HLA-driven Optimization of an HIV Vaccine Immunogen Using Epitomes

N. Jojic, V. Jojic, C. Kadie, C. Meek, and D. Heckerman
Microsoft Research

M. John, C. Moore, and S. Mallal
Royal Perth Hospital and Murdoch University

Contact email: jojic@microsoft.com

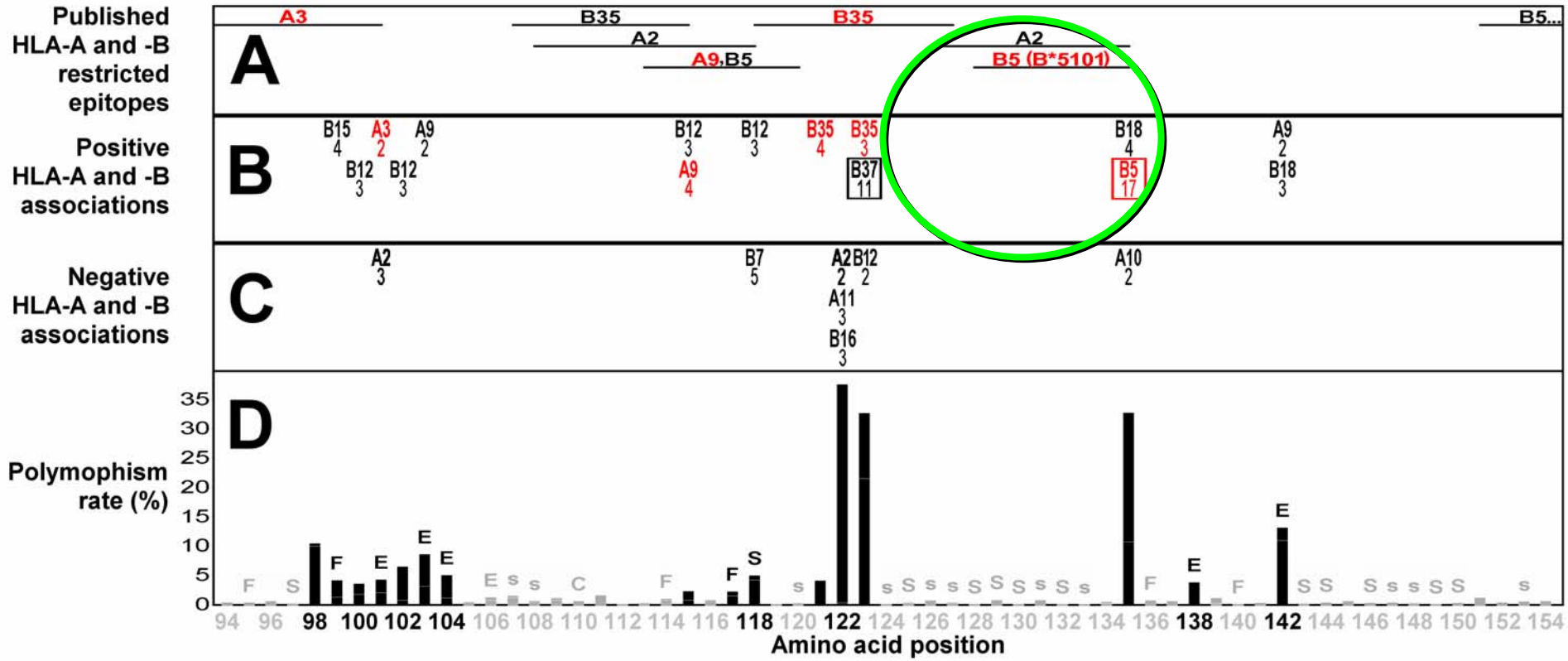
A good vaccine immunogen: Our assumptions

- More epitopes are better BUT longer immunogens are worse
- Epitope must have proper flanking region to be presented
- Cross reactivity should be exploited if possible
- Immunodominance must be addressed

Overview

- Use discovered associations between HLA and HIV adaptation to identify new epitopes
- Use newly identified and previously known epitopes to design vaccine immunogens (for Nef) using multiple optimization criteria
  - How many new epitopes are there?
  - How long should the immunogen be?
  - What is the correct model for cross reactivity?
- Analyze sensitivities to the various criteria

