130 ± 0.015; p17, 0.104 ± 0.180; and p24, 0.039 ± 0.010). The high rate of substitution found in Tat is particularly important in the light of recent interest in Tat as a early and potent CTL target[13].

## Influenza vaccines: responding to change

### *Influenza diversity and vaccine design*

Influenza A viruses infect many mammalian and avian hosts, including duck, swine, human, geese, quail, whale, seal, and chicken. The viral genome has eight segments and is categorized by the serological and genetic characteristics of its two surface glycoproteins, the neuraminidase gene (NA) and the hemagglutinin gene (HA)[70]. To date, 15 subtypes of the HA gene and 9 subtypes of the NA gene have been isolated from mammalian and avian hosts; all 15 HA genes occur naturally within the

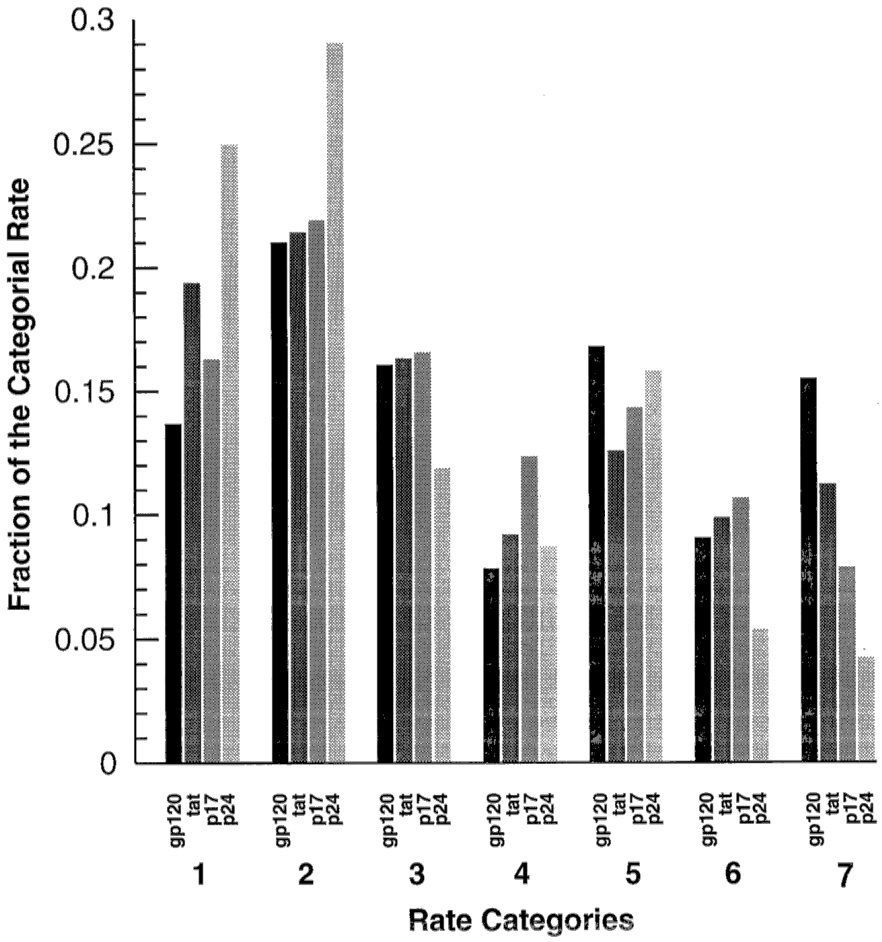


**Fig. 3** Histogram showing frequency of mutation rate categories assigned to different sites in HIV-1 proteins coding regions, with category one being the slowest evolving sites, category seven the fastest. The rates were calculated based on a full genome maximum likelihood tree, excluding recombinant sequences, using a gamma distribution, and the frequency of assigned rates within gp120, tat, p17 and p24 are shown.

avian population[71]. Two kinds of viral evolution have been observed in human hosts over the last century, which have been termed antigenic drift and antigenic shift. Antigenic drift is the gradual change in the nucleotide sequence which is due to point mutations accumulating over time. Normally, changes made in the influenza vaccine are made in response to the antigenic drift of the virus. Antigenic shift, on the other hand, is an abrupt change in the serological and genetic properties of the virus due to the introduction of new HA or NA subtypes into previously unexposed populations. Since 1900, four HA genes (H1, H3, H5 and H9) and two NA (N1 and N2) have been isolated from human hosts[72–75]. Currently, two influenza A subtypes are found co-circulating in the human population, H3N2, a result of an antigenic shift in 1968, and H1N1, a result of a re-introduction of an H1N1 strain in 1977 similar to H1N1 strains circulating around 1950[76,77].

Because of the annual cost and the threat of particularly dangerous world-wide influenza pandemics, the World Health Organization (WHO) global influenza network was established in 1948. The network comprises over 110 laboratories in 80 countries, and four WHO Collaborative Laboratories[78]. Both unusual and representative samples of influenza virus are sent by the national laboratories year-round to a Collaborative Laboratory where antigenic and molecular analyses are performed. New viruses are antigenically screened for cross-reactivity with the current vaccine strains by the HA inhibition assay (HI), using antiserum stimulated by the vaccine. Members of the WHO network meet to recommend strains for inclusion in the vaccine for the coming year, currently two influenza A strains (H3N2 and H1N1) and one B strain. Recommendations are made based on the antigenic and molecular analyses of recent isolates, epidemiological data, and post-vaccination serological studies in humans[79]. In the years spanning 1968–2001, the H3N2 component of the influenza vaccine was changed a total of 17 times (Fig. 2), in one case necessitated by antigenic drift of as little as 3 amino acids substitutions in the 307 amino acid HA1 component of HA (A/Shanghai/16/89 to A/Beijing/353/89).

### *Influenza and HIV compared*

Although influenza, like HIV, is a rapidly evolving virus, a comparison of HIV and influenza A evolution reveals very different patterns. HIV evolution is characterized by a radial spread outward from an ancestral node, while influenza is characterized by bottlenecks and global drift from year to year. Within a single 9 month flu season, little variation was typically found between geographically distinct influenza isolates after the emergence of the epidemic strain. HIV, on the other hand, shows increasing genetic diversity within a population through time. A phylogenetic tree of HA sequences

sampled world-wide in 1996 (Fig. 1b) shows much less diversity than a sampling of subtype B HIV-1 envelope sequences from a single city, Amsterdam in 1990–1991 (Fig. 1e). In stark contrast, a sampling of HIV sequences from the Democratic Republic of the Congo in 1997 shows extraordinary diversity, with virtually all HIV subtypes co-circulating in one geographic region (Fig. 1f). The diversity of influenza sequences world-wide in any given year appears to be roughly comparable to the diversity of HIV sequences found within a single infected individual at one time point (Fig. 1c). Thus, while influenza does have a relatively fast rate of mutation when measured over decades (Fig. 2), the vaccine for any given year is targeted towards a relatively homogeneous viral population. Small numbers of changes in the viral amino acid sequence at antigenic sites require a change in the vaccine strain to induce an immunologically appropriate response for currently circulating strains. Thus the diversity which an HIV-1 vaccine must counter through stimulating a broadly reactive immune response is far greater than the diversity countered by the influenza vaccine. If current evolutionary trends continue, the situation will only become worse from an HIV vaccine design standpoint. This is daunting when considered in the context of the small number of amino acid substitutions that result in a loss of antigenic cross-reactivity sufficient to diminish influenza vaccine efficacy (Fig. 2). Thus innovative approaches may be critical for success with HIV-1 vaccines.

