Within-host evolution of HIV-1

From basic multi-site theory to antigenic escape

I.M.Rouzine

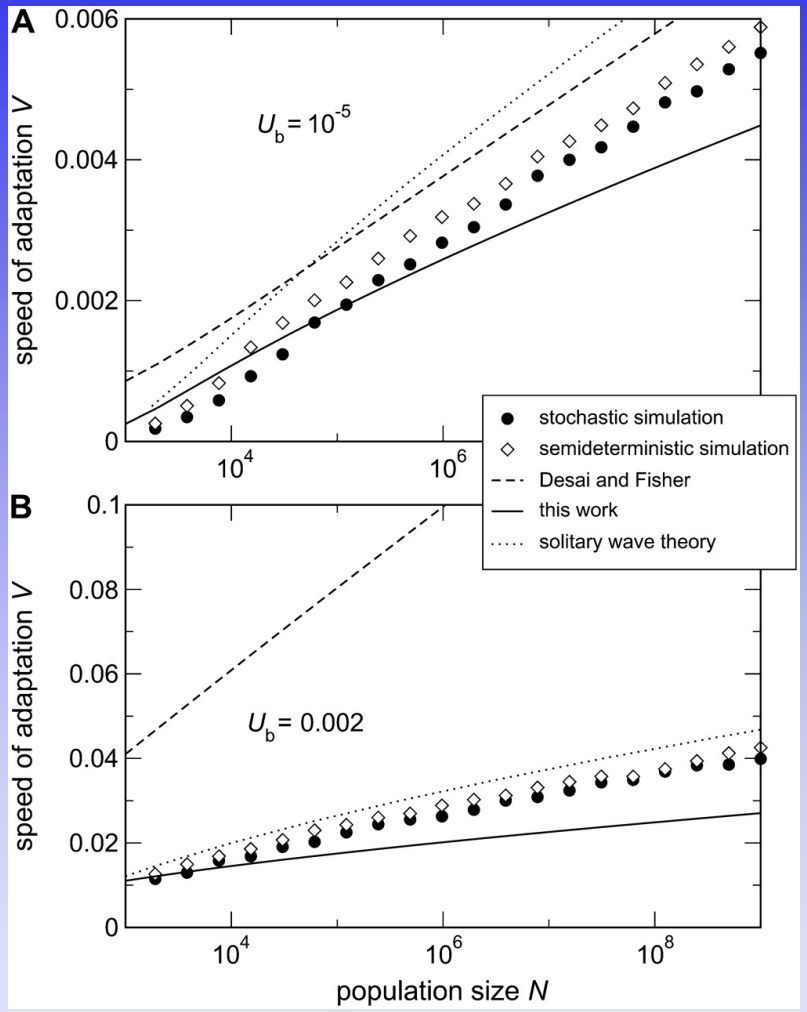
Tufts University, Boston

Part 1A: Asexual theory

Adaptation at fixed selection advantage s in "multiple mutation" regime -->BB

Name	"Continuous landscape"	"Stochastic edge"	"Traveling wave"
Parameter interval	$s << U_b$, $V-U_b << U_b$	$s >> U_b$ $s/\ln(s/U_b) << V << s$	$V >> \max(s, U_b)$
Landscape	Multiplicative	Multiplicative	Multiplicative
Mutation in bulk	Diffusion (linear)	Neglected	Continuous in log mutation load (non-linear)
Stochastic class number	Many	1	1
Treatment of stochastic tip	Smooth cutoff (heuristic or exact)	Branching process	Stochastic threshold
Refs	Tsimring et al 1996 Brunet et al 1997 Hallatschek 2010	Desai and Fisher 2007 Brunet et al 2008	Rouzine et al 2003 Rouzine et al 2008

Stochastic egde vs Traveling wave ($V >> U_b$)



Details
On black board

Brunet, E. et al. Genetics 2008;179:603-620

GENETICS

Summary 1A

- Basic asex theory for multi-site regime and fixed s is OK
- We need:

(i) more detailed predictables (coalescent, including ratchet and ss) (Cf Desai & Walczak 2011)

(ii) distributed s (cf 2-site: Gerrish, Schiffels)