HIV adaptation to HLA-restricted immune reponses is evident at population level



within published epitopes

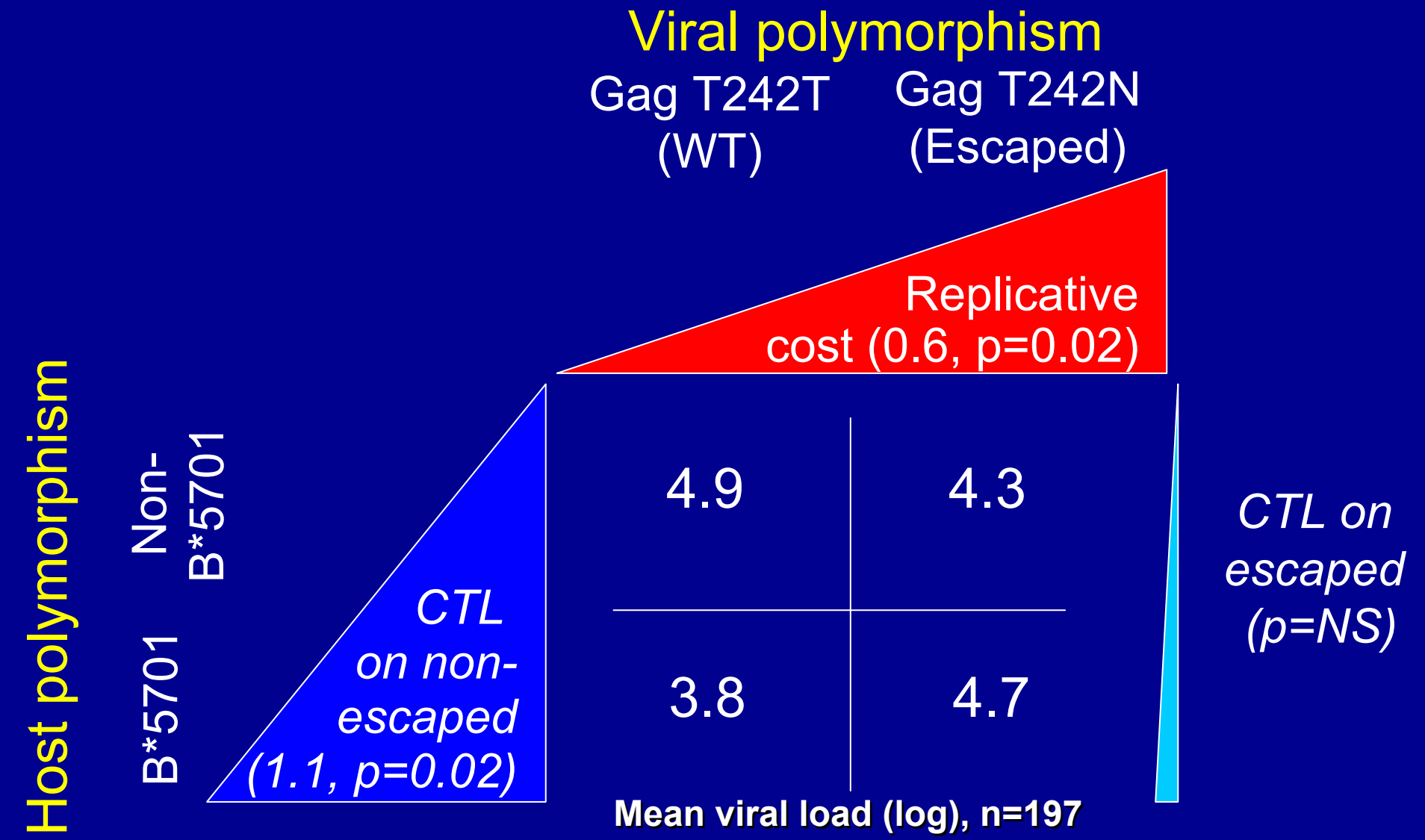
putative epitopes

flanking residues

☐ =Significant after correction

Moore et al Science 2002, 296; 1439

Example

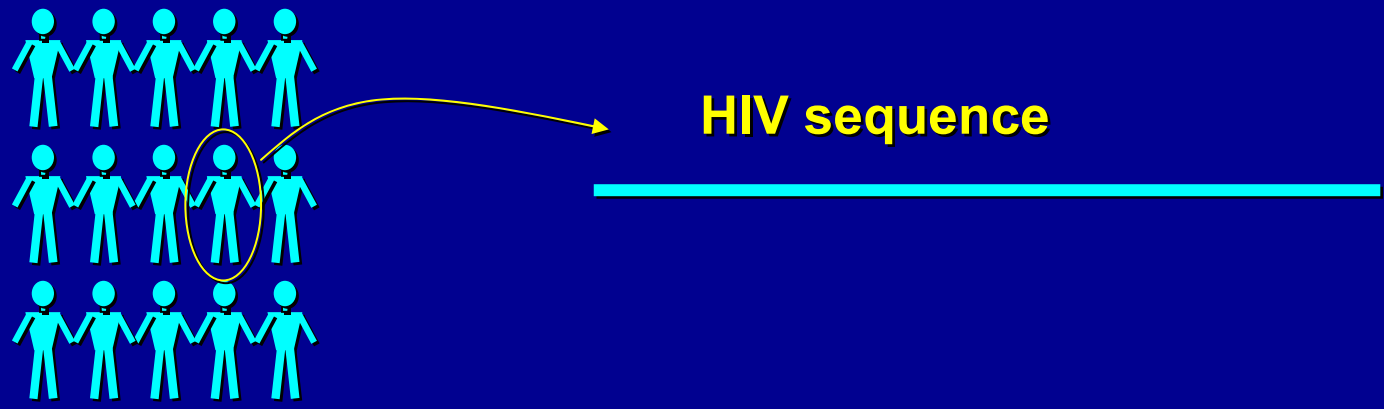

Viral polymorphism
Gag T242T (WT)
Gag T242N (Escaped)
Replicative cost (0.6, p=0.02)
Host polymorphism
B*5701
Non-B*5701
CTL on non-escaped (1.1, p=0.02)
4.9
4.3
3.8
4.7
CTL on escaped (p=NS)
Mean viral load (log), n=197

Predicting new epitopes

- Assume that there is one epitope for every escape found in a population of viral strains.
- For every escape, find the kmer in its vicinity that is most likely to contain the epitope (if the escape had not occurred).
- Replace the escaped amino acid with the non-escaped amino acid to obtain the epitope

Viral strains are 245 Nef sequences from WA HIV Cohort Study with ambiguities resolved by choosing the most likely amino acids at each position.

Predicting new epitopes



Start with escape, replace with non-escape, run window, select most likely

What about the adapted (escaped) kmers?

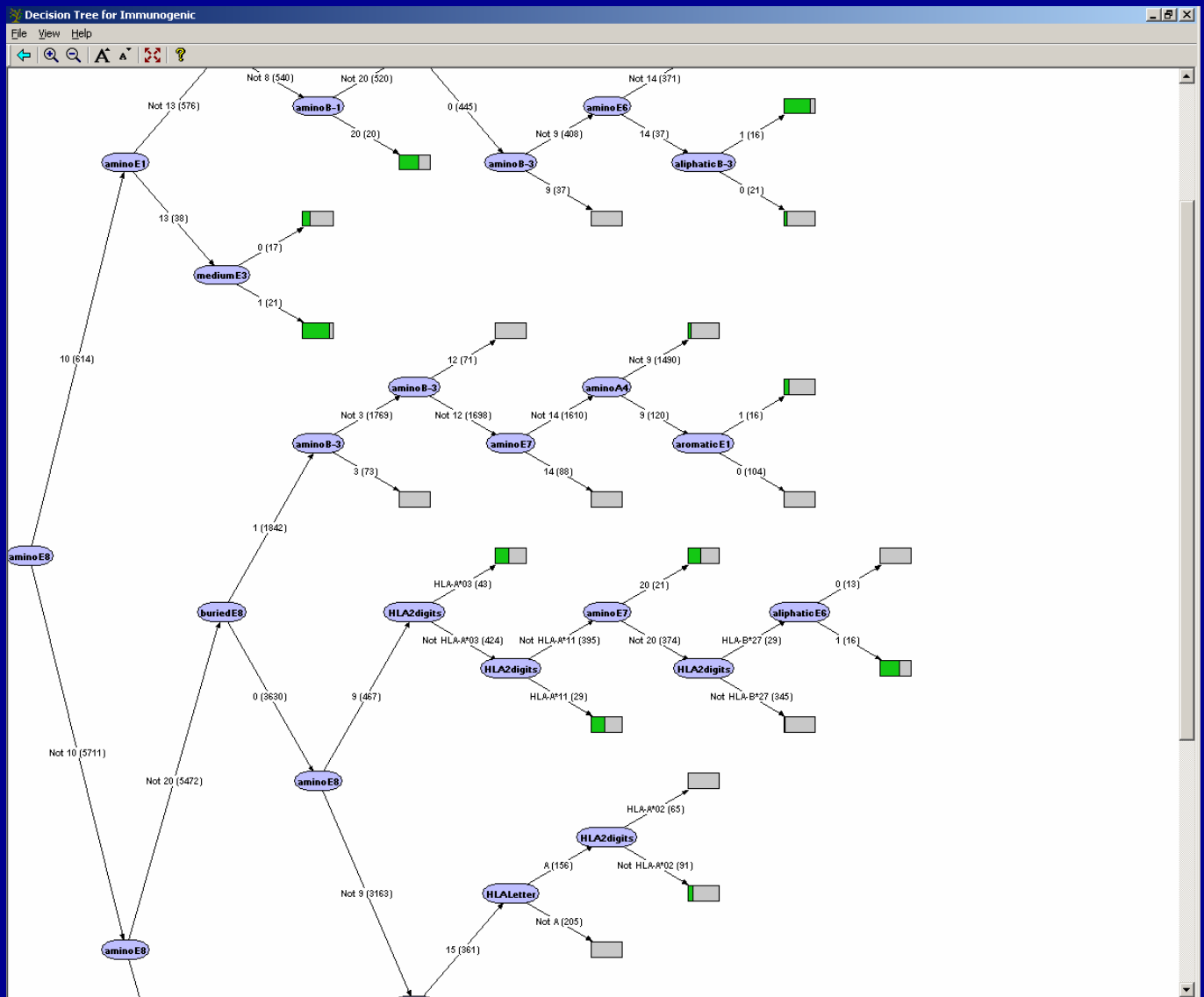
Will responses against the escaped amino acid help clearance of virus?

Our assumption in this work: Use only the non-escaped kmers (focus on protective vaccine)

Prediction Model

- Mixture of decision tree and logistic regression model
- Trained with 8-, 9-, and 10-mers from LANL
- Length 6 flanking regions determined by consensus amino acids from WA HIV Cohort Study