## Immunological responses to a rapidly evolving pathogen

### *Strategies for eliciting cross-reactive B-cell responses*

Three broadly cross-reactive neutralizing HIV-1 Env-directed monoclonal antibodies have been found (2F5, IgG1b12, and 2G12)[80,81]. These bind to conserved parts of Env essential for viral entry, so are promising as vaccine targets. The epitopes tend to be poorly revealed, best exposed during key transitional periods, and the antibodies have high affinity for the native trimer suggesting that they were raised in response to the oligomer on the virion surface rather than dissociated subunits[82]. This contrasts with the most exposed regions on Env gp120, the V3 and C5; antibodies directed at these sites tend to have weak cross-reactive neutralizing ability[83]. This could explain the limited neutralizing antibody responses elicited by gp120 vaccines[84,85]. Novel strategies for designing HIV Env in open conformations, intended to expose conserved critical regions, are being explored, with the hope that such reagents may elicit cross-reactive neutralizing responses[81]. Examples of such strategies include: (i) disulphide linked gp120–gp41 that mimics a native oligomeric conformation of Env[82]; (ii) glycosylation-deficient forms of Env that leave neutralizing epitopes exposed and increase their antigenicity[68]; (iii) Env linked to other immunogenic proteins[86]; (iv)

locking Env into a conformation adopted part way through the fusion process[87,88]; and (v) deleting variable loops to expose key epitopes[89].

### *CTL responses in HIV vaccines*

Several lines of evidence support the notion that CTL contribute to the control of natural infections. HIV CTL have been found in a significant fraction of HIV-1 seronegative high-risk sexual partners of HIV-1 positive individuals[90,91]. In particular, HIV-specific gamma-interferon secreting cells were detected in cervical mucosa and vaginal washes of exposed seronegative women[92,93], suggesting protective immune responses may occur in women chronically exposed to HIV-1, in whom HIV infection cannot be detected virologically or clinically. Broadly cross-reactive CTL were found in highly-exposed seronegative sex-workers in two regions in Africa[94], suggesting that broadly cross-reactive CTLs provide protection. Strong CTL responses and T-helper responses are also associated with long-term survival and non-progression in HIV-1 human[95,96] and SIV macaque infections[97]. Some human HLA types, including HLA B51[98], are associated with slow disease progression; HLA B51 presents epitopes located in highly conserved regions of HIV-1[99]. Finally, CTL escape mutations have been noted to arise in conjunction with progression to AIDS[100].

### *CTL cross-reactivity and implications for vaccine design*

Various degrees of CTL cross-subtype reactivity have been documented, including examples of complete cross-reactivity, diminished cross-subtype responses, as well as cases where the response is completely abrogated[30–34,94,101,102]. Different target proteins can yield different experimental results in terms of CTL responses to proteins[103]. CTL experiments often probe the immune response with a single variant, and may favour the detection of responses to conserved epitopes – in other words, a CTL response could be experimentally silent if the probe was sufficiently different from the infecting and stimulating strain in a CTL epitope, and such differences are more likely to occur in highly variable regions. CTL specific for early expressed proteins like Tat may be particularly important for controlling viral replication[13,104], but there are few defined HIV-1 Tat specific CTL responses[105]. Tat responses may be hard to detect[106] due to the variability of the Tat protein (see above) and the potential for rapid escape[13]. It is not clear what the practical implications of viral variation will be for CTL-stimulating vaccines, and how the complexity of host genetics and HLA alleles will factor-in, but people are beginning to attempt to address these issues[107–109].

Defining an optimum vaccine to stimulate protective CTL responses against the breadth of circulating viruses in a complex host population is a challenging problem, but DNA vaccines allow great flexibility[110,111], and so have great potential for rational design. The biology, however, can be complex. For example, even in the most parallel of infections, identical twins infected with the same batch of factor VIII, profound differences in immune escape and immunodominant CTL responses were observed[19]. This is a humbling result. If it is characteristic of host-viral interactions, then rational design of a vaccine based on discerning common patterns of response in diverse individuals will be very difficult. One very basic vaccine approach that may be broad enough to over-ride such issues is simply the inclusion of multiple whole HIV-1 protein coding regions to maximize the potential breadth of the response[112,113]. Another strategy is to concatenate DNA encoding epitopes tailored to maximize potential responses for a population with a particular HLA distribution into a single contiguous DNA-vaccine[107]. Alternatively, one could use the HIV-1 immunology database to define conserved epitope-rich domains[105,108], or to experimentally identify maximally reactive peptide regions[109] – such approaches are not exclusive, and can show significant overlap (Fig. 4). Yet another strategy, not directly addressing cross-reactivity but rather potency (which may in turn enhance cross-reactive potential), explores optimization of vaccination with immuno-modulatory agents in addition to HIV antigens that both strengthen and tailor the nature of the immune response[114–116].

### *Consensus sequences are more central than modern sequences*

Consensus, or most recent common ancestor sequences, might provide a better reference strains than modern viral sequences for designing peptides for immunological testing, and for vaccine preparations. A consensus is 'central', and thus more similar to circulating strains than they are to each other, and can remain relatively stable over time[117]. To provide a sense of the magnitude of the overall differences within and between subtypes, and of the virtues of consensus sequences, diversity was estimated through tallying the number of amino acid substitutions in pairwise comparisons of subtype A and subtype B protein sequence alignments from the year 2000 HIV Sequence Database[66]. Subtypes A and B were selected because numerous well-characterized sequences were available (17 subtype A and 95 B Env sequences; 12 A and 37 B Tat sequences; and 7 A and 35 B Gag sequences are included in the following summary). Positions with insertions and deletions were excluded.

**Comparisons of actual sequences to their subtype consensus**
Median percentage of amino acid differences are 10% (range 7–14%) in Env, 9% (4–14%) in Tat, and 5% in Gag (1–10%).

**Pairwise comparisons of within subtype protein sequences**
Median percentage of amino acid differences are 17% (4–30%) in Env, 15% (3–30%) in Tat, and 8% (2-15%) in Gag. (Note that sequences within a subtype are generally much more distant from one another than they are from the consensus of their subtype.)

**Inter-subtype comparisons between A and B subtype sequences**
Median number of amino acid changes is 25% (20–36%) in Env, 35% (27–42%) in Tat, and 17% (15–22%) in Gag.

