

# Using CamGrid to model the evolution of Influenza

David Burke  
Antigenic Cartography Group  
Department of Zoology  
University of Cambridge

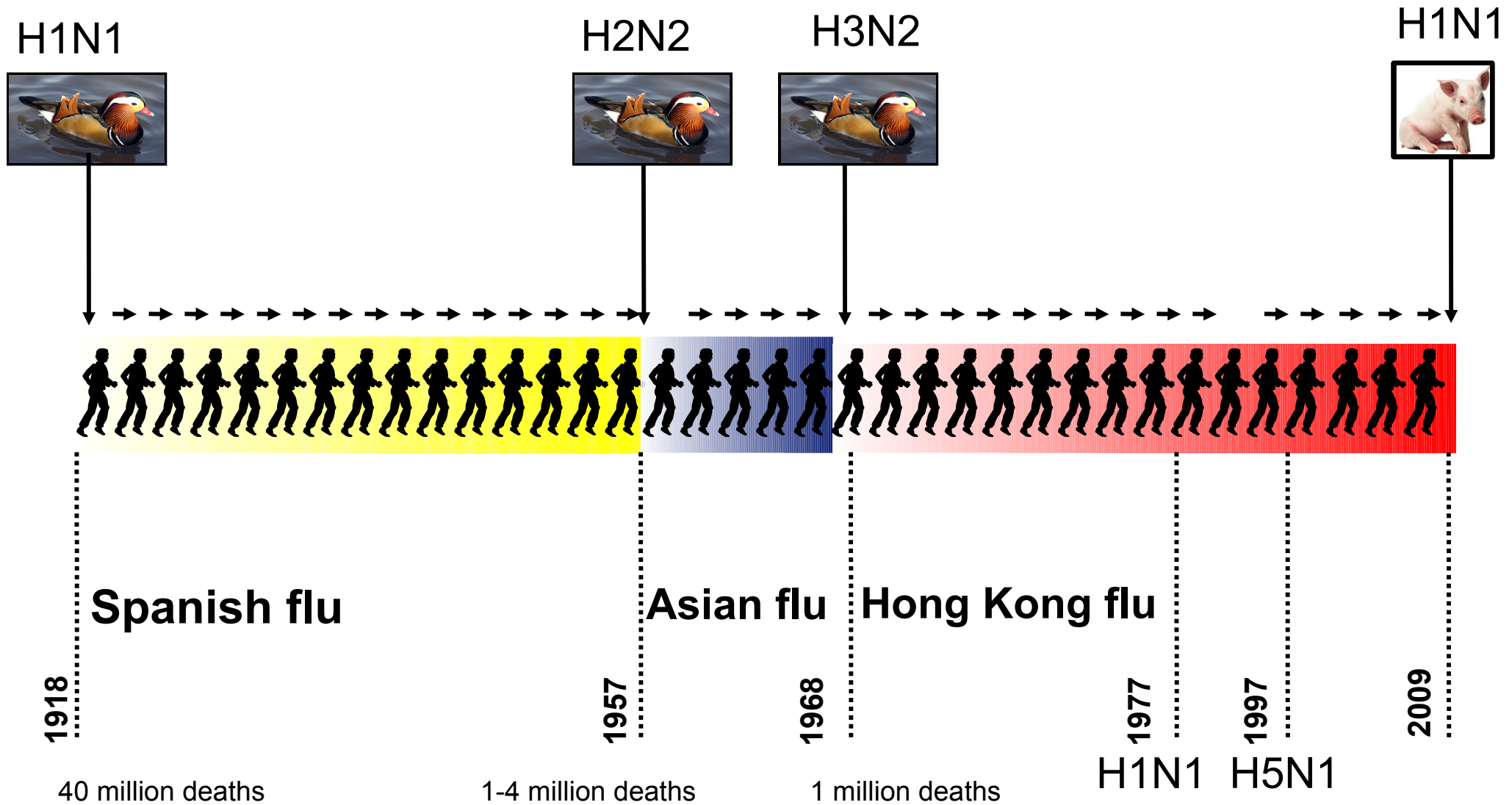
<http://www.antigenic-cartography.org>  
<http://www.zoo.cam.ac.uk/zoostaff/burke.htm>