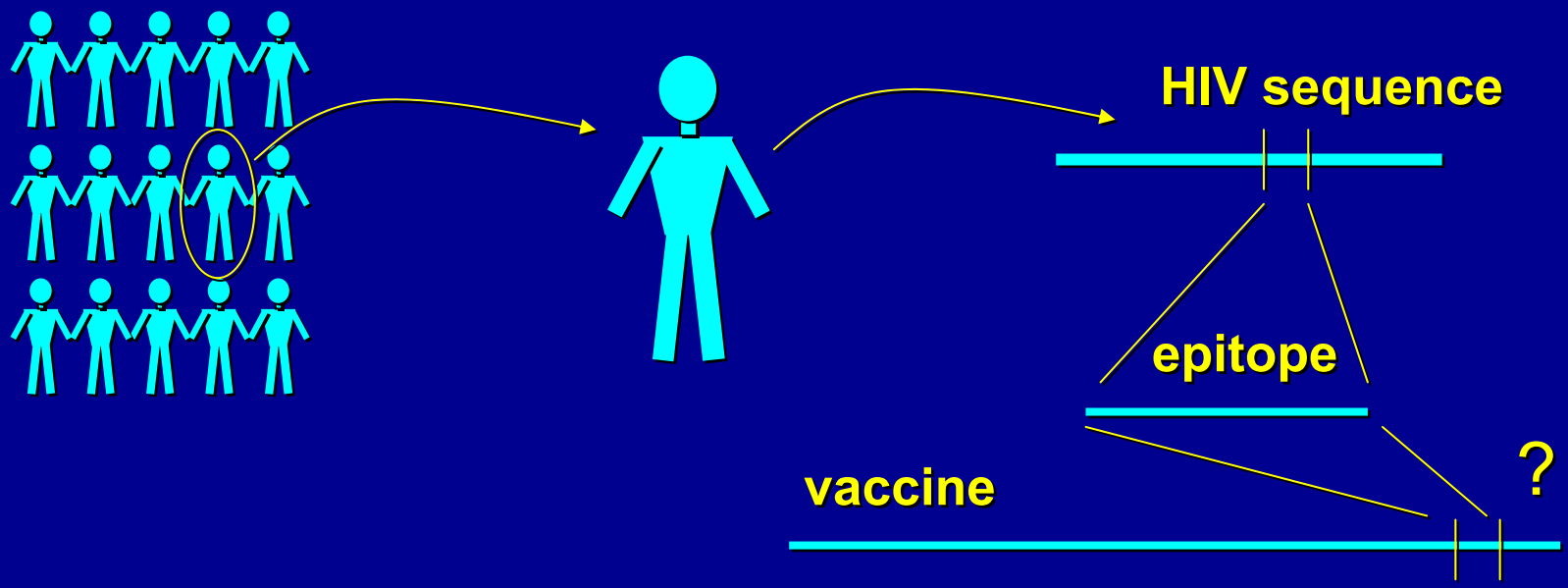
Decision Tree



Vaccine Immunogen Optimization

- Choose optimization criterion (score)
- Construct vaccines that optimize the score for various vaccine lengths
- Plot optimization score vs. length (allows for making tradeoffs)

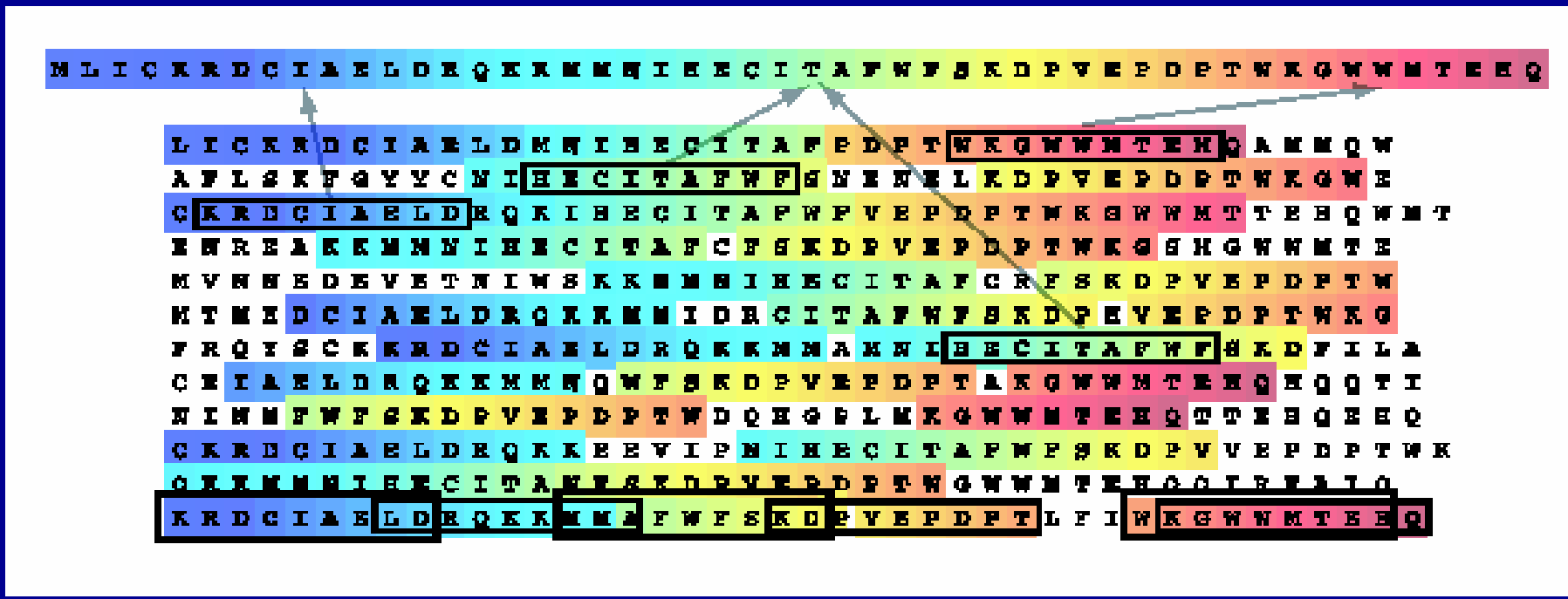
Optimization Score



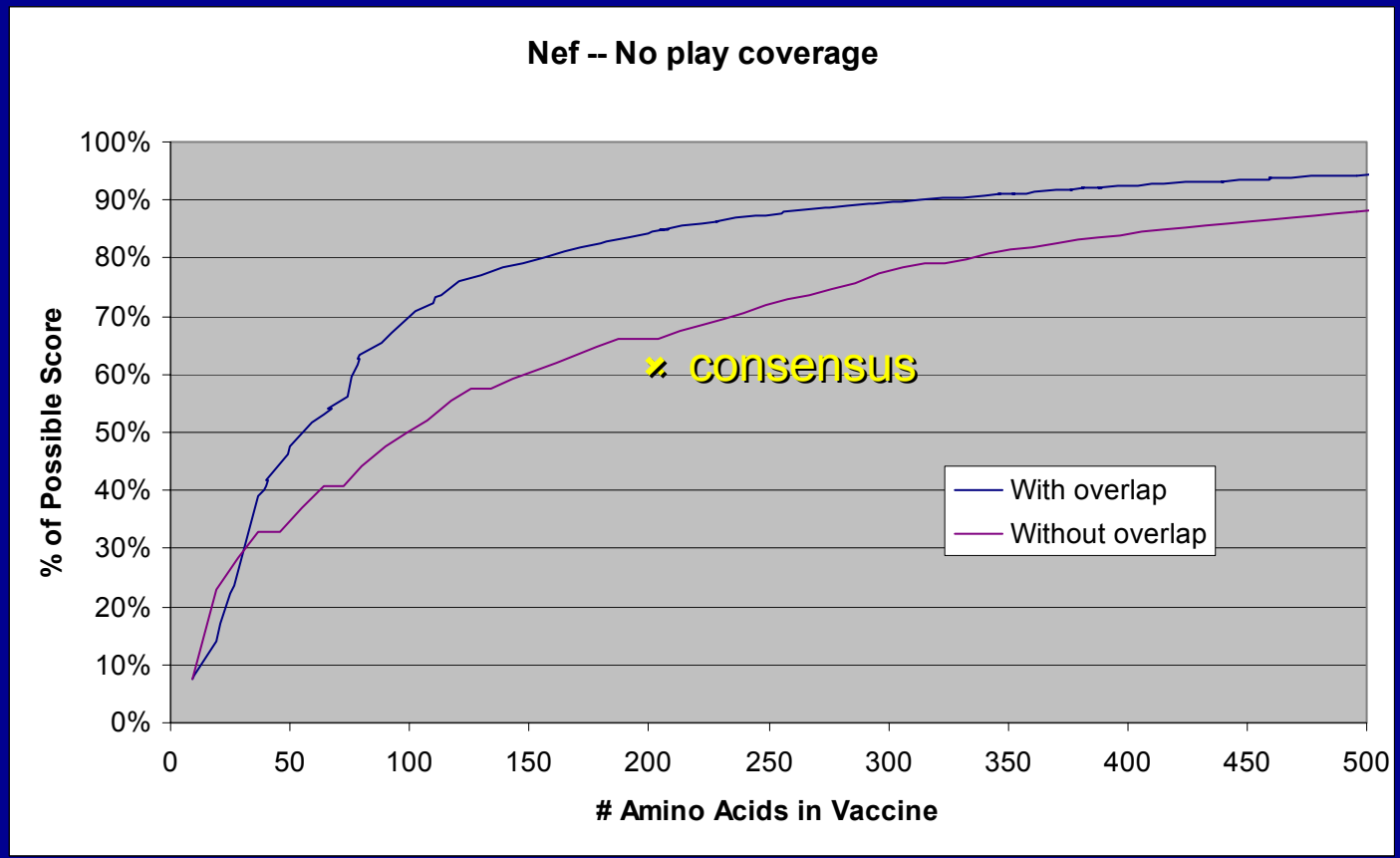
pop-average(fraction of epitopes that are covered)

Viral strains are 245 Nef sequences from WA HIV Cohort Study with ambiguities resolved by choosing the most likely amino acids at each position.