Variable regions in all HIV proteins are not evenly distributed, but clustered with patterns that probably reflect conservation of functional domains, regions that tolerate change, and regions that are subject to immune pressure. Still, 11% amino acid substitution roughly translates to an average of about one change per epitope. A detailed summary of the number of substitutions found for actual experimentally defined epitopes between and within subtypes can be found in Korber *et al*[118].

## *Epitope processing and variation*

Viral proteins are not recognized by CTL in their native form, but are cleaved into peptides in the cytosol of infected cells[119,120], transported to the endoplasmic reticulum[120,121], and loaded into a groove on class I HLA molecules[120,122]. Peptide-loaded class I molecules are then expressed on the surface of infected cells, accessible to antigen-specific receptors of CTLs. The class I peptide binding groove accommodates peptides of length 8–11 amino acids, and peptide binding to a particular class I molecule depends most critically upon amino acids known as anchor residues[122].

Virus variability may influence any step in epitope processing[123]. Immune escape driven by mutations in epitope flanking regions has been documented *in vivo*[124,125], showing that escape through processing abrogation can be of immunological significance. Protein cleavage is usually carried out by the proteasome[119] and can be sequence-dependent[119,126]; thus, viral variation might alter cleavage points, and positions with potential to serve as epitope boundaries may differ among viral strains. There are also alternatives to proteasomal cleavage pathways[127], and proteasomal cleavage products can require additional trimming by cytosolic proteases for epitope generation[128]. After cleavage and trimming, the transporter associated with antigen processing (TAP) translocates peptides into the

endoplasmic reticulum for loading onto HLA class I molecules[121]. TAP affinity for peptides also can contribute to epitope selection[129].

### *Epitope clustering and the potential importance of processing*

As noted before, there are regions in proteins in which experimentally defined CTL epitopes cluster (Fig. 4), and the biology underlying this non-uniform distribution is not yet understood. One possible contributing factor is the presence of conserved proteasomal cleavage sites at the boundaries of the epitope rich regions. We explored this possibility through use of an artificial neural network program that predicts C-terminal epitope boundaries (http://www.cbs.dtu.dk/services/NetChop/)[130]. This program correctly predicts 65% of cleavage sites and 85% of non-cleavage sites, using a training set of human peptides eluted from class I proteins, and a window of 19 residues surrounding the putative cleavage site as a basis for the prediction[130]. A value is assigned to each position in a protein, such that higher values indicate sites that are most likely to be cleaved, and to be the C-terminus of an epitope. Predicted cleavage sites for HXB2 (Fig. 4) often coincided with, or are within one or two amino acids of, experimentally-defined C-terminal amino acids of epitope or epitope rich regions. Sites with the highest cleavage site prediction scores vary between HIV-1 sequences and subtypes (Fig. 4).

Detailed examples of the influence of variation on cleavage site predictions are shown in Figure 5. The predicted cleavage scores for positions in the B subtype HXB2 p17 protein sequence are highly correlated with predicted cleavage scores for same position in the B subtype consensus sequence ($R^2 = 0.95$). But for the more variable Nef protein, the correlation between cleavage scores breaks down somewhat ($R^2 = 0.63$), even for these two very closely related sequences. For both p17 and Nef, HXB2 and consensus sequences from other subtypes have very different predicted C-terminal cleavage site scores at most positions. This suggests that cleavage may be subtype-specific and processing might be important. It is worth noting that this effect might go undetected in many of the cross-reactivity experiments that have been conducted to date, as often the optimal peptide is defined, the sequence variants within the optimal peptide are synthesized and tested for recognition, but processing is not tested.

Epitope clustering was apparent from the first compilation of experimentally defined HIV-1 CTL epitopes in the 1995 HIV Immunology Database[105], and has become increasingly evident with each new release of the HIV Immunology Database. This, we believe, might have broad practical implications. One could preferentially exclude from vaccines regions that do not encode known epitopes, and focus instead on

