

**Predicting the Evolution of Influenza  
HA1 (H3N2) Hemagglutinin**

**Walter M. Fitch & Robin M. Bush**

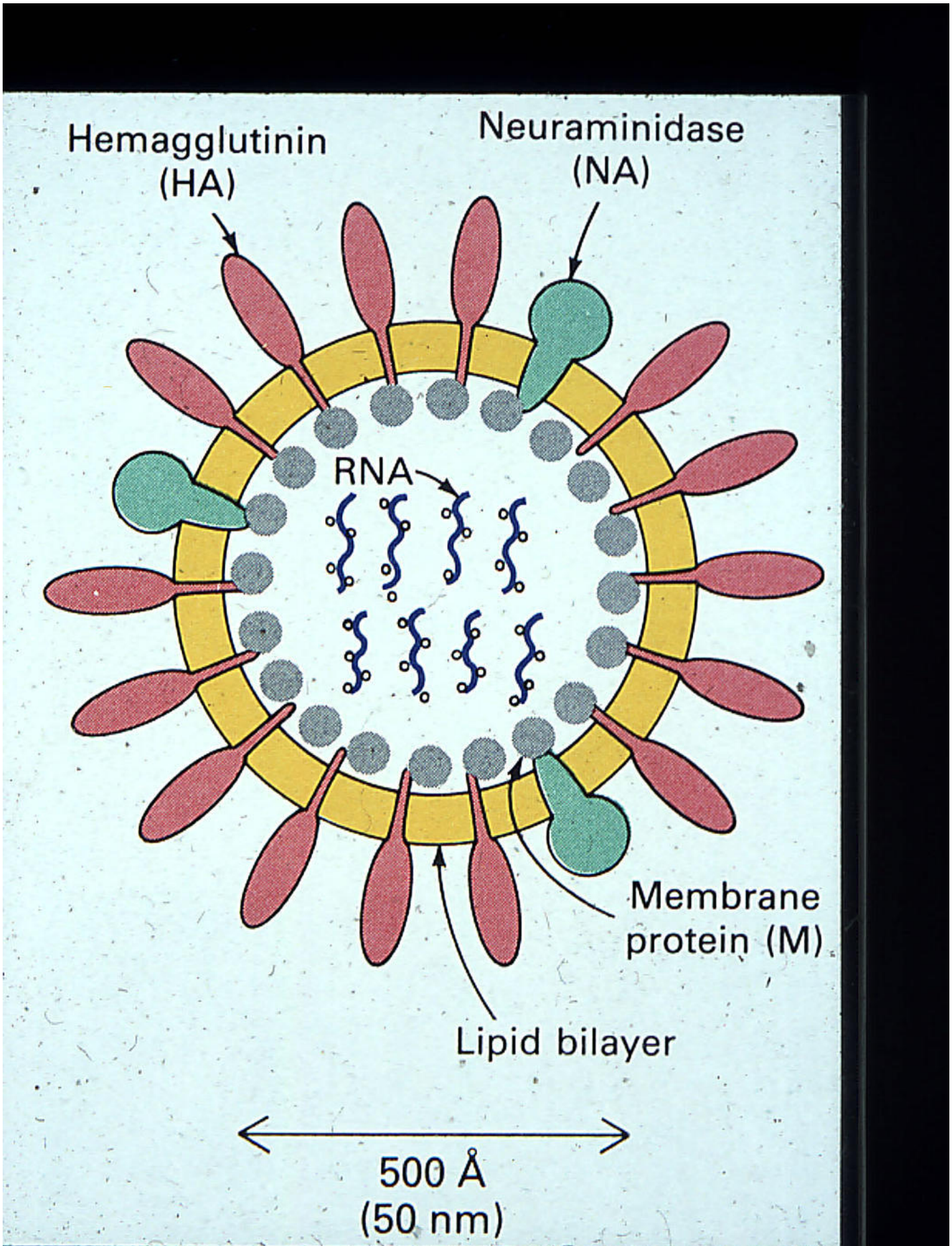
Department of Ecology and Evolutionary  
Biology  
University of California, Irvine