Epitomes: vaccine immunogens that overlap epitopes



Exploiting overlap yields immunogens that cover more epitopes per unit length

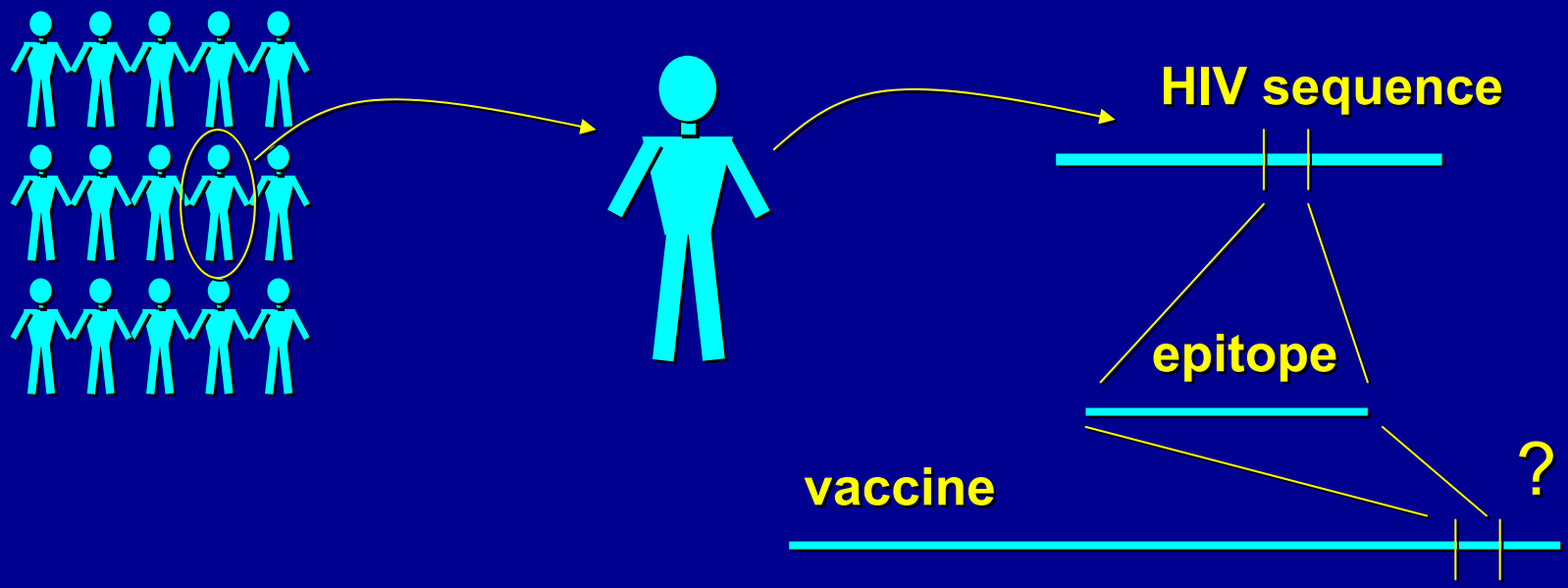


Cross Reactivity

Not much in the literature (e.g., McKinney et al. 2004):

One to two conservative or semi-conservative changes are typically well tolerated (provided the epitope still binds MHC-I)

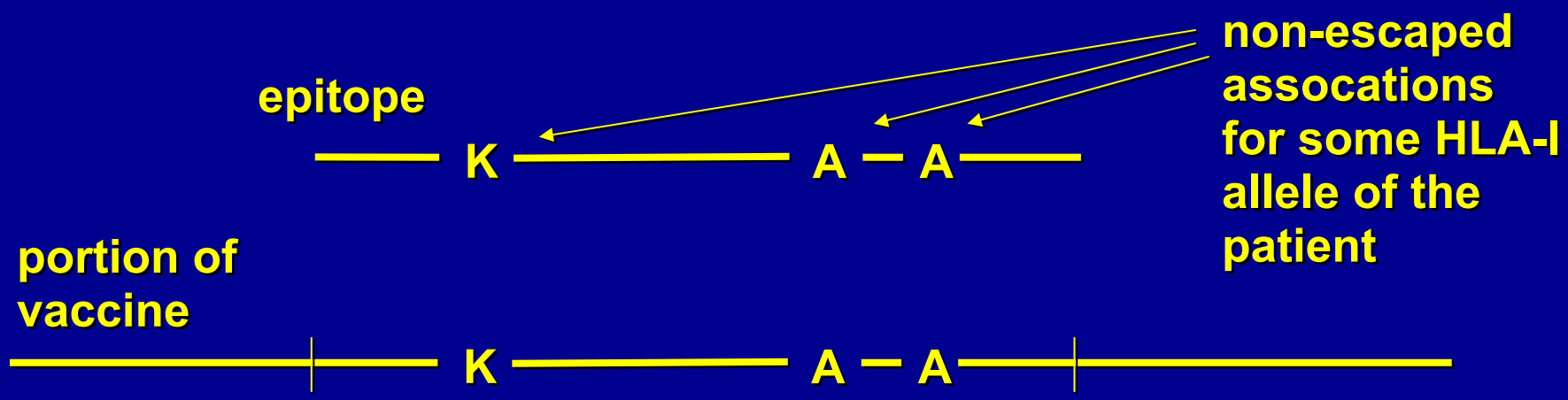
Optimization Score



pop-average(fraction of epitopes that are covered)

Viral strains are 245 Nef sequences from WA HIV Cohort Study with ambiguities resolved by choosing the most likely amino acids at each position.

Models of cross-reactivity



No-play: epitope is covered if, for some HLA of the patient, both epitope and vaccine segment have non-escaped associations and an exact match on every other amino acid

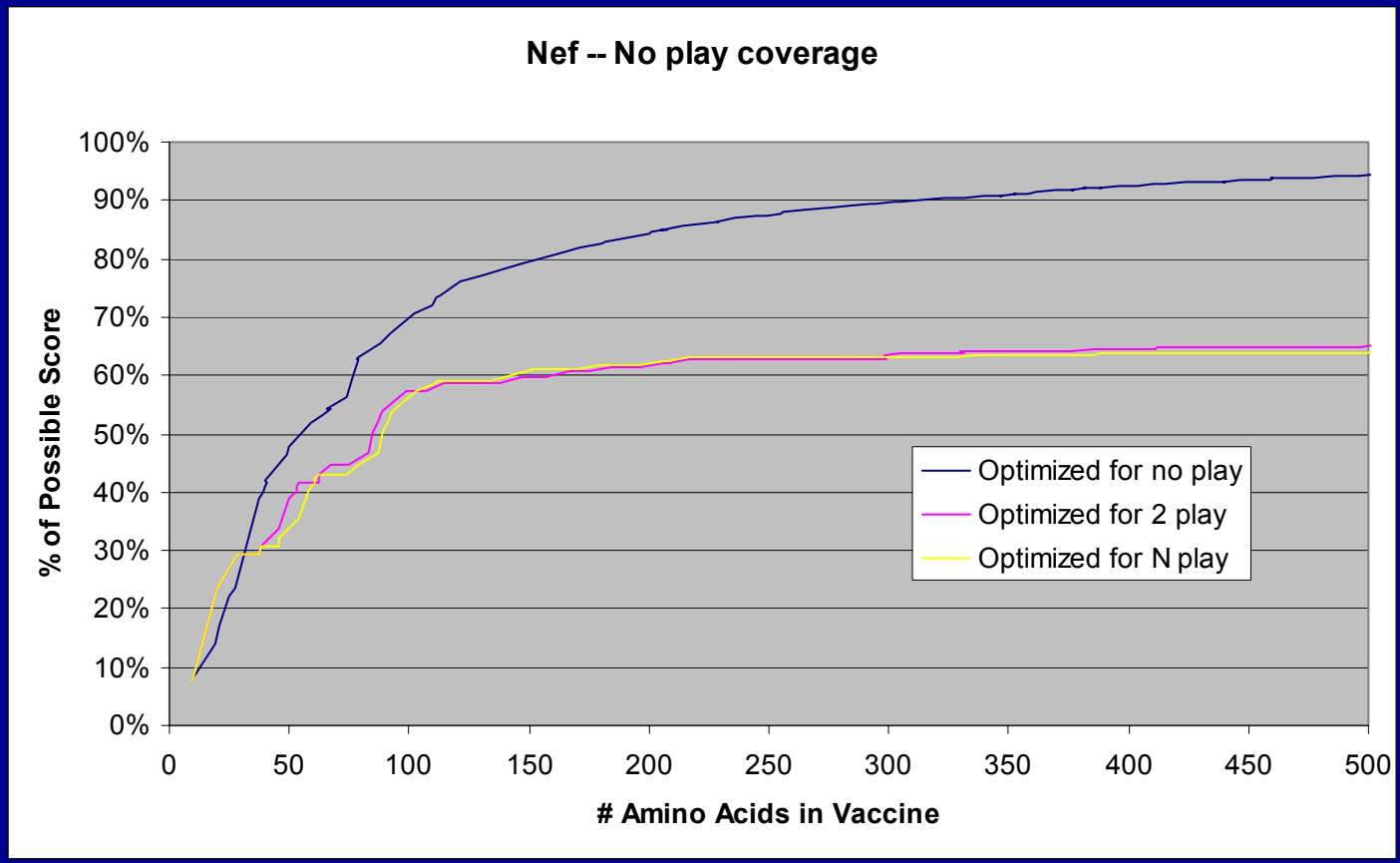
2-play: epitope is covered if, for some HLA of the patient, both epitope and vaccine segment have non-escaped associations and at most two amino acids differ only by conservative amino-acid substitutions elsewhere at non-anchor sites