(iii) simplest epistatic models

Rare recombination

- Selection, drift, and mostly asexual individuals recombine with small probability *r*, *M* >= 1 crossovers per genome (NB: M does not matter in the multi-site case for many properties)
- Traveling wave with quasi-stationary profile: deterministic bulk+stochastic edge
- The **engine** of adaptation is extension of the edge by new highly fit recombinants with parents far from the edge: long shot in fitness space (not short range, as by mutation)
- Alleles preexist and/or added by mutation (**fuel**)

General framework

- Distribution of progeny log fitness around parental, per each generation, is Gaussian with the variance $\sigma^2 = s^2/4$ x (genetic distance). Applied in a very general situation

- This is not to be confused with the population fitness distribution, which is derived from evolution equation, is more narrow (distributed linkage disequilibrium), has a cutoff, can have non-Gaussian prefactor etc

- Given the Gaussian width σ, to be found later self-consistently, properties of the wave are general (Rouzine & Coffin 2005; 2007; 2010) ----->details on blackboard and next slide

Main properties

Rouzine & Coffin TPB 2010

- Fitness distribution has high log fitness cutoff (edge) x_0, $x_0^2 \sim 2\sigma^2 \ln(Nr)$
- Wave is driven by new recombinants being formed at the edge
- Shape stays nearly Gaussian with variance $p\sigma^2$, where $p <\sim 1$ reflects distributed linkage disequilibrium (fitness correlations among genomes
- $v = p\sigma^2$ is the adaptation rate (FFT)
- Value of p, clone structure, ancestral relationship, are all controlled by the clone decay parameter $\beta = rx_0/v$ showing how much a clone born at edge, x_0 decays due to recombination with other sequences until it becomes average, x=0
- At $\beta << 1$, $p \sim 1 - 2\ln(1/\beta)/\ln(Nr)$, single clone born at edge x_0 dominates each fitness class. It's likeliest ancestors are in the middle of the tail, $x_0/2$, i.e., atypically well fit.
- At $\beta >> 1$, $p = 1 + O[\exp(-\beta)]$, a class with fitness x is comprised of many smaller clones born far from the edge, $x > x_0$. The likeliest parents of the next generation of wave are also shifted towards rthe center, $x > 0.5x_0$
- Ancestral gene tree can be derived from the above clone structure, because fitness classes are well mixed by recombination in time (unlike in few-locus models)
- The tree is nearly neutral in shape but compressed in time, with minimum $T_{MCRA} \sim 1/s$ at $\beta \sim 1$.
- Coalescent events come mostly from rare clones born far ahead of the typical edge

Width of recombination kernel σ^2

Some special cases:

1) Stationary case with beneficial mutation

Self-consistency: $NU_b p_{fix} = V = v/s$

(Neher, Shraiman & Fisher 2010, approximating fitness distribution with exact Gaussian of width σ)

2) Transitory dynamics, standing variation at L identical sites (Rouzine & Coffin 2010):

iInter-genomic correlations accumulating in time:

$\sigma^2(t) = s^2 L[1-C(t)][f-C_{loss}(t)][1-f(t)]/[1-C_{loss}(t)]/2$

$C(t)$ pairs of correlated sites (with same ancestor)

$C_{loss}(t)$ sites where all population is correlated (good alleles lost)

$1-f(t)$ good allele frequency

Self-consistency: $dC/dt = (1-C)/N_{anc}(\sigma^2)$

C_{loss} vs C is the neutral tree relation

$N_{anc}(\sigma^2)$ and the coalescence density derived from the clone structure of fitness classes

Stationary case revisited: Connecting the car RC 2010 with fuel injection NSF 2010

- NSF 2010: self-consistent condition for σ has been solved in the simple approximation that the fitness distribution of the population is exact, constant Gaussian with width σ, same as in the starting recombination kernel
- The validity is an open issue (cf. Fig. 2A in NSF 2010)
- Various features of the real general solution obtained in RC 2010 for given σ are potentially important and are being discussed

Connection to 2-site approach to full sex?

- 2-site theory by Weissman and Barton for $r = 1$: the crossover number per genome is important for the adaptation rate.
- In multi-site theory, it is not: only outcrossing rate matters (crossover number matters for some measures of LD)
- Any hope of connection? At 30% 50% 90% sex?