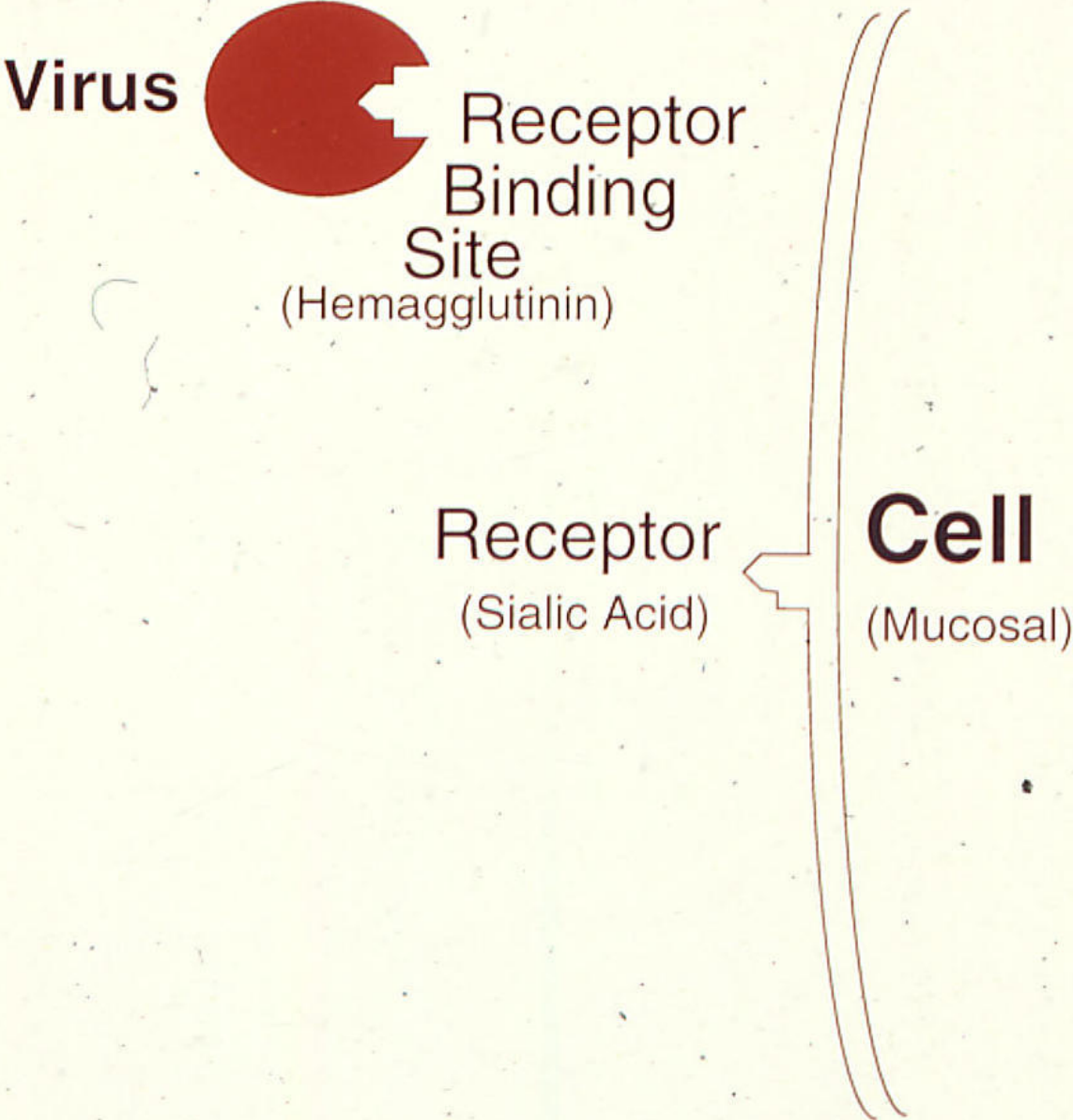
**Catherine A. Bender & Nancy J. Cox**

Influenza Branch  
Centers for Disease Control and Prevention