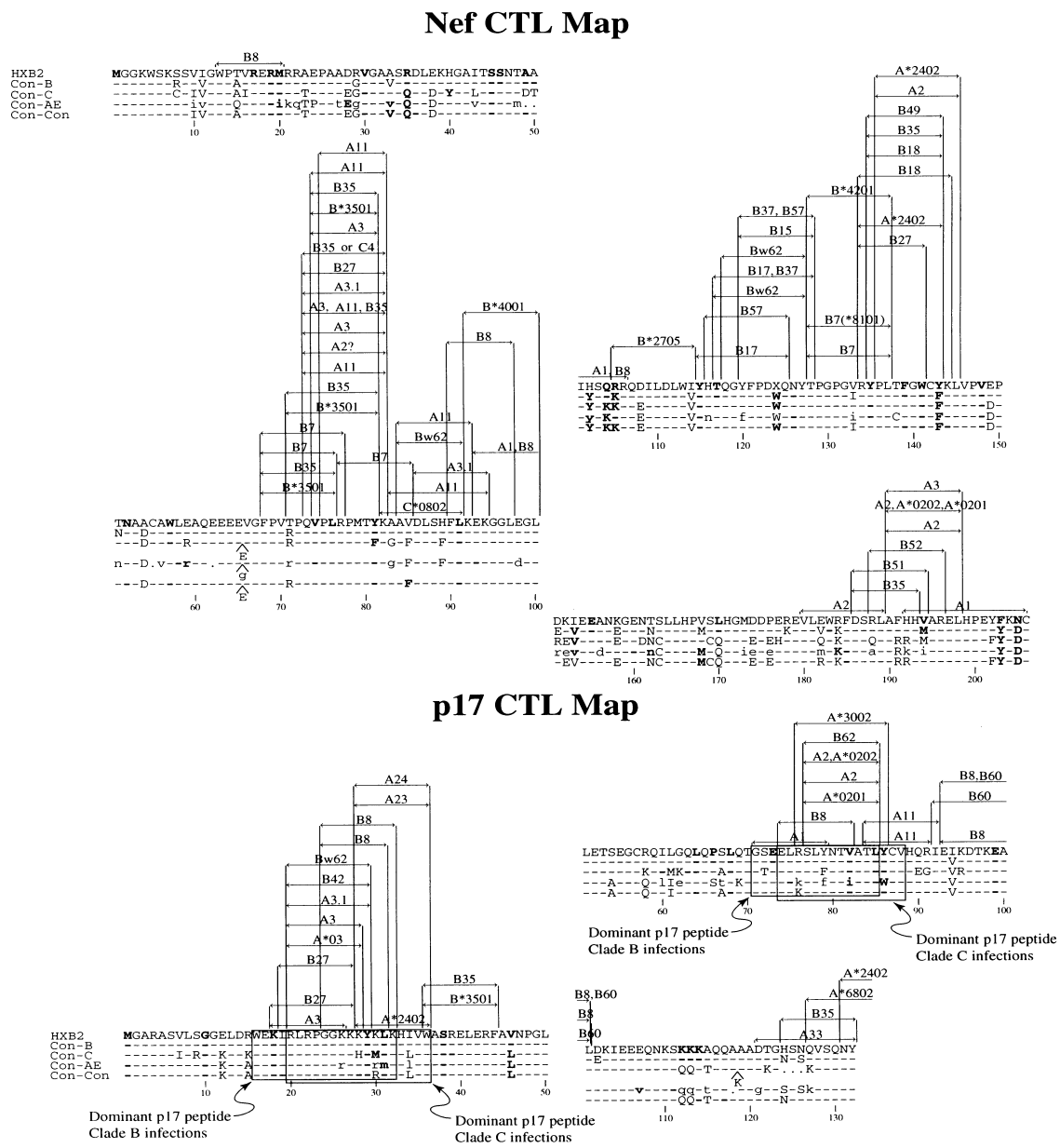


**Fig. 4** HXB2 based maps of experimentally-defined CTL epitopes in p17 and Nef from the HIV immunology compendium (http://hiv-web.lanl.gov/). p17 and Nef have very distinctive epitope clustering patterns. The clustering patterns coincide with peptides that were found to serve as a stimulating antigen for CTL from many individuals (although the frequency depended on the HLA types in the population); the maximally reactive peptides are shown in grey[109]. HLA presenting molecules are shown above delineated epitopes from the database. Alignments show amino acid variation between the HXB2 reference strain and consensus sequences of different subtypes; subtype C and CRF01 were selected as they are important vaccine candidates. Predicted C-terminal proteasome cleavage sites are marked as bold letters, and differ between subtypes. Note that subtype consensus sequences are more closely related to each other than would be modern circulating strains.

epitope dense regions. Such constructs may produce safer vaccines than coding regions of full proteins, while preserving essentially full immunological potential for CTL. Shorter gene regions might also allow

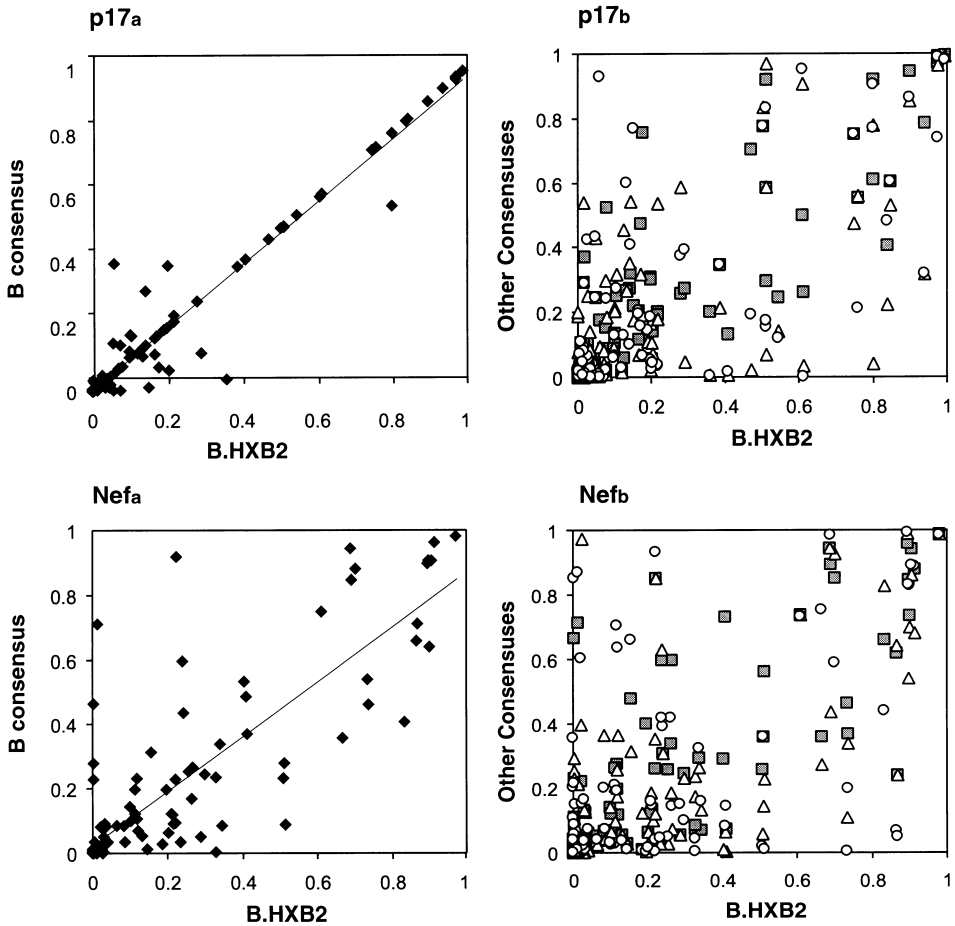


**Fig. 5** Distinctions between proteasomal cleavage sites within and between subtypes. This scatter diagram shows the correlation between the predicted probability of a site being a proteasome cleavage site based on a program developed by Can Kesmir *et al* (www.cbs.dtu.dk/services/NetChop/). The HXB2 sequence (x-axis) is compared to the subtype B, C, and AE consensus sequences as well as consensus of subtype consensus sequences (y-axis). The consensus sequences were aligned to HXB2 (see Fig. 4). Each point on the graph corresponds to an amino acid position and has two co-ordinates: cleavage prediction for this position in HXB2 and the cleavage prediction for the appropriate consensus sequence. Points representing the B consensus sequence cleavage prediction as ordinate are denoted by filled diamonds; C consensus, empty triangles; CRF01_AE consensus sequence, empty circles; the main group consensus, grey rectangles. p17a and Nefa: cleavage predictions for B consensus versus B.HXB2 in p17 and Nef, respectively. $R^2$ values for best fit lines are 0.95 for p17 and for 0.63 for Nef. p17b and Nefb: cleavage predictions for other consensus sequences versus HXB2 in p17 and Nef, respectively. $R^2$ values are 0.52 for p17 and 0.55 in Nef for the C consensus; 0.54 in p17 and 0.38 in Nef for the CRF01_AE consensus; and 0.77 for p17 and 0.62 for Nef for the M group consensus.

for inclusion of a cocktail of variants. Furthermore, epitope dense regions may make possible natural processing of epitopes that polyepitope vaccines could not. Thus it may be worth considering incorporating clustering information in vaccine design strategies.