Part II. Applications to HIV-1
Long-term evolution (years):

- Untreated chronic in typical patients: $U_b \sim 0.05$, $\langle s \rangle \sim 0.005$, $r \sim 0.01$ ($M = 10$), $V \sim U_b \sim 2V_{r=0} = V_{large\ r}/4$

(Batorsky et al 2011, submitted)

- Atypically low virus load: $r = 0$ => Asex close to the "continuous fitness" case, $s << V <\sim U_b$
- Treated patients, hidden pockets: farther in due to low N => lower V

Short-term evolution: less sites, more selection advantage

- Short-time scale, compensatory for drug-resistance : s > 1%, smaller U_b ~ 10^{-3}-10^{-4} “Traveling wave” or “stochastic edge” regime asex; or with recombination.
- Primary drug resistance: even larger s and smaller U_b.

Antigenic escape: basics

- Genome has several dozens of 8-codon long cytotoxic epitopes, individual for each patient.
- CTL clones for many of them are activated by virus
- Escape mutants occur at 5-25 epitopes, most of them within 6 months.
- Time gap between escapes is increasing
- 20-70 escape mutants per epitope are observed (singles and doubles). Only one is fixed.
- CTL clone number is increasing

Variable sites are in epitopes: Goonetilleke et al 2009

CH40

reactive 18mer / optimal 9mer	a. Gag	b. Gag	d. Pol	f. Vif	g. Vif	h. Vif	i. Vpr	j. Tat	k. Env	l. Env	m. Env	n. Env	o. Rev	q. Env	r. Nef	s. Nef	t. Nef
		Gag (389-406) (395-403)			Vif (113-130) (119-127)								Rev (49-66)	Env (830-847)			Nef (185-202) (187-195)
0									0%	0%	0%	0%		0%	0%	0%	0%
16	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	55%
45	100%	100%	58%	7%	0%	93%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%
111	100%	100%	38%	22%	11%	67%	100%	11%	100%	33%	0%	0%	22%	11%	33%	22%	100%
181	100%	100%	73%	71%	71%	100%	100%	0%	100%	71%	0%	14%	14%	14%	86%	14%	100%
412	100%	100%	91%	100%	75%	100%	100%	67%	100%	100%	92%	100%	100%	16%	92%	92%	100%
	(A120D) *(V128G) # (126-*PGS)	*(I401L) *(R403K)	(K76R)	*(E51G) (D62E/ G/A)	*(E117D) (K122T/E) (I124L) (L125V)	(T167I)	(I61L) (I63M) (V70I) (H78R) (I81L) (S84*R/G) (Q86H)	*(Q54P) (R57G)	(N136K) *(N139T) (G145E) *(E146K/ G/T) *(M147I/L)	(T295N) (N300H) *(T324I) *(R326K) *(Y331S) (E331 K/A) (E332L)	(G410D) (T411R) (H413M) (K417R)	(I439V) (K442*E /*K/G) (R444S)	(R50K) (L60F) *(S61N)	(V833I) (L842F R/C) (R847K)	(W57R) *(G64E)	(P147L) *(G151E) (K158*E/T)	(S188R/*N) (R192H) *(R196Q)

CH77

reactive 18mer / optimal 9mer	b. Gag	c. Gag	f. Pol	g. Pol	i. Pol	j. Vpr	k. Tat	l. Tat	l. Rev	m. Env	n. Env	o. Env	p. Env	q. Env	r. Env	s. Env	t. Env	u. Env	v. Env
	Gag IW9 (140-157) (146-155)	Gag TW10 (236-253) (240-249)							Rev (9-26)						Env (334-351)	Env (350-368) (352-360)			Env (597-614)
0								0%		4%		0%	0%		0%	0%			0%
14	0%	0%	0%	0%	47%	0%	0%	100%	13%	7%	14%	0%	0%	0%	0%	87%	0%	0%	7%
32	0%	0%	17%	11%	17%	0%	0%	100%	8%	8%	0%	0%	0%	0%	0%	100%	0%	0%	0%
102	50%	0%	25%	75%	0%	40%	20%	100%	40%	80%	20%	100%	0%	0%	20%	100%	40%	0%	0%
159	100%	100	55%	100%	0%	100%	100%	100%	8%	100%	42%	100%	92%	17%	68%	100%	0%	17%	0%
592	100%	100	88%	100%	0%	100%	100%	100%	100%	100%	100%	100%	0%	0%	100%	100%	100%	0%	100%
	(I147L)	*(T242N) *(V247I) *(G248E/A)	*(R3K) *(N5D)	(N478D)	(N673D)	(A59D) (T61I) (I63T/*M/N) (Q165H)	(K29Q/N) *(F32Y)	(R57Q) (R57K) (P58L/S/T) (A59S) (D61N/*G) (K63E) (I64*T/N/S)	(K14R) (T15I/*A) (R17G) (L18V/I)	(A3V) (R4K) *(V5E) (I6V) (R9K) *(C10Y) (Q11R) (H12Y) (L13S)	*(N130D) (S134F) *(N135D) *(D137N) (S143N) (S145N) (S146R) *(K150R) (E151Q)	(T186A) (K187T/*E) (T188*N/A) (N189D/S) (T190I) (K192E)	(D230 E) (K232Q/R) *(K243N)	(S121A)	(S344N/G) *(H345Y) (V346I) (D348A/N)	(R355*K/ S/N) (K357R/G) (T359A) (V360G/A)	*(D386N) (W395R) (R396W/G) (R397K)	(N460S) (N461D/T/S) (D462E/G) (S463N/G)	(G597D) (V612*T/*S) (N616S)