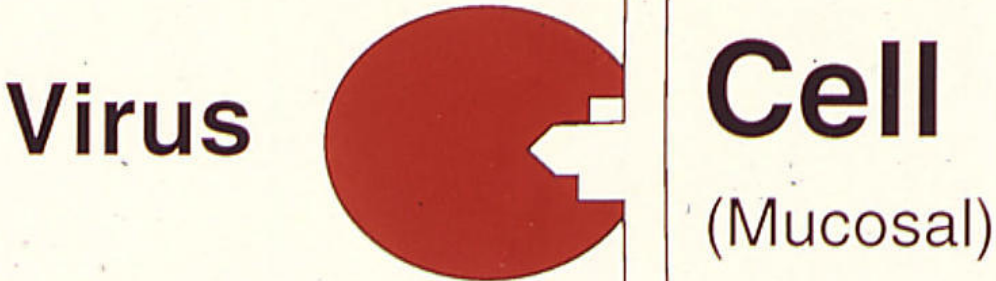


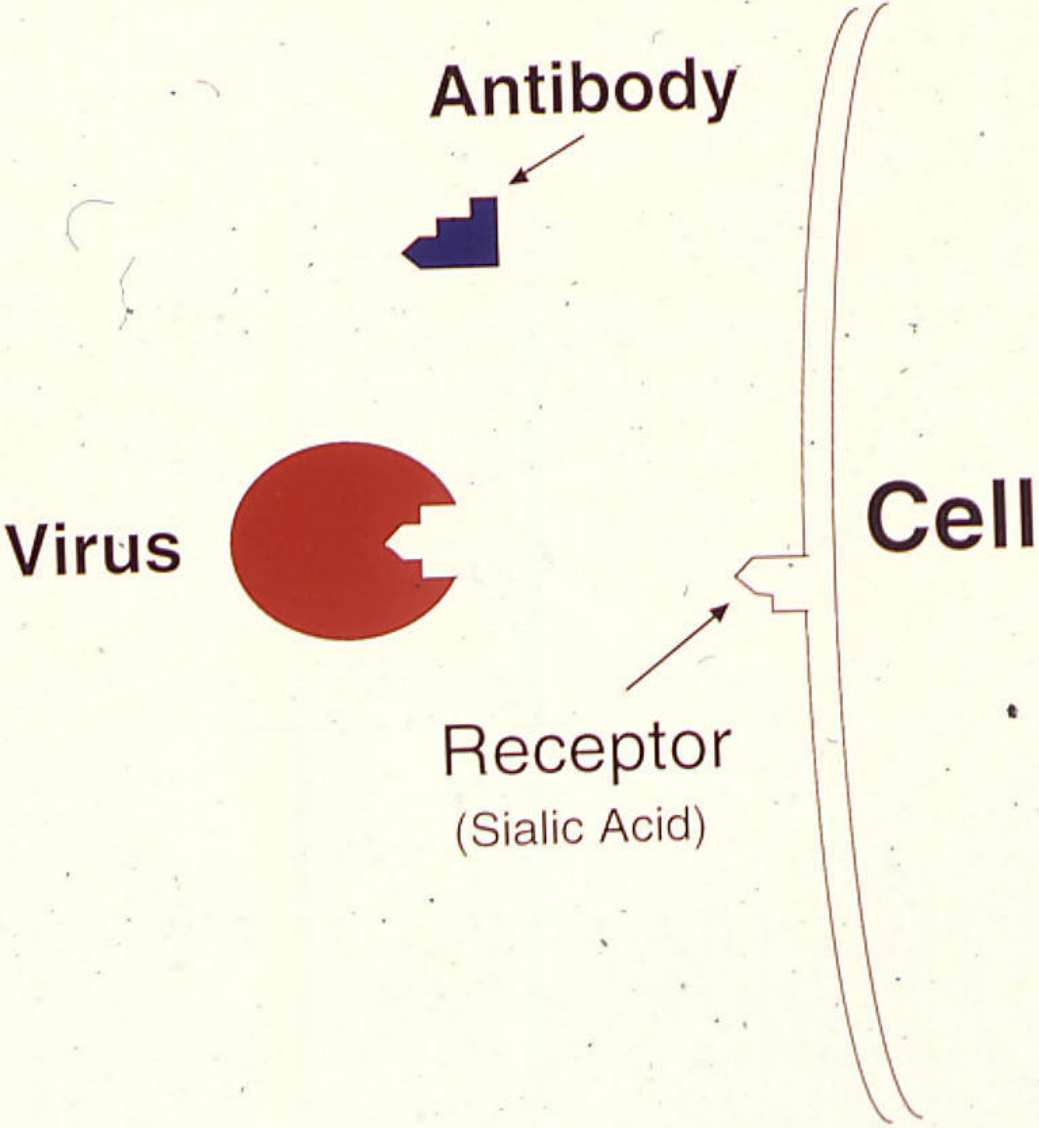


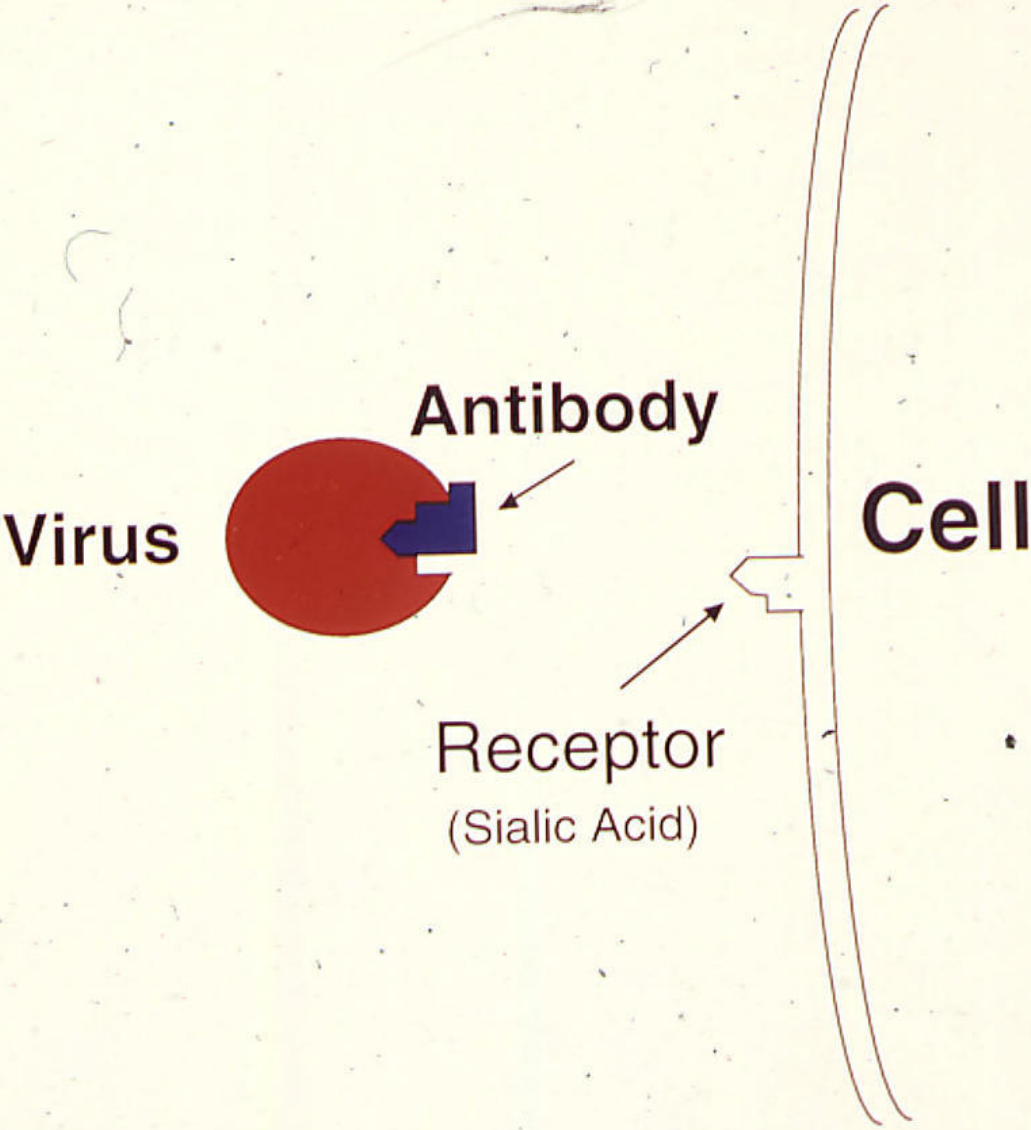


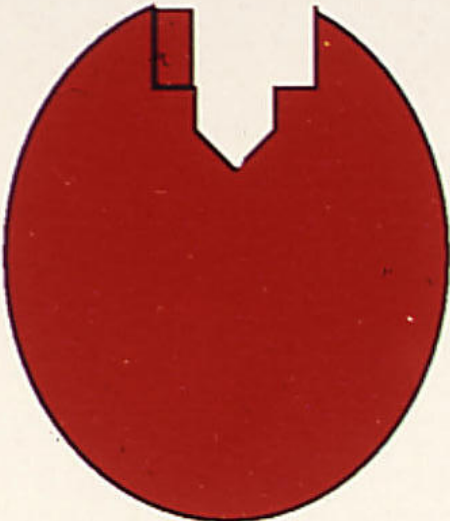