## *References*


1 Kaul R, Rowland-Jones S. Methods of detection of HIV-specific CTL and their role in protection against HIV infection. In: Korber BTM, Moore JP, Brander BFHC, Walker BD, Koup R. (eds) *HIV Molecular Immunology Database*, part IV. Los Alamos, NM: Theoretical Biology and Biophysics, Los Alamos National Laboratory, 1999 (http://hiv-web.lanl.gov)
2 Drake JW. Rates of spontaneous mutation among RNA viruses. *Proc Natl Acad Sci USA* 1993; **90**: 4171–5
3 Wodarz D, Bangham C. Evolutionary dynamics of HTLV-I. *Mol Evol* 2000; **50**: 448–55
4 Perelson AS, Essunger P, Cao Y *et al*. Decay characteristics of HIV-1-infected compartments during combination therapy. *Nature* 1997; **387**: 188–91
5 Zhang LQ, MacKenzie P, Cleland A, Holmes EC, Leigh-Brown AJ, Simmonds P. Selection for specific sequences in the external envelope protein of human immunodeficiency virus type 1 upon primary infection. *J Virol* 1993; **67**: 3345–3356
6 Wolinsky S, Korber B, Neumann A *et al*. Adaptive evolution of human immunodeficiency virus type 1 during the natural course of infection. *Science* 1996; **277**: 537–42
7 Shankarappa R, Margolick J, Gange S *et al*. Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *J Virol* 1999; **73**: 10489–502
8 Lukashov VV, Kuiken CL, Goudsmit J. Intrahost human immunodeficiency virus type 1 evolution is related to length of the immunocompetent period. J Virol *1995*; **69**: 6911–6
9 Eigen M. On the nature of virus quasispecies. *Trends Microbiol* 1996; **4**: 216–8
10 Condra JH. Resistance to HIV protease inhibitors. *Haemophilia* 1998; **4**: 610–5
11 Price DA, Goulder PJ, Klenerman P *et al*. Positive selection of HIV-1 cytotoxic T lymphocyte escape variants during primary infection. *Proc Natl Acad Sci USA* 1997; **94**: 1890–5
12 Poignard P, Sabbe R, Picchio GR *et al*. Neutralizing antibodies have limited effects on the control of established HIV-1 infection *in vivo*. *Immunity* 1999; **10**: 431–8
13 Allen TM, O'Connor DH, Jing P *et al*. Tat-specific cytotoxic T lymphocytes select for SIV escape variants during resolution of primary viraemia. *Nature* 2000; **407**: 386–90
14 Korber B, Muldoon M, Theiler J *et al*. Timing the ancestor of the HIV-1 pandemic strains. *Science* 2000; **288**: 1789–96
15 Colgrove R. A combinatorial ledge: reverse transcriptase fidelity, total body viral burden, and the implications of multiple-drug HIV therapy for the evolution of antiviral resistance. *Antiviral Res* 1999; **41**: 45–56
16 Lubaki NM, Ray SC, Dhruva B, Quinn TC, Siliciano RF, Bollinger RC. Characterization of a polyclonal cytolytic T lymphocyte response to human immunodeficiency virus in persons without clinical progression. *J Infect Dis* 1997; **6**: 1360–7
17 Tang J, Costello C, Keet IP *et al*. HLA class I homozygosity accelerates disease progression in human. *AIDS Res Hum Retroviruses* 1999; **15**: 317–24
18 Borrow P, Lewicki H, Wei X *et al*. Anti-viral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus. *Nat Med* 1997; **3**: 205–11
19 Goulder P, Sewell A, Lalloo D *et al*. Patterns of immunodominance in HIV-1-specific cytotoxic T lymphocyte responses in two human histocompatibility leukocyte antigens (HLA)-identical siblings with HLA-A*0201 are influenced by epitope mutation. *J Exp Med* 1997; **8**: 1423–33
20 Sala M, Wain-Hobson S. Are RNA viruses adapting or merely changing? *Mol Evol* 2000; **51**: 12–20
21 Yang Z, Nielsen R, Goldman N, Pedersen AM. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 2000; **155**: 431–49

22 Gojobori T, Yamaguchi Y, Ikeo K, Mizokami M. Evolution of pathogenic viruses with special reference to the rates of synonymous and non-synonymous substitutions. *Jpn J Genet* 2000; **69**: 481–8
23 Simmonds P, Balfe P, Ludlam CA, Holmes EC, Leigh-Brown AJ. Analysis of sequence diversity in hypervariable regions of the external glycoprotein of human immunodeficiency virus type 1. *J Virol* 1990; **64**: 5840–50
24 Korber B. HIV signature and sequence variation analysis. In: Rodrigo AG, Learn GH. (eds) *Computational Analysis of HIV Molecular Sequences*. Dordrecht, The Netherlands: Kluwer, 2000
25 Yamaguchi-Kabata Y, Gojobori T. Re-evaluation of amino acid variability of the human immunodeficiency virus type 1 gp120 envelope glycoprotein and prediction of new discontinuous epitopes. *J Virol* 2000; **74**: 4335–50
26 Zanotto PM, Kallas EG, de Souza RF, Holmes EC. Genealogical evidence for positive selection in the nef gene of HIV-1. *Genetics* 1999; **153**: 1077–89
27 Moore JP, Cao Y, Leu J, Qin L, Korber B, Ho DD. Inter- and intra-subtype neutralization of human immunodeficiency virus type 1: the genetic subtypes do not correspond to neutralization serotypes but partially correspond to gp120 antigenic serotypes. *J Virol* 1995; **70**: 427–444
28 Nyambi PN, Nkengasong J, Lewi P *et al*. Multivariate analysis of human immunodeficiency virus type 1 neutralization data. *J Virol* 1996; **70**: 6235–43
29 Menendez-Arias L, Mas A, Domingo E. Cytotoxic T-lymphocyte responses to HIV-1 reverse transcriptase. *Viral Immunol* 1998; **11**: 167–81
30 Dorrell L, Dong T, Ogg GS *et al*. Distinct recognition of non-clade B human immunodeficiency virus type 1 epitopes by cytotoxic T lymphocytes generated from donors infected in Africa. *J Virol* 1999; **73**: 1708–14
31 Cao H, Kanki P, Sankale JL *et al*. CTL cross-reactivity among different HIV-1 clades: implications for vaccine development. *J Virol* 1997; **71**: 8615–23
32 Durali D, Morvan J, Letourneur F *et al*. Cross-reactions between the cytotoxic T-lymphocyte responses of human immunodeficiency virus-infected African and European patients. *J Virol* 1998; **72**: 3547–53
33 Rowland-Jones SL, Dong T, Fowke KR *et al*. Cytotoxic T cell responses to multiple conserved HIV epitopes in HIV-resistant prostitutes in Nairobi. *J Clin Invest* 1988; **102**: 1758–65
34 Buseyne F, Chaix ML, Fleury B *et al*. Cross-clade-specific cytotoxic T lymphocytes in HIV-1-infected children. *Virology* 1998; **250**: 316–24
35 Beer BE, Bailes E, Sharp PM, Hirsch VM. Diversity and evolution of primate lentiviruses. In: Kuiken CL, B. Foley B, B. Hahn B *et al*. (eds) *AIDS and Human Retroviruses HIV Sequence Database*, part III. Los Alamos, NM: Theoretical Biology and Biophysics, Los Alamos National Laboratory, 1999 (http://hiv-web.lanl.gov)
36 Hahn B, Shaw G, De Cock K, Sharp P. AIDS as a zoonosis: scientific and public health implications. *Science* 2000; **287**: 607–14
37 Chen Z, Telfier P, Gettie A *et al*. Genetic characterization of new West African simian immunodeficiency virus SIVsm: geographic clustering of household-derived SIV strains with human immunodeficiency virus type 2 subtypes and genetically diverse viruses from a single feral sooty mangabey troop. *J Virol* 1996; **70**: 3617–-27
38 Chen A, Luckay A, Sodora DL *et al*. Human immunodeficiency virus type 2 (HIV-2) seroprevalence and characterization of a distinct HIV-2 genetic subtype from the natural range of simian immunodeficiency virus-infected sooty mangabeys. *J Virol* 1997; **71**: 3953–60
39 Peeters M, Honore C, Huet T *et al*. Isolation and partial characterization of an HIV-related virus occurring naturally in chimpanzees in Gabon. *AIDS* 1989; **10**: 625–30
40 Gao F, Bailes E, Robertson DL *et al*. Origin of HIV-1 in the chimpanzee Pan troglodytes. *Nature* 1999; **397**: 436–41
41 Mauclere P, Loussert-Ajaka I, Damond F *et al*. Serological and virological characterization of HIV-1 group O infection in Cameroon. *AIDS* 1997; **11**: 445–53
42 Peeters M, Gueye A, M'Boup S *et al*. Geographical distribution of HIV-1 group O viruses in Africa. *AIDS* 1997; **11**: 493–8
43 Simon F, Mauclere P, Roques P *et al*. Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. *Nat Med* 1998; **4**: 1032–7
44 Zhu T, Korber BT, Nahmias AJ, Hooper E, Sharp PM, Ho DD. An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. *Nature* 1998; **391**: 594–6