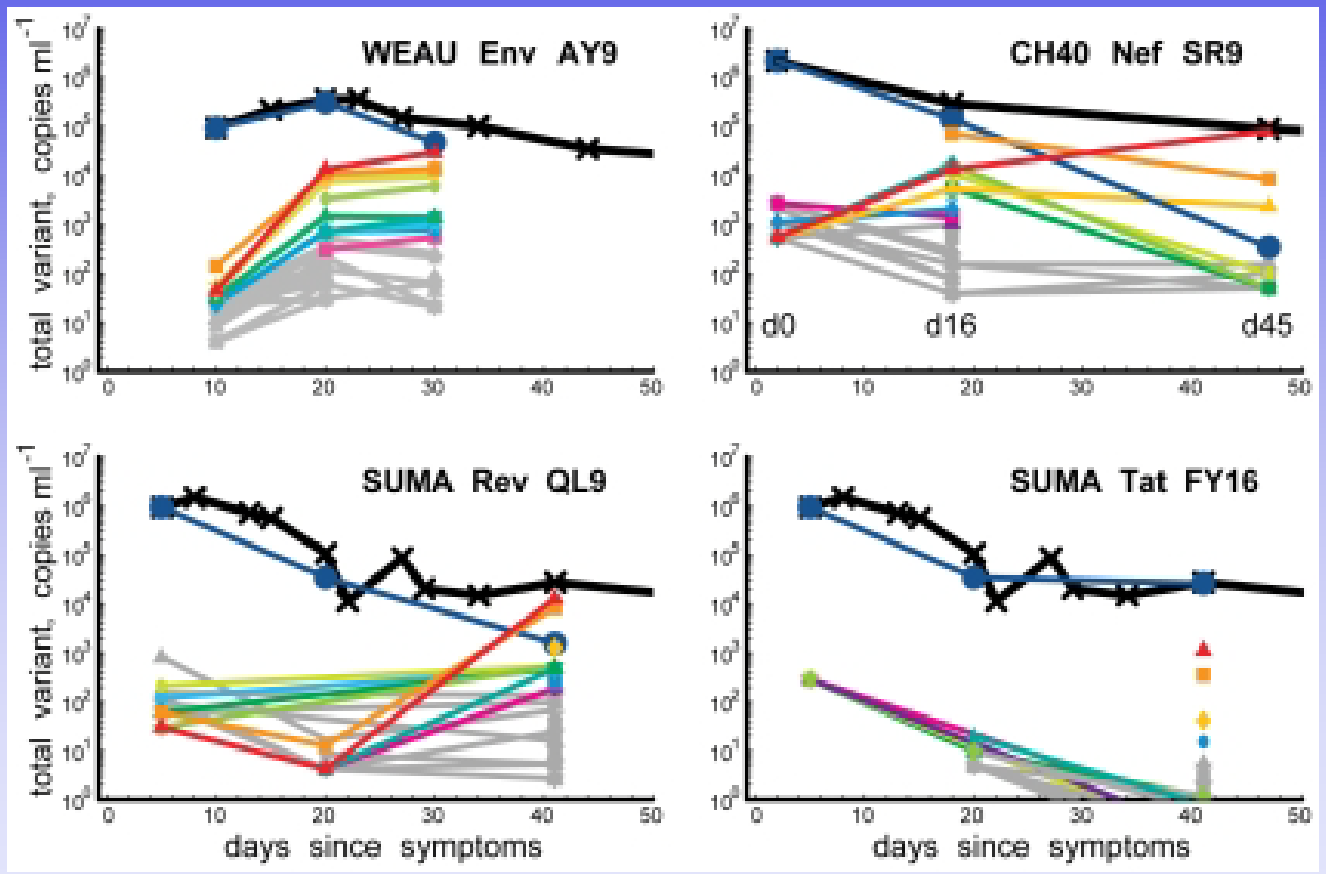
CH77 cont.