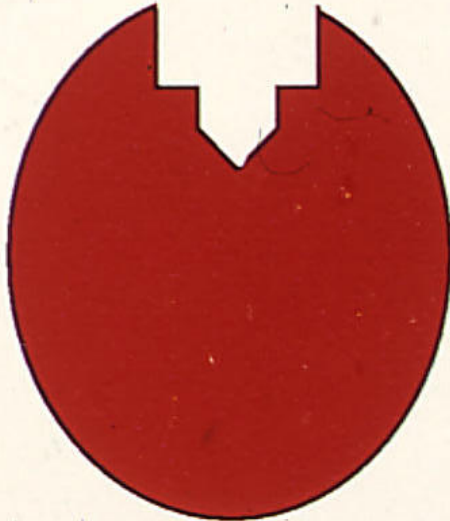








**Mutated  
Virus**



**Virus**

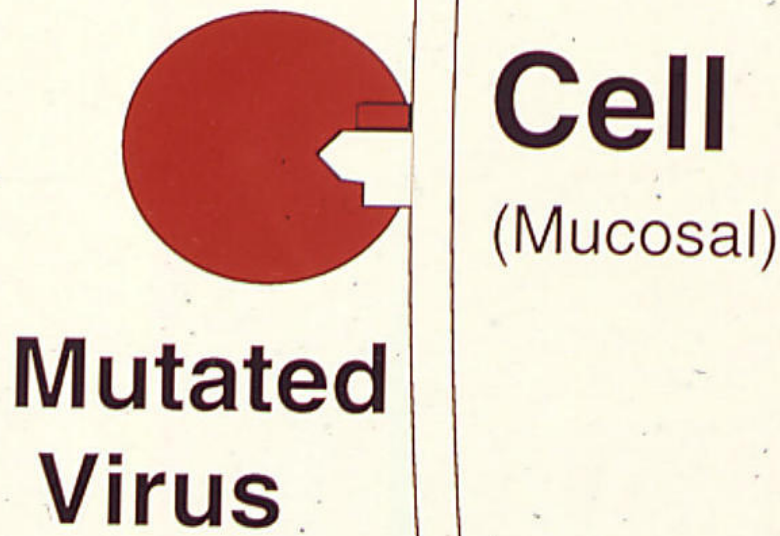


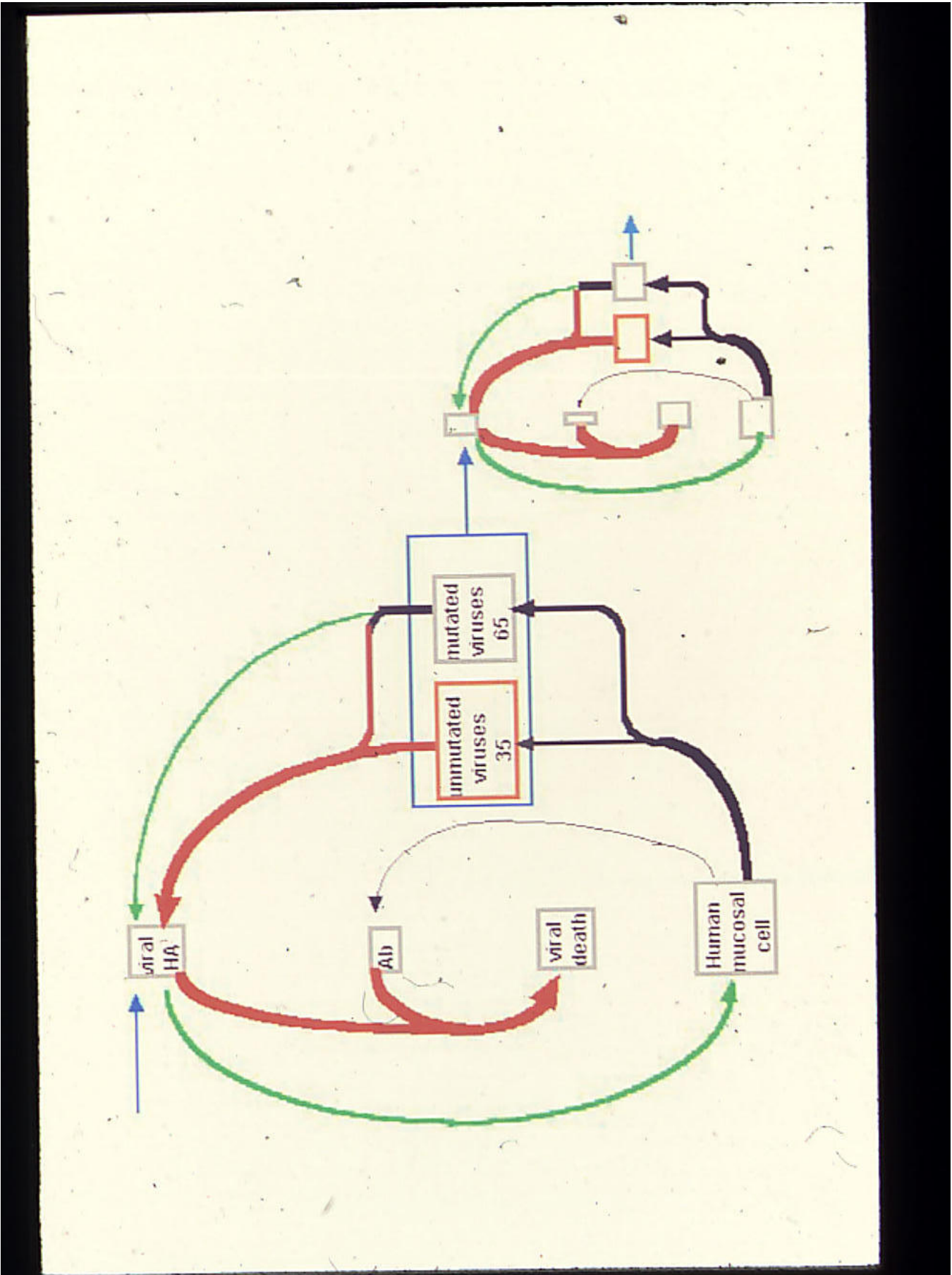