N-play: epitope is covered if, for some HLA of the patient, both eptiope and vaccine segment have non-escaped associations and differ only by conservative amino-acid substitutions elsewhere at non-anchor sites

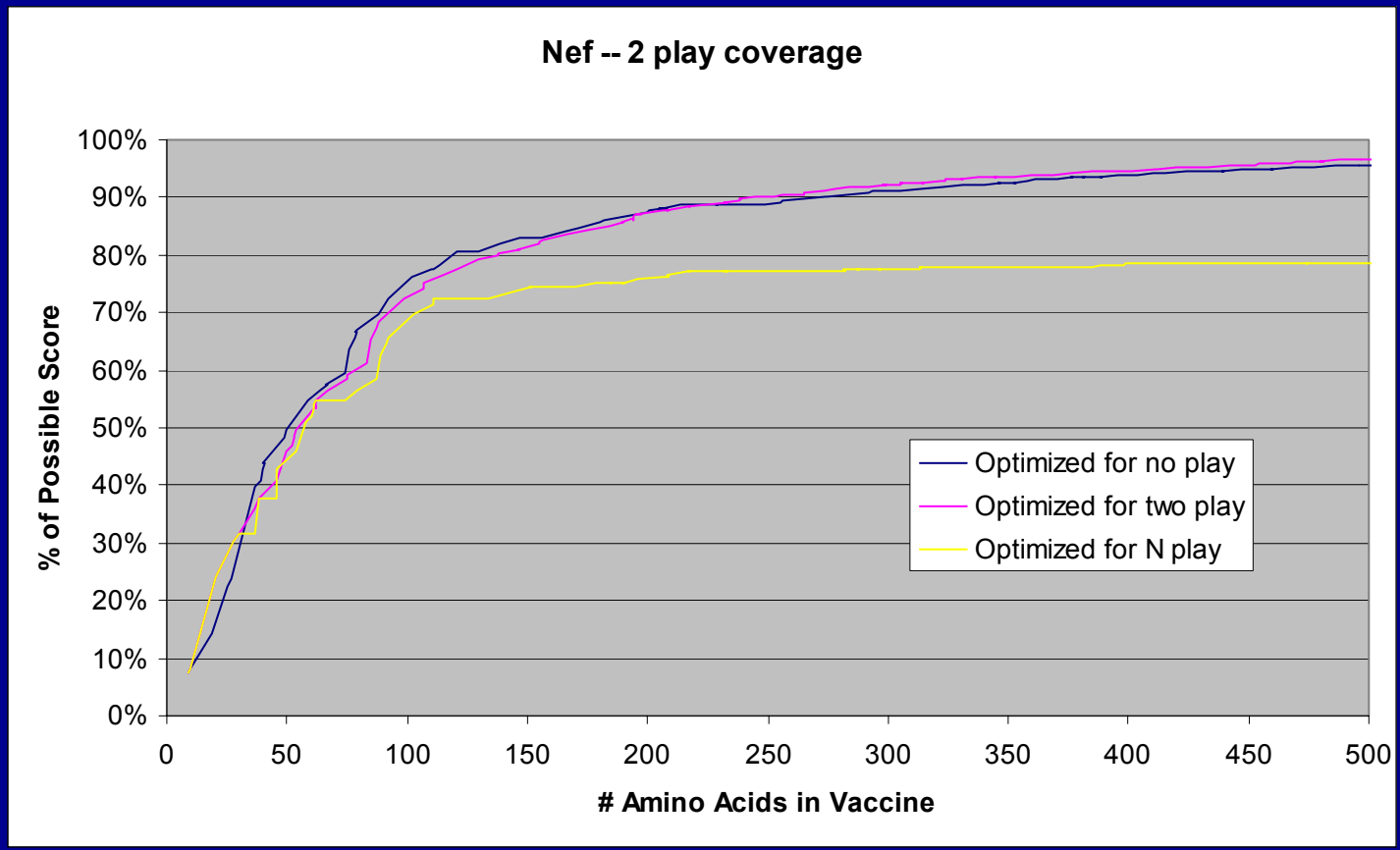
Cross-reactivity: Sensitivity analysis

- Construct vaccines (of various lengths) to optimize for each model of cross reactivity
- Score each set of vaccines according to each of the cross reactivity models
- For each combination of optimization criterion and score, plot coverage vs. length
- Is there a model of cross reactivity that produces a good vaccine regardless of which model is true?