reactive 18mer / optimal 9mer	w. Env	x. Env	y. Env	z. Env	aa. Nef	bb. Nef	cc. Nef
			Env (822-839)	Env (838-855)	Nef (17-34)		Nef (73-90) (82-90)
0			0%	0%			
14	0%	0%	7%	0%	21%	0%	0%
32	0%	0%	0%	0%	100%	0%	100%
102	0%	40%	100%	0%	100%	80%	100%
159	0%	0%	100%	0%	100%	100%	100%
592	60%	100%	100%	100%	100%	100%	100%
	(L721R)	(E739K)	*(A821T)	(I843V)	(R21K/G)	*(V49A)	(V76L)

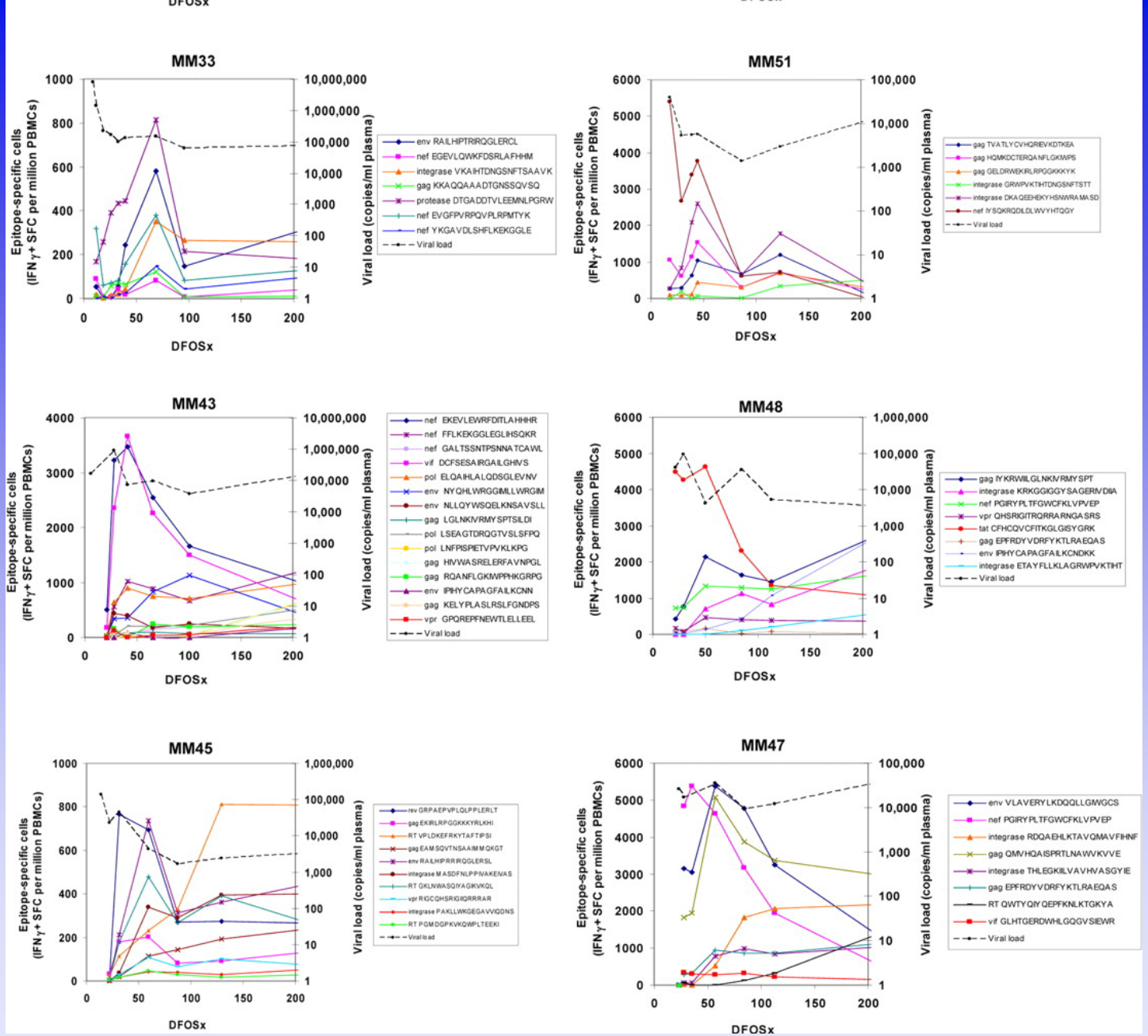
CH58

reactive 18mer / optimal 9mer	a. Gag	c. Gag	d. Env	e. Env	f. Rev	f. Env	g. Env	h. Nef
		Gag TW10 (236-253) (240-249)		Env (576-595) (584-592)				Nef (105-122)
0			0%	0%		0%	0%	
9	0%	0%	14%	28%	0%	0%	0%	0%
45	0%	23%	0%	89%	22%	22%	89%	33%
85	11%	89%	11%	100%	33%	33%	100%	89%
154	50%	100%						
350	100%	100%	100%	100%	100%	100%	100%	100%
	*(V83A)	*(T242N)	(R334K)	(Y586H)	(R50I)	(D736Y)	(F829V/S/L)	(K110R)

Many escape mutants (EM), one fixed



Fischer et al 2010



Turnbull et al 2009