**Antibody**

**Mutated**

**Virus**







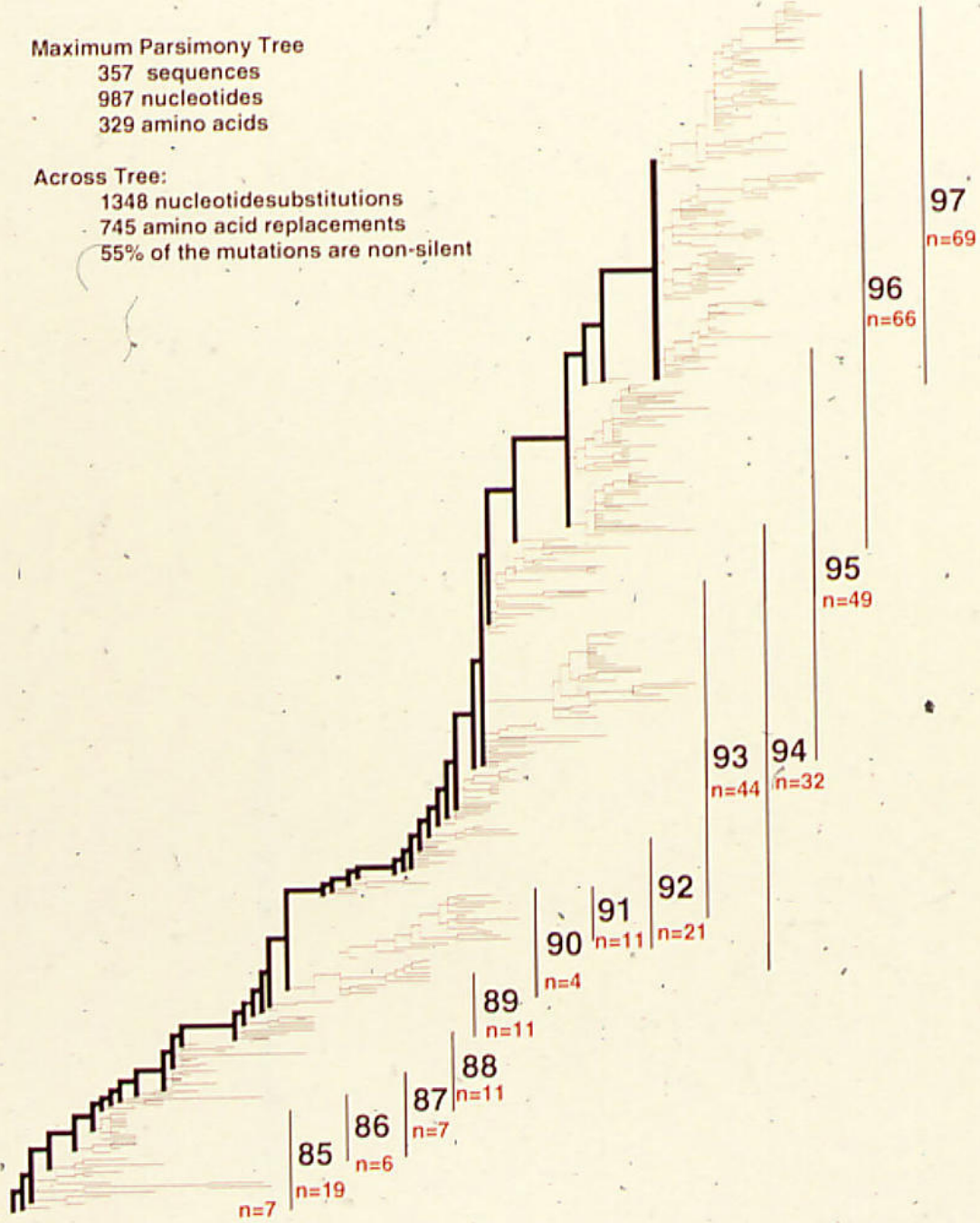
Number of (unrooted) trees, T,  
for s sequences