45 Salemi M, Strimmer K, Hall WW *et al*. Dating the common ancestor of SIVcpz and HIV-1 group M and the origin of HIV-1 subtypes by using a new method to uncover clock-like molecular evolution. *FASEB J Express* 2000; On line, Dec 8: Article 10.1096/fj.00–0449fje
46 Sharp PM, Bailes E, Gao F, Beer BE, Hirsch VM, Hahn BH. Origins and evolution of AIDS viruses: estimating the time-scale. *Biochem Soc Trans* 2000; **28**: 275–82
47 Hillis DM, Mable BK, Moritz C. Application of molecular systematics: the state of the field and a look to the future. In: Hillis DM, Moritz C, Mable BK. (eds) *Molecular Systematics*, 2nd edn. Sunderland, MA: Sinauer, 1996; 515–43
48 Korber B, Theiler J, Wolinsky S. Limitations of a molecular clock applied to considerations of the origins of HIV-1. *Science* 1998; **280**: 1868–71
49 Grassly NC, Harvey PH, Holmes EC. Population dynamics of HIV-1 inferred from gene sequences. *Genetics* 1999; **151**: 427–38
50 Rodrigo AG, Shpaer EG, Delwart EL *et al*. Coalescent estimates of HIV-1 generation time *in vivo*. *Proc Natl Acad Sci USA* 1999; **96**: 2187–91
51 Pybus OG, Rambaut A, Harvey PH. An integrated framework for the inference of viral population history from reconstructed genealogies. *Genetics* 2000; **155**: 1429–37
52 Yusim K, Peeters M, Pybus OG, Bhattacharya T, Delaporte E, Mulanga C *et al*. Using HIV-1 sequences to infer historical features of the AIDS epidemic and HIV evolution. *Philos Trans R Soc Lond B Biol Sci* 2001; **356**: 855–65
53 Vidal N, Peeters M, Mulanga-Kabeya C, Nzilambi N, Robertson D, Ilunga W *et al*. Unprecedented degree of human immunodeficiency virus type 1 (HIV-1) group M genetic diversity in the Democratic Republic of Congo suggests that the HIV-1 pandemic originated in Central Africa. *J Virol* 2000; **74**: 10498–507
54 Louwagie J, McCutchan FE, Peeters M *et al*. Phylogenetic analysis of gag genes from 70 international HIV-1 isolates provides evidence for multiple genotypes. *AIDS* 1993; **7**: 769–80
55 Gao F, Robertson DL, Carruthers CD *et al*. A comprehensive panel of near-full-length clones and reference sequences for non-subtype B isolates of human immunodeficiency virus type 1. *J Virol* 1998; **72**: 5680–98
56 The European Commission and the Joint United Nations Programme on HIV/AIDS. HIV-1 subtypes: implications for epidemiology, pathogenicity, vaccines and diagnostics. *AIDS* 1997; **11**: 17–36
57 Robertson D, Anderson J, Bradac J *et al*. HIV-1 nomenclature proposal. In: Kuiken CL, Foley B, Hahn B *et al*. (eds) *AIDS and Human Retroviruses HIV Sequence Database*, part III. Los Alamos, NM: Theoretical Biology and Biophysics, Los Alamos National Laboratory, 1999 (http://hiv-web.lanl.gov)
58 Robertson D, Anderson J, Bradac J *et al*. HIV-1 nomenclature proposal. *Science* 2000; **288**: 55–6
59 Gao F, Robertson DL, Morrison SG *et al*. The heterosexual human immunodeficiency virus type 1 epidemic in Thailand is caused by an intersubtype (A/E) recombinant of African origin. *J Virol* 1996; **70**: 7013–29
60 Carr JK, Salminen MO, Koch C *et al*. Full-length sequence and mosaic structure of a human immunodeficiency virus type 1 isolate from Thailand. *J Virol* 1996; **70**: 5935–43
61 McCutchan FE, Carr JK, Bajani M *et al*. Subtype G and multiple forms of A/G intersubtype recombinant human immunodeficiency virus type 1 in Nigeria. *Virology* 1999; **254**: 226–34
62 Cornelissen M, van Den Burg R, Zorgdrager F, Goudsmit J. Spread of distinct human immunodeficiency virus type 1 AG recombinant lineages in Africa. *J Gen Virol* 2000; **81**: 515–23
63 Carr JK, Laukkanen T, Salminen MO *et al*. Characterization of subtype A HIV-1 from Africa by full genome sequencing. *AIDS* 1999; **13**: 1819–26
64 Su L, Graf M, Zhang Y *et al*. Characterization of a virtually full-length human immunodeficiency virus type 1 genome of a prevalent intersubtype (C/B') recombinant strain in China. *J Virol* 2000; **74**: 11367–76
65 Liitsola K, Tashkinova I, Laukkanen T *et al*. HIV-1 genetic subtype A/B recombinant strain causing an explosive epidemic in injecting drug users in Kaliningrad. *AIDS* 1998; **12**: 1907–19
66 Kuiken C. HIV-1 sequence alignments. In: Kuiken CL, Foley B, Hahn B *et al*. (eds) *AIDS and Human Retroviruses HIV Sequence Database*, part I. Los Alamos, NM: Theoretical Biology and Biophysics, Los Alamos National Laboratory, 1999 (http://hiv-web.lanl.gov)
67 Hahn BH, Shaw GM, Arya SK, Popovic M, Gallo RC, Wong-Staal F. Molecular cloning and characterization of the HTLV-III virus associated with AIDS. *Nature* 1984; **312**: 166–9