Slope of escape
(Perelson's and Lee's groups 2009)

- Escape slope is decreasing from 0.3 for earliest escapes, to almost none (~ 0.02) within a year
- Escape slope ε is measured by fitting $1/\{1+\exp[-\varepsilon(t-t_{50})]\}$ to the winner frequency.
- Predicted ε >~ 1 for full CTL pressure per site

Major players

- Depletion of target cells
- Avidity (antigen sesitivity) of CTL subsets
- Mutation costs

Who is responsible for what feature?

Model equations

Simplified model by Althaus & de Boer 2008:

$$\frac{dT}{dt} = \lambda - \left[\delta_T - p_R \sum_{seq} I_{seq} f_{seq} \right] T, \quad p_R = \beta p / \delta_V$$

I_{seq} : Number of cells infected with a sequence

$f_{seq} = \exp[-\sum_{EM} s_{EM}]$: Sequence fitness

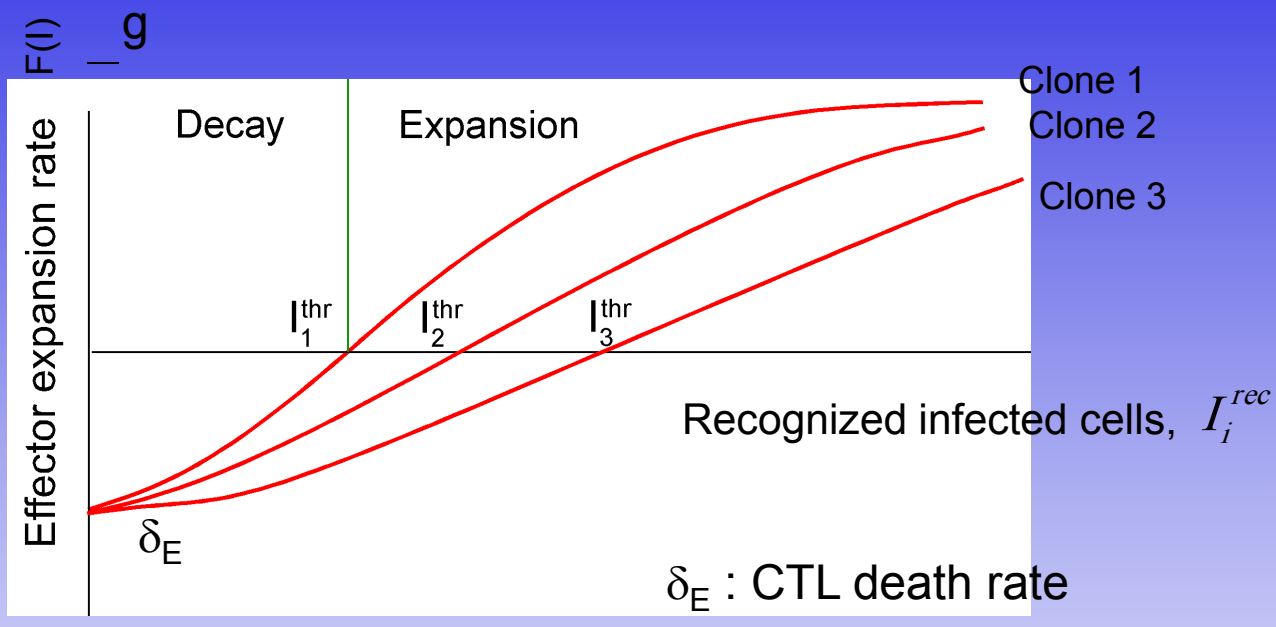
$$\frac{dI_{seq}}{dt} = \left(p_R T f_{seq} - \delta_I - k \sum_{i:\text{recognize}} E_i \right) I_{seq} + \sum_{seq'} M_{seq,seq'} I_{seq'}$$