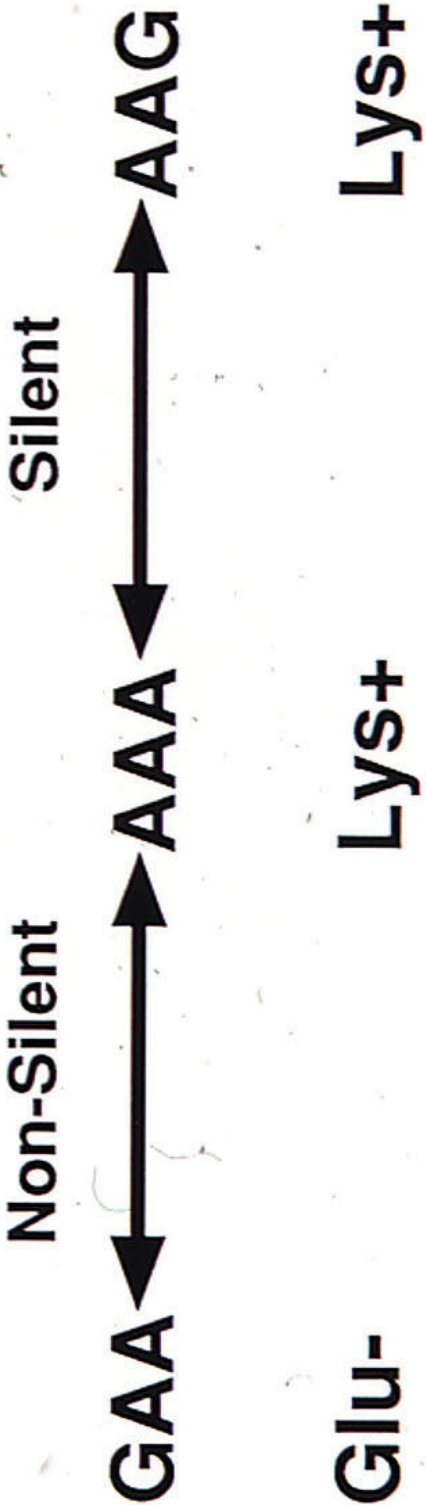
s	T
3	1
4	3
5	15
6	105
7	945
8	10,395
9	135,135
10	2,027,025
11	34,459,425
12	654,729,075
13	$1.37 \times 10^{10}$
14	$3.16 \times 10^{11}$
.	
..	
20	$3.19 \times 10^{23}$

$$T = \prod_{i=1}^{s-2} (2i - 1)$$

Maximum Parsimony Tree  
357 sequences  
987 nucleotides  
329 amino acids

Across Tree:  
1348 nucleotidesubstitutions  
745 amino acid replacements  
55% of the mutations are non-silent







What is probability of getting exactly  
eight non-silent mutations in  
examining ten mutations?

$$P = a! p^n q^s / n! s!$$

$$P = 10! (0.45)^8 (0.55)^2 / (8! 2!) = 0.023$$

**Positively selected codons offtips**

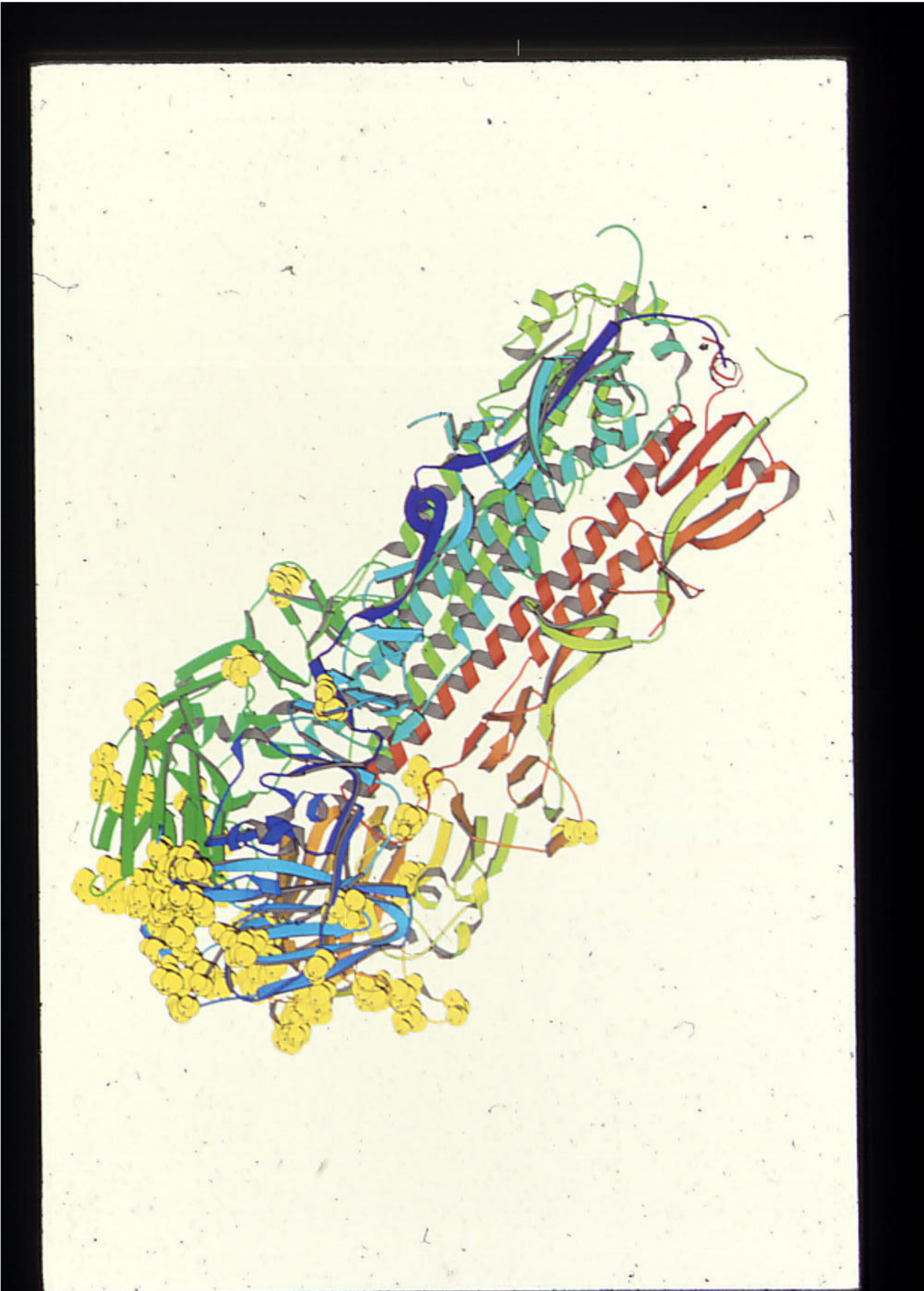
	pos	all	non	sil	prob	ifsig=1
1:	42*	3	3	0	0.0361	1
2:	56*	5	4	1	0.0400	1
3:	73	6	6	0	0.0013	1
4:	75	9	9	0	0.0000	1
5:	76*	3	3	0	0.0361	1
6:	121	7	6	1	0.0061	1
7:	127*	3	3	0	0.0361	1
8:	129*	7	5	2	0.0371	1
9:	136	8	6	2	0.0164	1
10:	137	6	5	1	0.0158	1
11:	146*	3	3	0	0.0361	1
12:	149	11	10	1	0.0001	1
13:	150	6	6	0	0.0013	1
14:	154	4	4	0	0.0119	1
15:	162*	5	4	1	0.0400	1
16:	164	4	4	0	0.0119	1
17:	165	12	8	4	0.0142	1
18:	175*	3	3	0	0.0361	1
19:	183	4	4	0	0.0119	1
20:	197	15	15	0	0.0000	1
21:	198	6	6	0	0.0013	1
22:	199	19	18	1	0.0000	1
23:	202	5	5	0	0.0039	1
24:	203	4	4	0	0.0119	1
25:	235*	3	3	0	0.0361	1
26:	267	4	4	0	0.0119	1
27:	293	4	4	0	0.0119	1