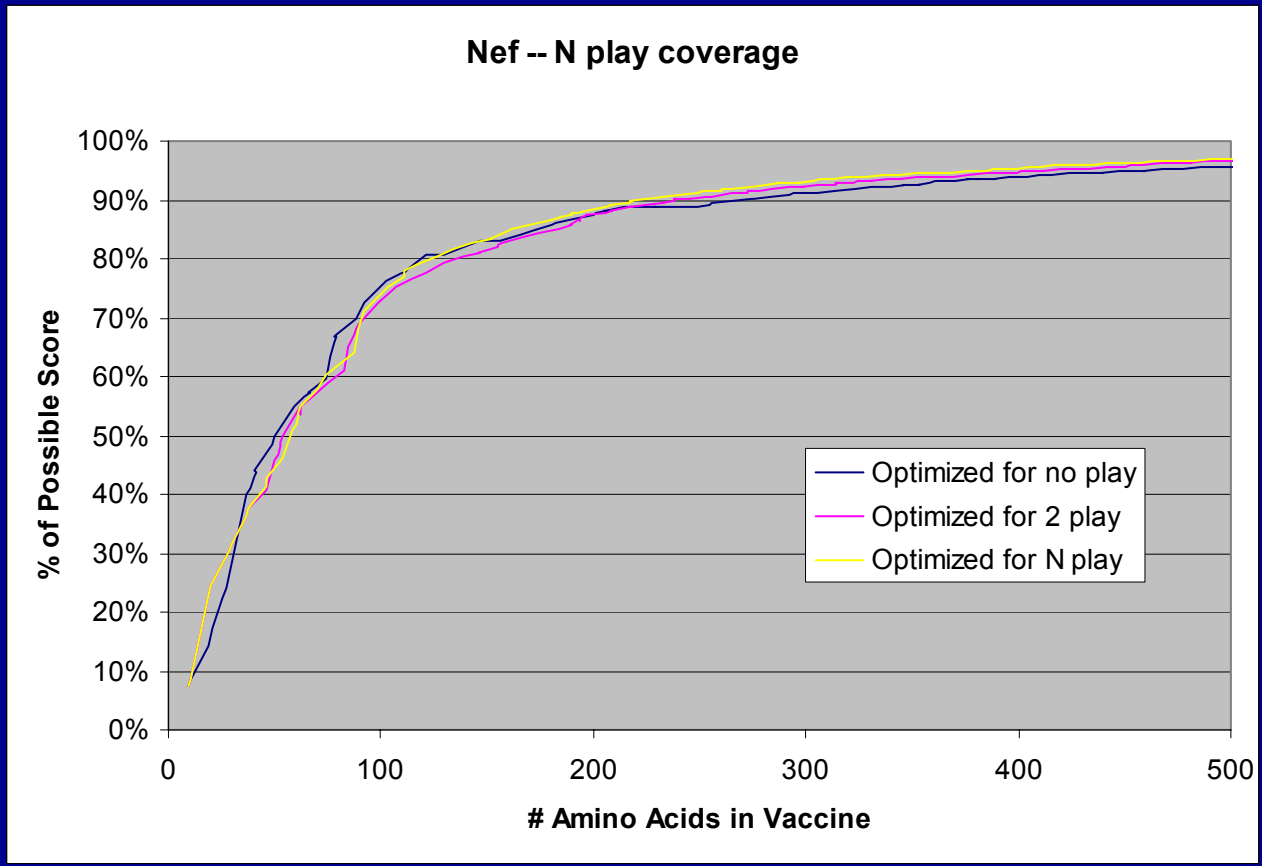
Cross-reactivity: Sensitivity analysis



Cross-reactivity: Sensitivity analysis



Cross-reactivity: Sensitivity analysis



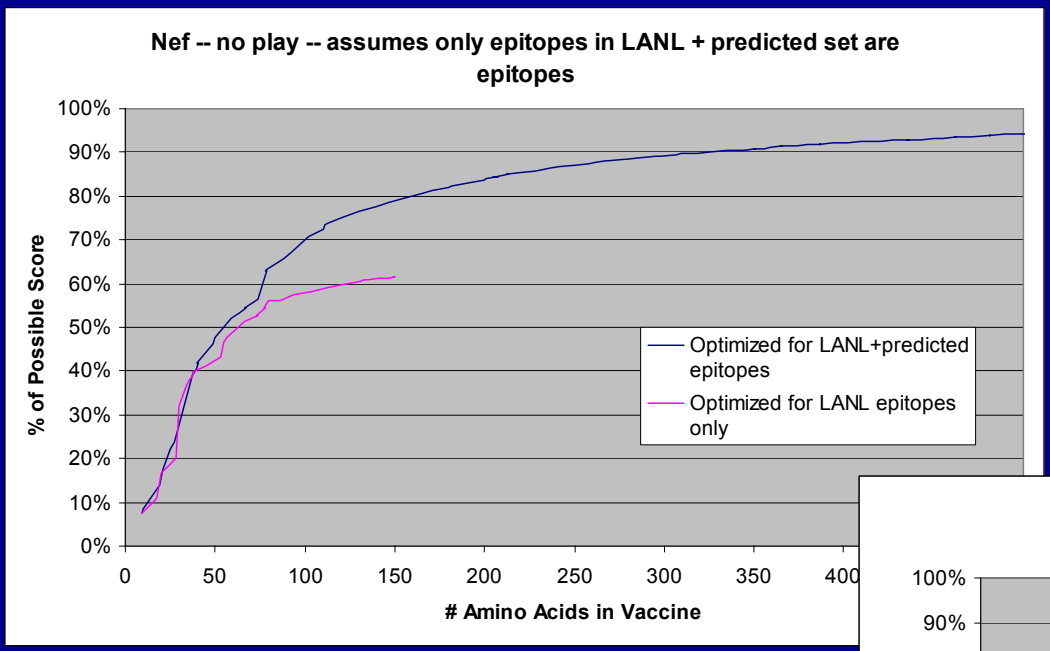
Vaccine optimized for no-play model does well by all three cross-reactivity models.

Number of epitopes: Sensitivity analysis

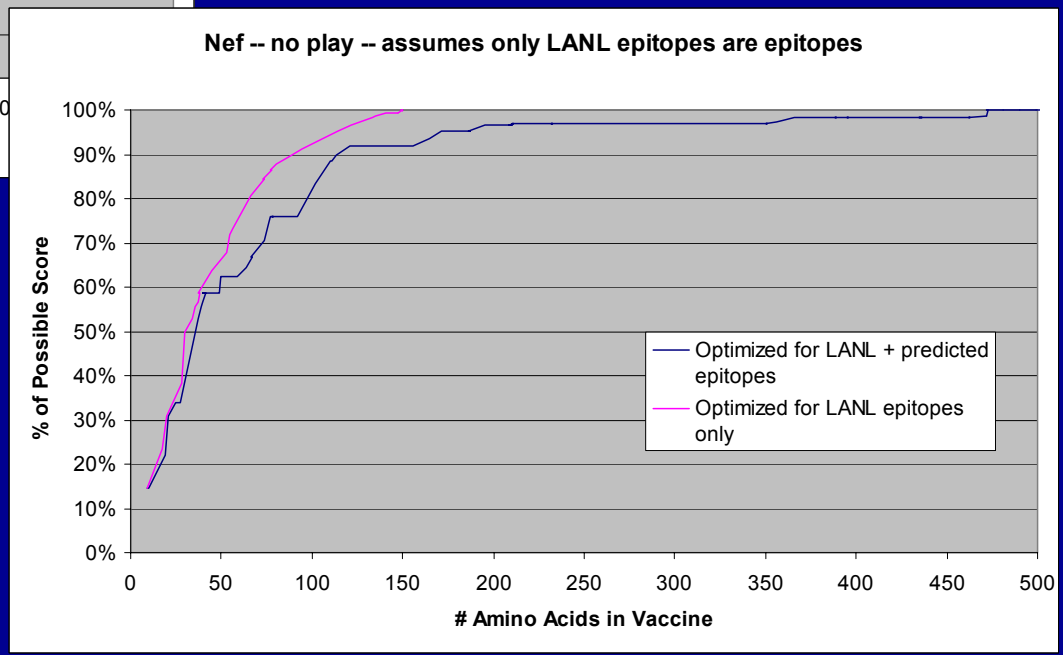
- Baseline: LANL+ predicted
- Less: LANL only
- More: All 9mers are epitopes

Is there a vaccine optimized for one of these that does well regardless of the true number?

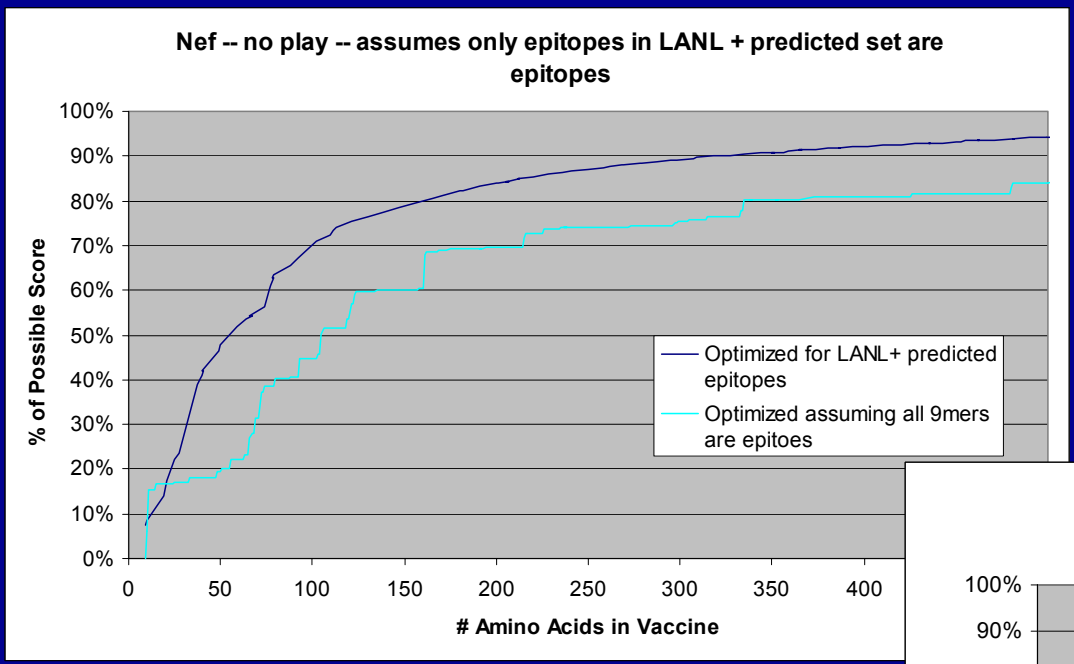
What if there are fewer epitopes?