68 Reitter J, Means R, Desrosiers R. A role for carbohydrates in immune evasion in AIDS. *Nat Med* 1998; **4**: 679–84
69 Yang Z. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J Mol Evol* 1994; **39**: 306–14
70 Lamb RA. Genes and proteins of the influenza viruses. In: Krug RM. (ed) *The Influenza Viruses*. New York: Plenum, 1989; 1–87
71 Alexander DJ. A review of avian influenza in different bird species. *Vet Microbiol* 2000; **74**: 3–13
72 Smith FI, Palese P. Variation in influenza virus genes. In: Krug RM. (ed) *The Influenza Viruses*. New York: Plenum, 1989; 319–59
73 Subbarao K, Klimov A, Katz J *et al*. Characterization of an avian influenza A (H5N1) virus isolated from a child with a fatal respiratory illness [see comments]. *Science* 1998; **279**: 393–6
74 Guo Y, Li J, Cheng X *et al*. Discovery of men infected by avian influenza A (H9N2) virus. *Chin J Exp Cline Virol* 1999; **13**: 105–8
75 Peiris M, Yuen KY, Leung CW *et al*. Human infection with influenza H9N2 [letter]. *Lancet* 1999; **354**: 916–7
76 Nakajima K, Desselberger U, Palese P. Recent human influenza A (H1N1) viruses are closely related genetically to strains isolated in 1950. *Nature* 1978; **274**: 334–9
77 Kendal AP, Noble GR, Skehel JJ, Dowdle WR. Antigenic similarity of influenza A (H1N1) viruses from epidemics in 1977–1978 to 'Scandinavian' strains isolated in epidemics of 1950–1951. *Virology* 1978; **89**: 632–6
78 Stamboulian D, Bonvehi PE, Nacinovich FM, Cox N. Influenza. *Infect Dis Cline North Am* 2000; **14**: 141–66
79 Hay A, Gust I, Hampson A *et al*. Update: influenza activity – United States and worldwide, 1999–2000 season, and composition of the 2000–01 influenza vaccine. *MMWR* 2000; **78**: 1913–8
80 D'Souza MP, Livnat D, Bradac JA, Bridges SH. Evaluation of monoclonal antibodies to human immunodeficiency virus type 1 primary isolates by neutralization assays: performance criteria for selecting candidate antibodies for clinical trials. *Infect Dis* 1997; **175**: 1056–62
81 Parren PW, Moore JP, Burton DR, Sattentau QJ. The neutralizing antibody response to HIV-1: viral evasion and escape from humoral immunity. *AIDS* 1999; **13** (**Suppl A**): S137–62
82 Binley J, Sanders R, Clas B *et al*. A recombinant human immunodeficiency virus type I envelope glycoprotein complex stabilized by an intermolecular disulfide bond between the gp120 and gp41 subunits is an antigenic mimic of the trimeric virion-associated structure. *J Virol* 2000; **74**: 627–43
83 Nyambi P, Mbah H, Burda S, Gorny M, Nadas A, Zolla-Pazner S. Conserved and exposed epitopes on intact, native, primary human immunodeficiency virus type I virions of group M. *J Virol* 2000; **74**: 7096–107
84 Connor R, Korber BTM, Cao Y *et al*. Immunological and virological analyses of persons infected by human immunodeficiency virus type 1 while participating in trials of recombinant gp120 subunit vaccines. *J Virol* 1998; **72**: 1552–76
85 Locher CP, Grant RM, Collisson EA *et al*. Antibody and cellular immune responses in breakthrough infection subjects after HIV type 1 glycoprotein 120 vaccination. *AIDS Res Hum Retroviruses* 1999; **15**: 1685–9
86 Collado M, Rodriguez D, Rodriguez J, Vazquez I, Gonzalo R, Esteban M. Chimeras between the human immunodeficiency virus (HIV-1) Env and vaccinia virus immunogenic proteins p14 and p39 generate in mice broadly reactive antibodies and specific activation of $CD8^+$ T cell responses to Env. *Vaccine* 2000; **18**: 3123–33
87 LaCasse R, Follis K, Trahey M, Scarborough JD, Littman DR, Nunberg J. Fusion-competent vaccines: broad neutralization of primary isolates of HIV. *Science* 1999; **283**: 357–62
88 Salzwedel K, Smith ED, Dey B, Berger EA. Sequential CD4-coreceptor interactions in human immunodeficiency virus type 1 Env function: soluble CD4 activates Env for coreceptor-dependent fusion and reveals blocking activities of antibodies against cryptic conserved epitopes on gp120. *J Virol* 2000; **74**: 326–33
89 Stamatatos L, Cheng-Mayer C. An envelope modification that renders a primary, neutralization-resistant clade B human immunodeficiency virus type 1 isolate highly susceptible to neutralization by sera from other clades. *J Virol* 2000; **72**: 7840–5

90 Bernard NF, Yannakis CM, Lee JS, Tsoukas CM. Human immunodeficiency virus (HIV)-specific cytotoxic T lymphocyte activity in HIV-exposed seronegative persons. *J Infect Dis* 1999; **179**: 538–47

91 Goh WC, Markee J, Akridge RE *et al*. Protection against human immunodeficiency virus type 1 infection in persons with repeated exposure: evidence for T cell immunity in the absence of inherited CCR5 coreceptor defects. J Infect Dis 1999; **179**: 548–57

92 Biasin M, Caputo SL, Speciale L *et al*. Mucosal and systemic immune activation is present in human immunodeficiency virus-exposed seronegative women. *J Infect Dis* 2000; **182**: 1365–74

93 Kaul R, Plummer F, Kimani J *et al*. HIV-1-specific mucosal $CD8^+$ lymphocyte responses in the cervix of HIV-1-resistant prostitutes in Nairobi. *J Immunol* 2000; **164**: 1602–11

94 Rowland-Jones SL, Dong T, Dorrell L *et al*. Broadly cross-reactive HIV-specific cytotoxic T-lymphocytes in highly-exposed persistently seronegative donors. *Immunol Lett* 1999; **66**: 914

95 Greenough TC, Brettler DB, Kirchhoff F *et al*. Long-term non-progressive infection with human immunodeficiency virus type 1 in a hemophilia cohort. *J Infect Dis* 1999; **180**: 1790–802

96 Rosenberg E, Walker B. HIV type 1-specific helper T cells: a critical host defense. *AIDS Res Hum Retroviruses* 1998; **14** (**Suppl 2**): S143–7

97 Evans D, Knapp L, Jing P *et al*. Rapid and slow progressors differ by a single MHC class I haplotype in a family of MHC-defined rhesus macaques infected with SIV. *Immunol Lett* 1999; **613**: 53–9

98 Kaslow RA, Carrington M, Apple R *et al*. Influence of combinations of MHC genes on the course of HIV-1 infection. *Nat Med* 1996; **2**: 405

99 Tomiyama H, Sakaguchi T, Miwa K *et al*. Identification of multiple HIV-1 CTL epitopes presented by HLA-B*5101 molecules. *Hum Immunol* 1999; **60**: 177–86

100 Goulder P, Phillips RE, Colbert RA *et al*. Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS. *Nat Med* 1997; **3**: 212–6