The eighteen positions of H3 hemagglutinins from human influenza viruses under positive selection on the off-tip branches of the tree

Position	NS/S	Probability	Position	NS/S	Probability
121	5/0	0.0135	186	9/1	0.0025
124	5/0	0.0135	190	4/0	0.0320
133	8/0	0.0010	193	4/0	0.0320
135	5/0	0.0135	194	4/0	0.0320
138	6/0	0.0057	197	4/0	0.0320
142	4/0	0.0320	201	4/0	0.0320
145	8/0	0.0010	226	20/1	<0.0001
156	9/1	0.0025	262	4/0	0.0320
158	5/0	0.0135	275	6/0	0.0057

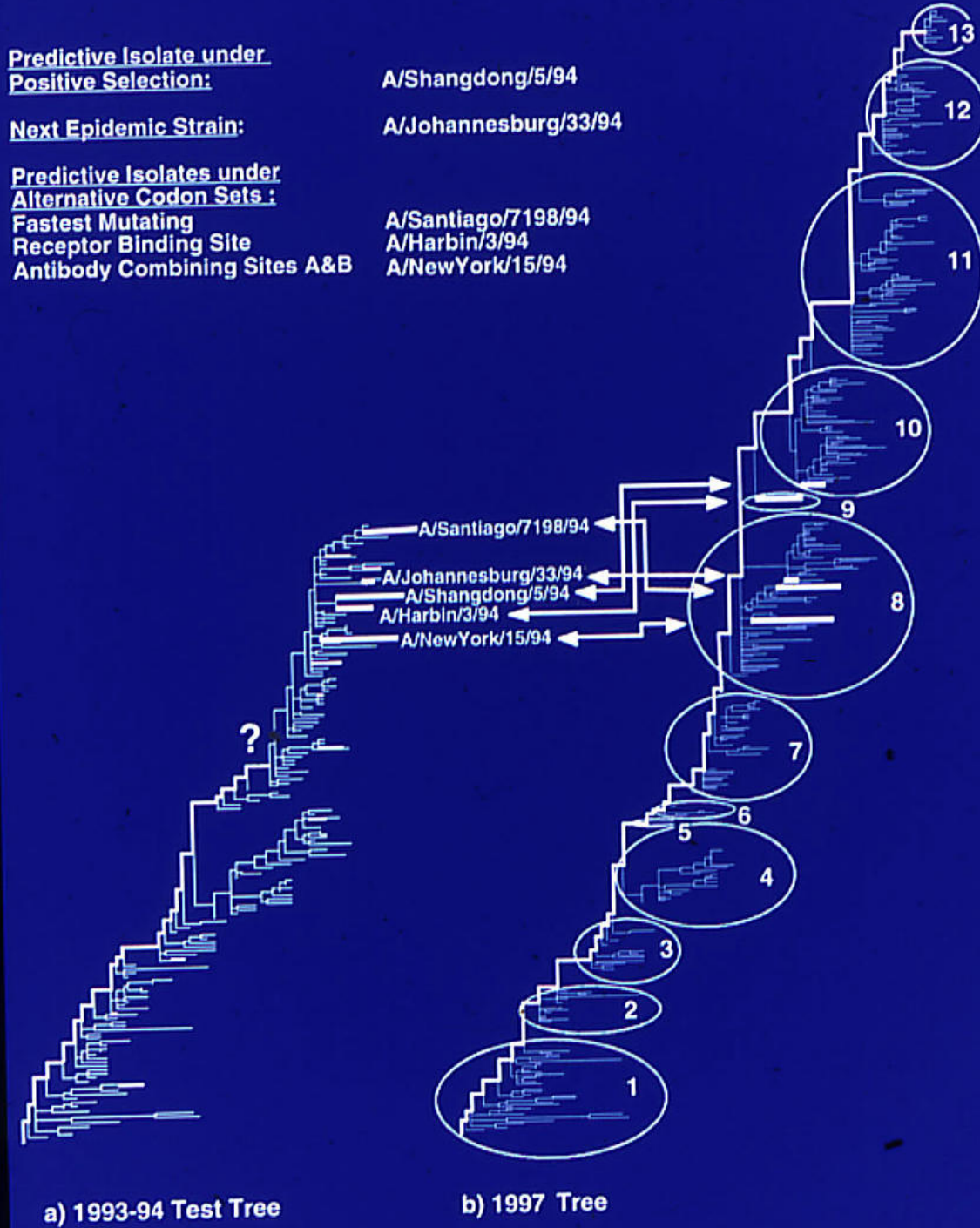
Position is the amino acid position in the protein. NS/S is the number of non-silent over the number of silent substitutions. Probability is the binomial probability that NS/S would be observed by chance. The letter t indicates that the position also had a significant excess of non-silent changes in the tip branches.