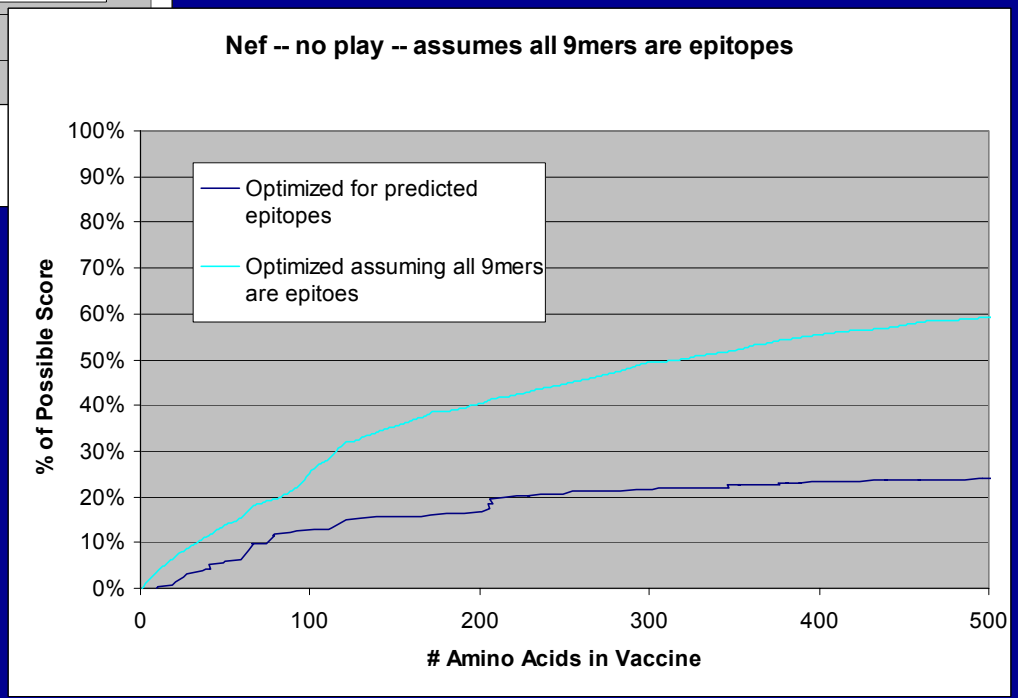
Fewer epitopes



What if there are more epitopes?

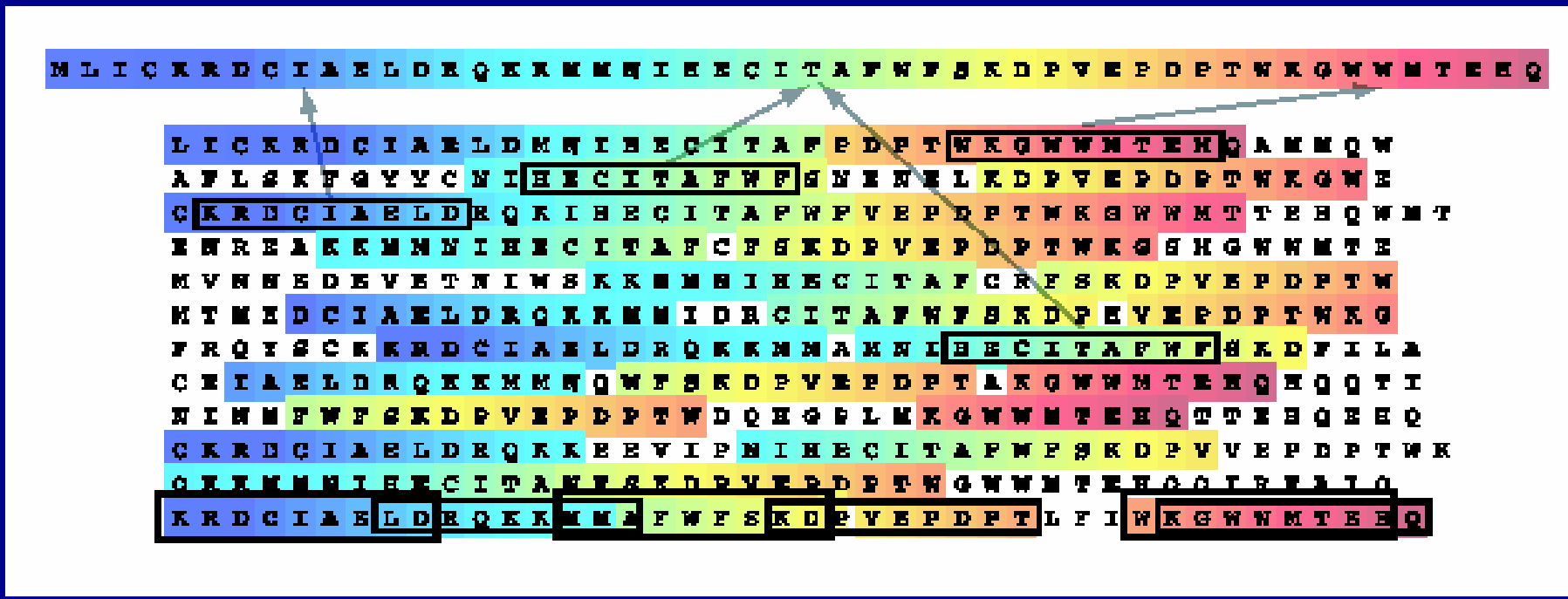


More epitopes

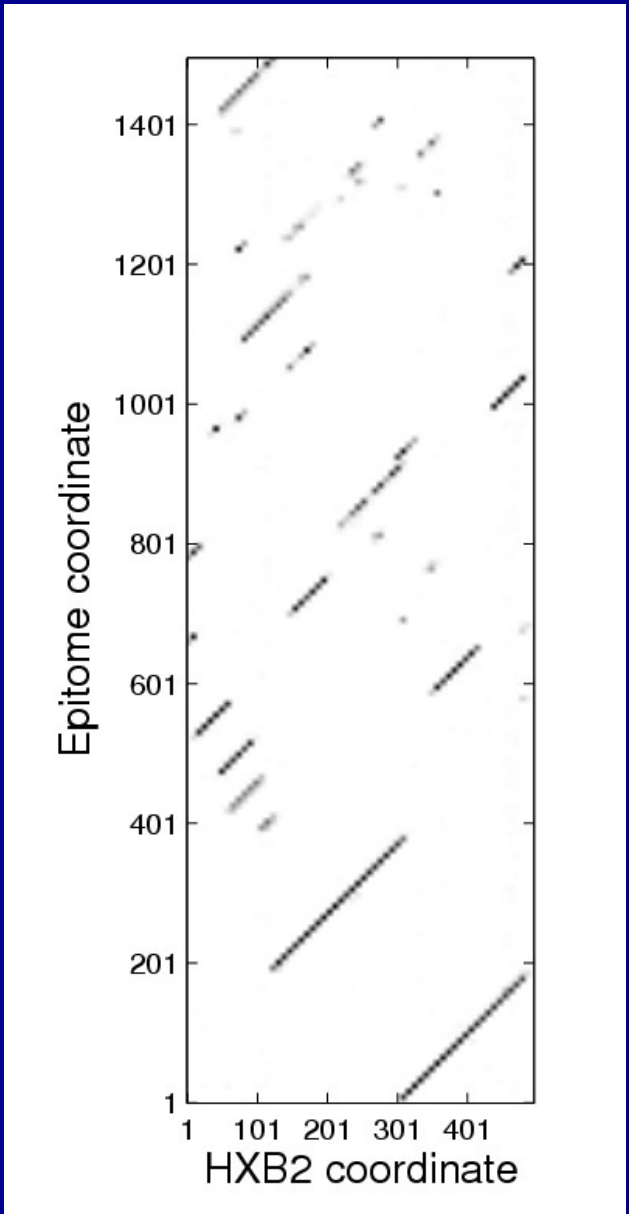


If uncertain, should err in favor of more epitopes (overlap provides some robustness)

Epitomes: vaccine immunogens that overlap epitopes



Gag epitome organization

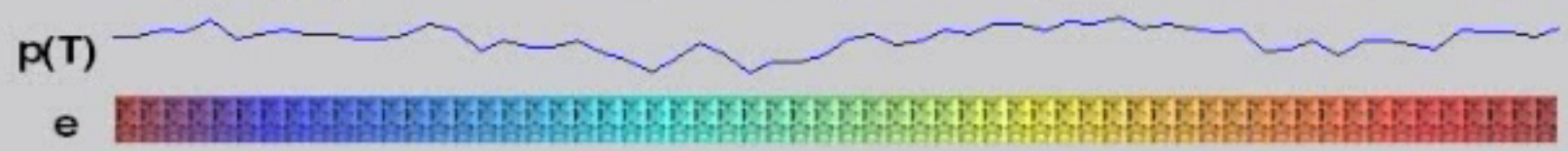


Each 9mer in each strain is matched to the closest matching positions in HXB2 and the epitome, and a dot is placed at these coordinates in the graph.