101 Paranjape RS, Gadkari DA, Lubaki M, Quinn TC, Bollinger RC. Crossreactive HIV-1-specific CTL in recent seroconverters from Pune, India. *Indian J Med Res* 2000; **108**: 35–41

102 Gotch F. Cross-clade T cell recognition of HIV-1. *Curr Opin Immunol* 2000; **10**: 388–92

103 Jin X, Roberts CG, Nixon DF *et al*. Longitudinal and cross-sectional analysis of cytotoxic T lymphocyte responses and their relationship to vertical human immunodeficiency virus transmission. *J Infect Dis* 1998; **178**:1317–26

104 van Baalen CA, Pontesilli O, Huisman RC *et al*. Human immunodeficiency virus type 1 Rev- and Tat-specific cytotoxic T lymphocyte frequencies inversely correlate with rapid progression to AIDS. *J Gen Virol* 1997; **78**: 1913–8

105 Korber B. CTL epitopes. In: Korber BTM, Moore JP, Brander BFHC, Walker BD, Koup R. (eds) *HIV Molecular Immunology Database*, part 1. Los Alamos, NM: Theoretical Biology and Biophysics, Los Alamos National Laboratory, 1999 (http://hiv-web.lanl.gov)

106 Legrand E, Pellegrin I, Neau D *et al*. Course of specific T lymphocyte cytotoxicity, plasma and cellular viral loads, and neutralizing antibody titers in 17 recently seroconverted HIV type 1-infected patients. *AIDS Res Hum Retroviruses* 1997; **13**: 1383–94

107 Hanke T, McMichael A. Pre-clinical development of a multi-CTL epitope-based DNA prime MVA boost vaccine for AIDS. *Immunol Lett* 1999; **66**: 177–81

108 Ferrari G, Kostyu DD, Cox J *et al*. Identification of highly conserved and broadly cross-reactive HIV type 1 cytotoxic T lymphocyte epitopes as candidate immunogens for inclusion in *Mycobacterium bovis* BCG-vectored HIV vaccines. *AIDS Res Hum Retroviruses* 2000; **16**: 1433–43

109 Goulder PJ, Brander C, Annamalai K *et al*. Differential narrow focusing of immunodominant human immunodeficiency virus gag-specific cytotoxic T-lymphocyte responses in infected African and Caucasoid adults and children. *J Virol* 2000; **74**: 5679-90

110 Fomsgaard A. HIV-1 DNA vaccines. *Immunol Lett* 1999; **65**: 127–31

111 Wagner R, Shao Y, Wolf H. Correlates of protection, antigen delivery and molecular epidemiology: basics for designing an HIV vaccine. *Vaccine* 1999; **17**: 17O6–10

112 Evans TG, Keefer MC, Weinhold KJ *et al*. A canarypox vaccine expressing multiple human immunodeficiency virus type 1 genes given alone or with rgp120 elicits broad and durable $CD8^+$ cytotoxic T lymphocyte responses in seronegative volunteers. *J Infect Dis* 1999; **180**: 290–8

113 Akahata W, Ido E, Shimada T *et al*. DNA vaccination of macaques by a full genome HIV-1 plasmid which produces non-infectious virus particles. *Virology* 2000; **275**: 116–24

114 Hamajima K, Fukushima J, Bukawa H *et al*. Strong augment effect of IL-12 expression plasmid on the induction of HIV-specific cytotoxic T lymphocyte activity by a peptide vaccine candidate. *Cline Immunol Immunopathol* 1997; **83**: 179–84
115 Xin KQ, Hamajima K, Sasaki S *et al*. IL-15 expression plasmid enhances cell-mediated immunity induced by an HIV-1 DNA vaccine. *Vaccine* 1999; **17**: 858–66
116 Barouch DH, Craiu A, Kuroda MJ *et al*. Augmentation of immune responses to HIV-1 and simian immunodeficiency virus DNA vaccines by IL-2/Ig plasmid administration in rhesus monkeys. *Proc Natl Acad Sci USA* 2000; **97**: 4192–7
117 Lukashov VV, Goudsmit J. Evolution of the human immunodeficiency virus type 1 subtype-specific V3 domain is confined to a sequence space with a fixed distance to the subtype consensus. *J Virol* 1997; **71**: 6332–8
118 Korber B, Foley B, Gaschen B, Kuiken C. Epidemiological and immunological implications of the global variability of HIV-1. In: Pantaleo G, Walker B (eds) *Retroviral Immunology: Immune Response and Restoration*. Totowa, NJ: Humana, 2001; In press
119 Orlowski M, Wilk S. Catalytic activities of the 20 S proteasome, a multicatalytic proteinase complex. *Arch Biochem Biophys* 2000; **383**: 1–16
120 Pamer E, Cresswell P. Mechanisms of MHC class I-restricted antigen processing. *Annu Rev Immunol* 1998; **16**: 323–58
121 Abele R, Tampe R. Function of the transport complex TAP in cellular immune recognition. *Biochim Biophys Acta* 1999; **1461**: 405–19
122 Rammensee H, Bachmann J, Stevanovic S. Norwell, MA: *MHC Ligands and Peptide Motifs*. Chapman & Hall, Kluwer Academic, 1999
123 Paradela A, Alvarez T, Garcia-Peydro M *et al*. Limited diversity of peptides related to an alloreactive T cell epitope in the HLA-B27-bound peptide repertoire results from restrictions at multiple steps along the processing-loading pathway. *J Immunol* 2000; **164**: 329–37
124 Beekman N, van Veelen P, van Hall T *et al*. Abrogation of CTL epitope processing by single amino acid substitution flanking the C-terminal proteasome cleavage site. *J Immunol* 2000; **164**: 1898–905
125 Chassin D, Andrieu M, Cohen W *et al*. Dendritic cells transfected with the nef genes of HIV-1 primary isolates specifically activate cytotoxic T lymphocytes from seropositive subjects. *Eur J Immunol* 1999; **29**: 196–202
126 Niedermann G, Geier E, Lucchiari Hartz M, Hitziger N, Ramsperger A, Eichmann K. The specificity of proteasomes: impact on MHC class I processing and presentation of antigens. *Immunol Rev* 1999; **172**: 29–48
127 Geier E, Pfeifer G, Wilm M *et al*. A giant protease with potential to substitute for some functions of the proteasome. *Science* 1999; **283**: 978–81
128 Stoltze L, Schirle M, Schwarz G *et al*. Two new proteases in the MHC class I processing pathway. *Nat Immunol* 2000; **1**: 413–8
129 Daniel S, Brusic V, Caillat Zucman S *et al*. Relationship between peptide selectivities of human transporters associated with antigen processing and HLA class I molecules. *J Immunol* 1998; **161**: 617–24
130 Kesmir C, Nussbaum A, Schild H, Detours V, Brunak S. Program available on the web: <http://www.cbs.dtu.dk/services/NetChop>. Manuscript submitted, 2001
131 Macken C, and The Influenza Sequence Database Staff. *Influenza Sequence Database*. 2000 <http://www.flu.lanl.gov>