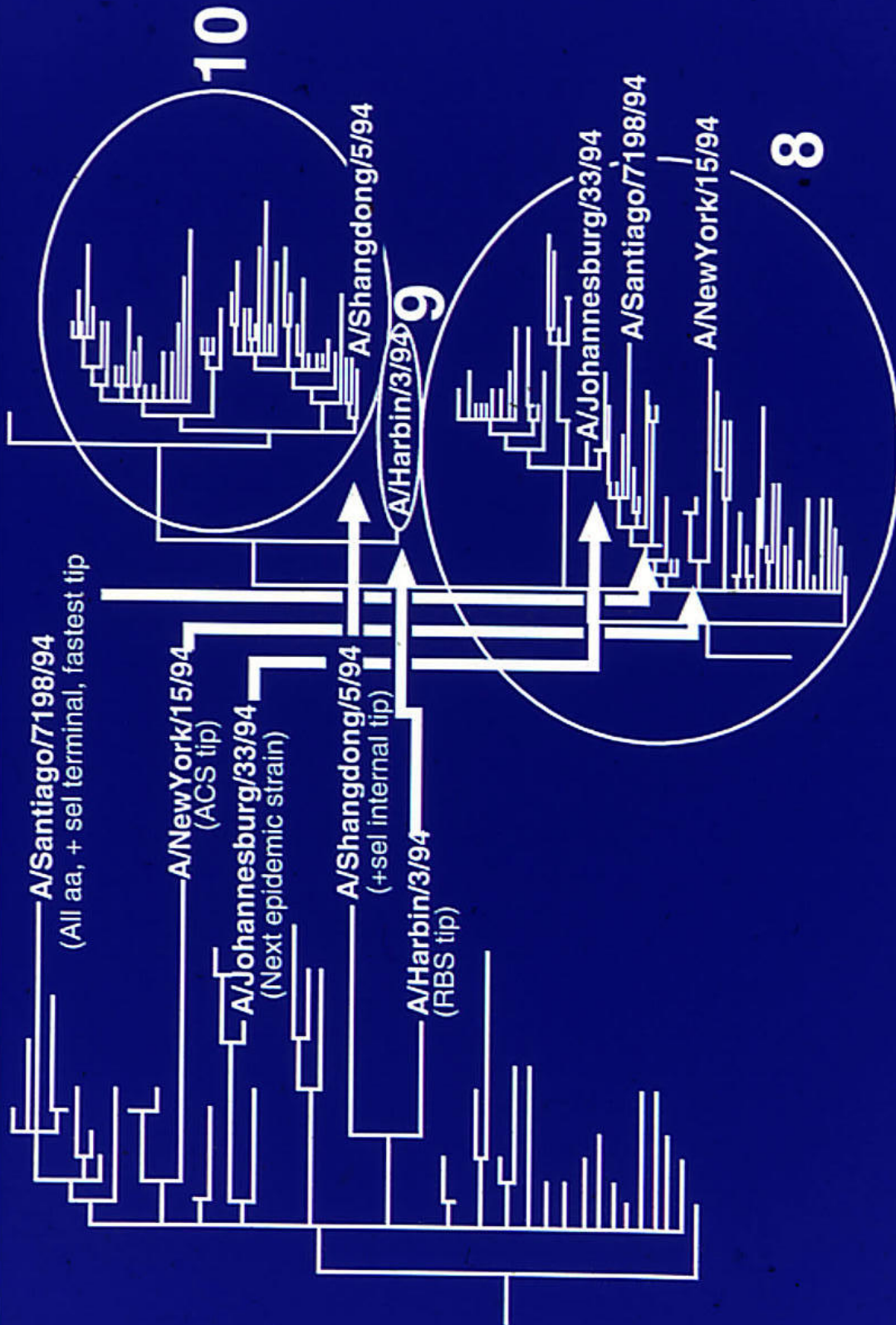


## Subsets of Positions used as Predictors

- Under Positive Selection using Off-Tip Changes Only
- Under Positive Selection using Tip Changes Only
- Under Positive Selection using All Changes
- RBS: Receptor Binding Site Positions
- Antigenic Site Positions
- Positions that undergo Host-Mediated Change (egg)
- Randomly Chosen Positions







Top of 93-94 Test Tree with Predictive Tips      Predictive Tips Mapped on to 1997 Tree

**TRUNK GROUP PICKED AS FUNCTION OF FLU-YEAR AND CODON SET**

Codon Sets	86-7	87-8	88-9	89-90	90-1	91-2	92-3	93-4	94-5	95-6	96-7	Upper
PosSel N=18	2	2	2	6	6	7	8	10	11	12	13	9
AB N=41	2	2	3	6	4,6	7	8	8	8,10	12	13	7
AB-Pos N=23	2	2	2,3	2,3	2,3	4,7	4,7,8	8	10	10	10	1
CDE N=90	1	1	3	3	3	6	7	8	8	11,12	12	1
RBS N=16	1,2	2	3	3	3	7	8	9	10,11	12	13	5
Fast N=20	1	1	3	4,6	4,6	7	8	8	11	12	13	6
Fast-Pos N=18	1	1	3	3	3	3	7	7,8	7,8	10	12	1
N isolates	39	50	61	65	76	97	141	173	222	288	357	
%Random ≥PosSel	4.4	24.6	49.0	0.8	1.3	14.0	10.5	0.6	0.1	1.4	3.3	
%Isolates in upper group	15.4	2.0	19.7	1.5	1.3	8.2	14.2	0.6	1.8	6.2	3.9	