This provides an ***alignment*** of the epitome to the HXB2. As could be seen by the darkness of the dots, the early part of the epitome is matched to a lot, and contains similar content as the consensus sequence. The tail of the epitome contains less and less frequent variations.

(Note: the darkness of the dot depends on how many 9mers match there and how well they match)

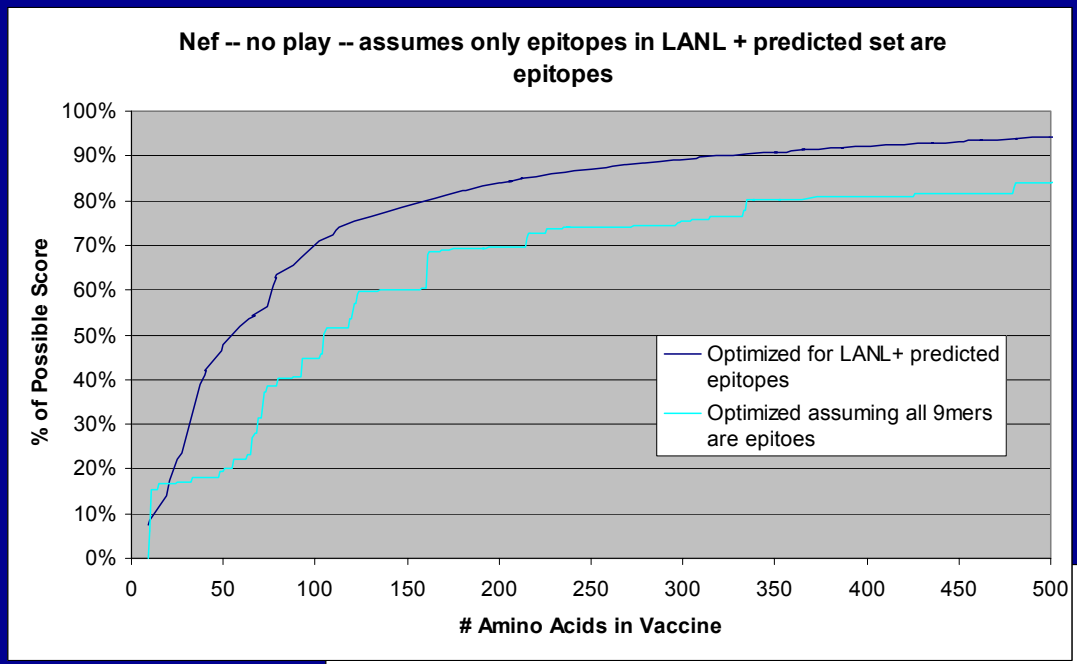
Learning epitome as a sequence of probability distributions over letters



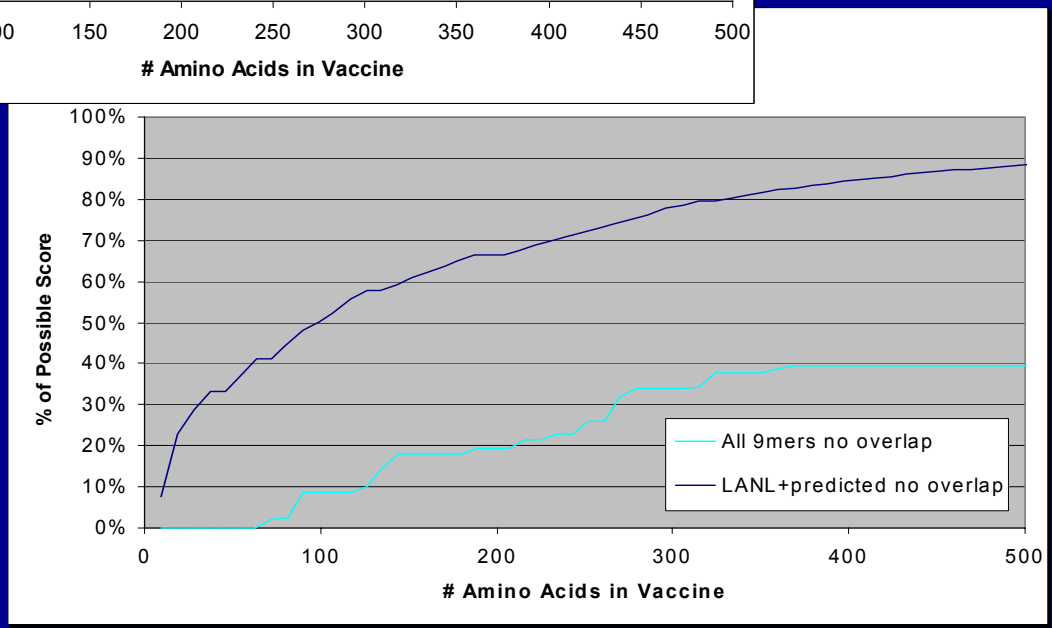
Data color-coded according to the current epitome mapping Q(T)

TFDCIAELDRQKKMMNMECITAFWFSKDPVEPTWKGWFFT
ELICKFDCIAELDRESVVDWPDLTWKGWWFTEHHEHQCQHG
QKKMMNIHECIRFWFSKDPVSPDWKGWWFTEHQHQQKQVTHQ
CCIAELDRQKKMHECITAFWFSKDPVEPDLTWKGWWMTP
MNIHECITAFWFTFKDPVSPDPTWKGWWFTEHQIWECQPS
HLDRQKKGMNIHDPVEPDPTWKGWWMKMTEHQGQQEHQCGPF
SELDRQKKMMNIHEAFWFTKDPVEPDPWKGWWFTEHQEEHQ
VQAELDRQGKMMNAFWFTKDPVSPDLTMTQWKGWWFTEHQRL
EICKRDCIAELDANFEECITAFWFSKDPVEPDPTWKGWFM
DCIAELDRQKKMMMDNIHECITAFWFTKDILAPVEPDPTWKG

What if there are more epitopes?



overlap

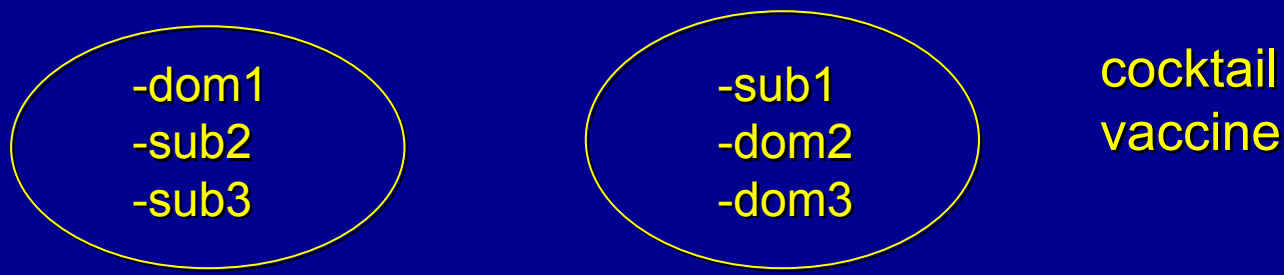


no overlap

Immunodominance

Rodriguez et al. 2002: If dominant and subdominant epitopes are delivered separately, then responses to both are elicited

Generalization to vaccine (covering numerous patients):



Becomes an HLA-sensitive optimization problem

Summary

- HLA-driven adaptation in HIV is evident at population level as associations between HLA alleles and HIV sequence
- Such HLA associations can be used to predict immunogenic epitopes in HIV in-vivo
- We have optimized HIV vaccine immunogen sequences for maximum coverage of all (non-adapted) epitopes by exploiting epitope overlap and knowledge of adaptation effects
- Assuming no cross reactivity yields near-optimal vaccine even if there is some cross-reactivity
- If uncertain about the identity of the epitopes, it is better to err in favor of more epitopes. The more the epitopes overlap, the less important it becomes to know their identity (and the less they overlap, the shorter the vaccine)