[dfb21@cam.ac.uk](mailto:dfb21@cam.ac.uk)

9 September 2011



# Influenza virus: pandemic and epidemic



# The Influenza Virus

Annually, 'flu infects 5-15% of the global population (~600 million people)

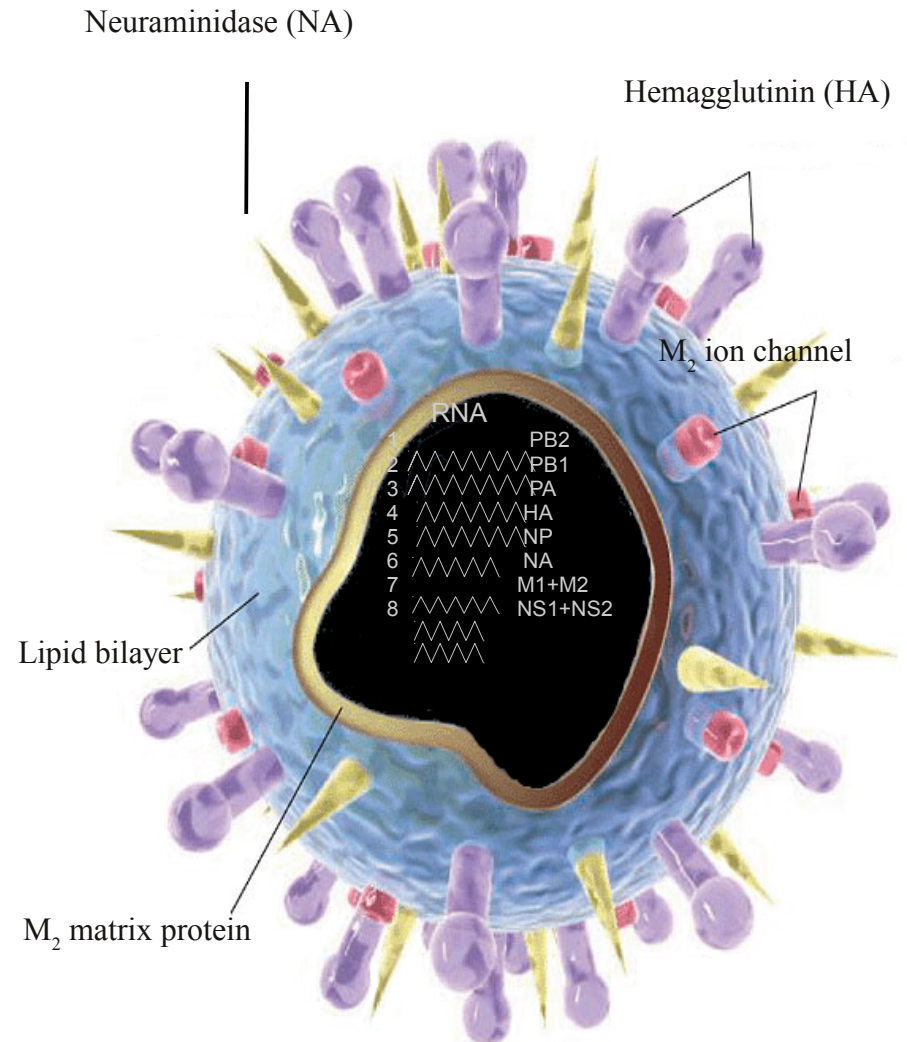
Virus genome contains 8 RNA segments which code 11 proteins

RNA polymerase makes a single nucleotide error roughly every 13 thousand nucleotides

Nearly every new influenza virus has a mutation thus is highly antigenically variable