$$\frac{dE_i}{dt} = \sigma + F(I_i^{rec}) E_i,$$

I_i^{rec} : Total # of cells infected with sequences recognized by ith clone, E_i

$F(I_i^{rec})$: CTL expansion rate (next slide)

Thresholds in antigen for CTL subsets



δ_E : CTL death rate

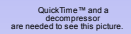
g: maximum possible CTL expansion rate

n CTL subsets ordered in descending avidity

n^{th} site cannot mutate

Dynamic parameters are mostly known from HIV dynamic studies, but the avidity (threshold) and mutation costs are ours to play with

Attempt of simulation of antigenic escape: one mutant per epitope, s=0, CTL thresholds equidistant in log



Caption for the previous slide

Preliminary simulation of dynamics of antigenic escape in the limit of low mutation cost and a fixed step in log avidity between epitopes. (A) Frequencies of escape variants at consecutive epitopes (numbers on curves) ordered in descending avidity. (B) Frequencies of multiple mutants, mutated sequential epitopes are shown. (C) Numbers of CTLs recognizing different epitopes (numbers on curves). (D) Depletion of permissive cells with respect to their level in uninfected host. Inset: Analytic and simulated predictions for the maximum number of mutated epitopes. *X*-axis: $x = (1/a) \log[0.7\lambda/(I_1^{thr}\delta_P)] + 1$. Here a is the step of log avidity: $h_n = h_1 \exp[-a(n-1)]$. Other parameters are defined in the legend to Fig. 5. Parameter values in A-D: δ_T = 1/day, p = 4.0 10^{-7} /day/cell, λ = 1.0 10^8 cell/day, δ_P = 1/day, k = 3.0 10^{-9} /day/cell, δ_E = 0.2/day, $\sigma_n = \sigma$ = 2.0 10^5 cell/day, g = 2/day, s = 0, μ = 3 10^{-5}, h_1 = 1.0 10^7 cell, a = 0.634 (varies in the inset on panel D). Initial conditions: $E_n(0)$ and $T(0)$ at their steady state levels in uninfected host, $I_{wt}(0)$ = 10 cell, other sequences absent.

Discrepancies

- CTL clone number decreases (cf. Turnbull et al 2009)
- CTL disappear and escape stops due to T cell depletion when death=replication w/o CTL (cf. rapid surge of virus under CD8 T cell depletion, Schmitz et al 1999 etc)
- Slopes of early escapes 3-fold large
- Fixing it with large s ~ 1 does not work: once s >~ 5%, WT rapidly comes back and we have steady diversity instead of full escapes as observed
- Small s do not help with slopes

Fixing our problems: Making use of multi-site effects

- Idea 1: Similar thresholds for a cluster of epitopes => distributed pressure
- Idea 2: Interference of EM within and between epitopes slows down escape and causes delay (works differently within and between)
- Both ideas seem to work (have simulation examples…)

Preliminary conclusions on antigenic escape

- One-locus model does not work, although remains a valuable bioinformatic tool)
- Constant selection pressure may effectively work in a time interval (epitope cluster with similar avidity)
- Mutation cost of fixed EM is small <~ 10%
- Order or escapes is set by CTL antigen thresholds
- Threshold ladder should converge up, then clonal interference and distribution of CTL pressure will cause more and more delay and smaller slopes
- Depletion of T cells is not really that important (for increasing time intervals between escapes)
- Need to replace deterministic simulation with cutoff 1/N by a correct analytic multi-ste model (Part 1)

Acknowledgements

Rebecca Batorsky

Collaboration with DRP NCI-NIH:

Mary Kerney
Fred Maldarelli
John Coffin

Support

NCI-NIH 5R37CA089441-09