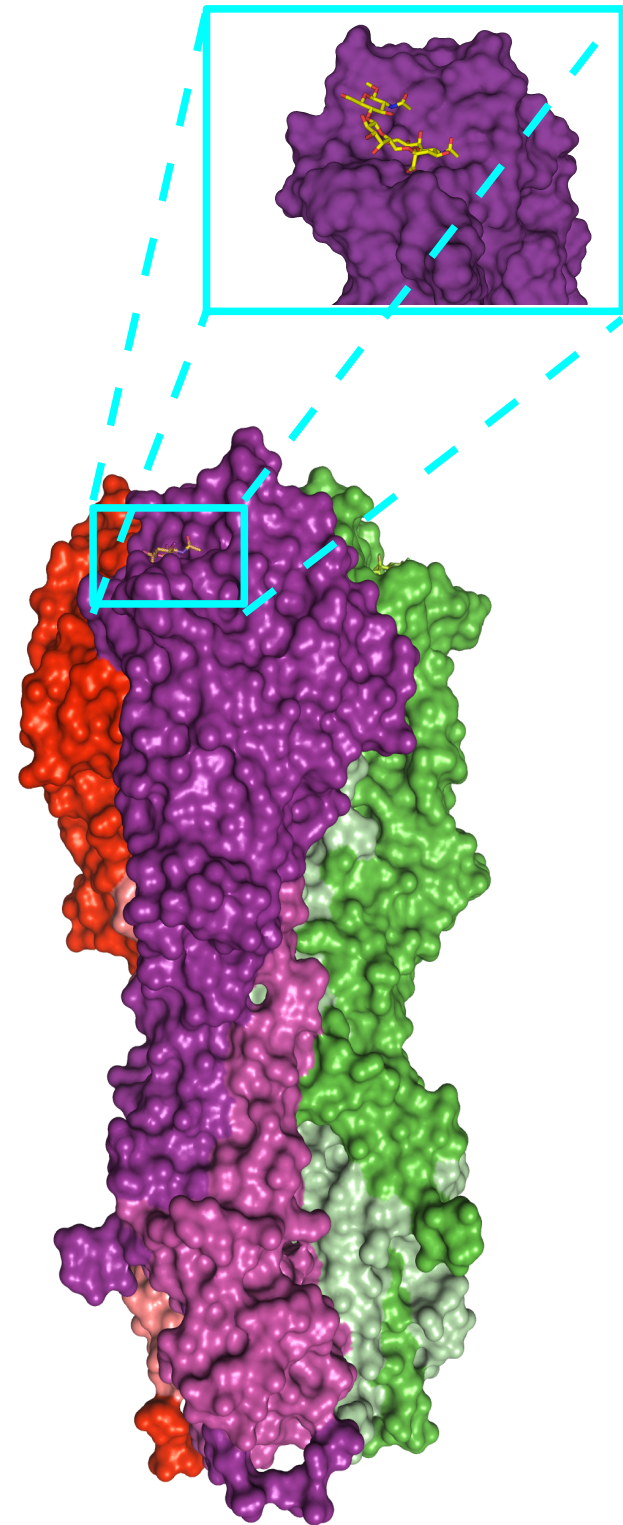
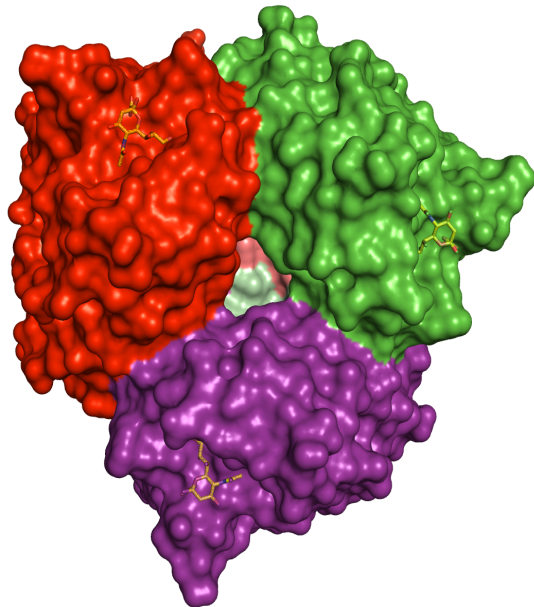
Over time, mutations build up and antibodies lose the ability to bind.

For this reason, the 'flu vaccine has had to be updated more than 20 times over the last 40 years.



# Haemagglutinin

- ▶ Protein expressed on surface of virus
- ▶ The major component of the 'flu vaccine
- ▶ The main focus of global 'flu surveillance
- ▶ There are ~500 HA proteins on surface of virus
- ▶ HA binds to sugars on the surface of the cell



# What are we interested in modelling?

## Antigenic Variation

Genetic changes which lead to vaccine breakdown

## Species Specificity

Binding preference (human/avian) and zoonotic potential

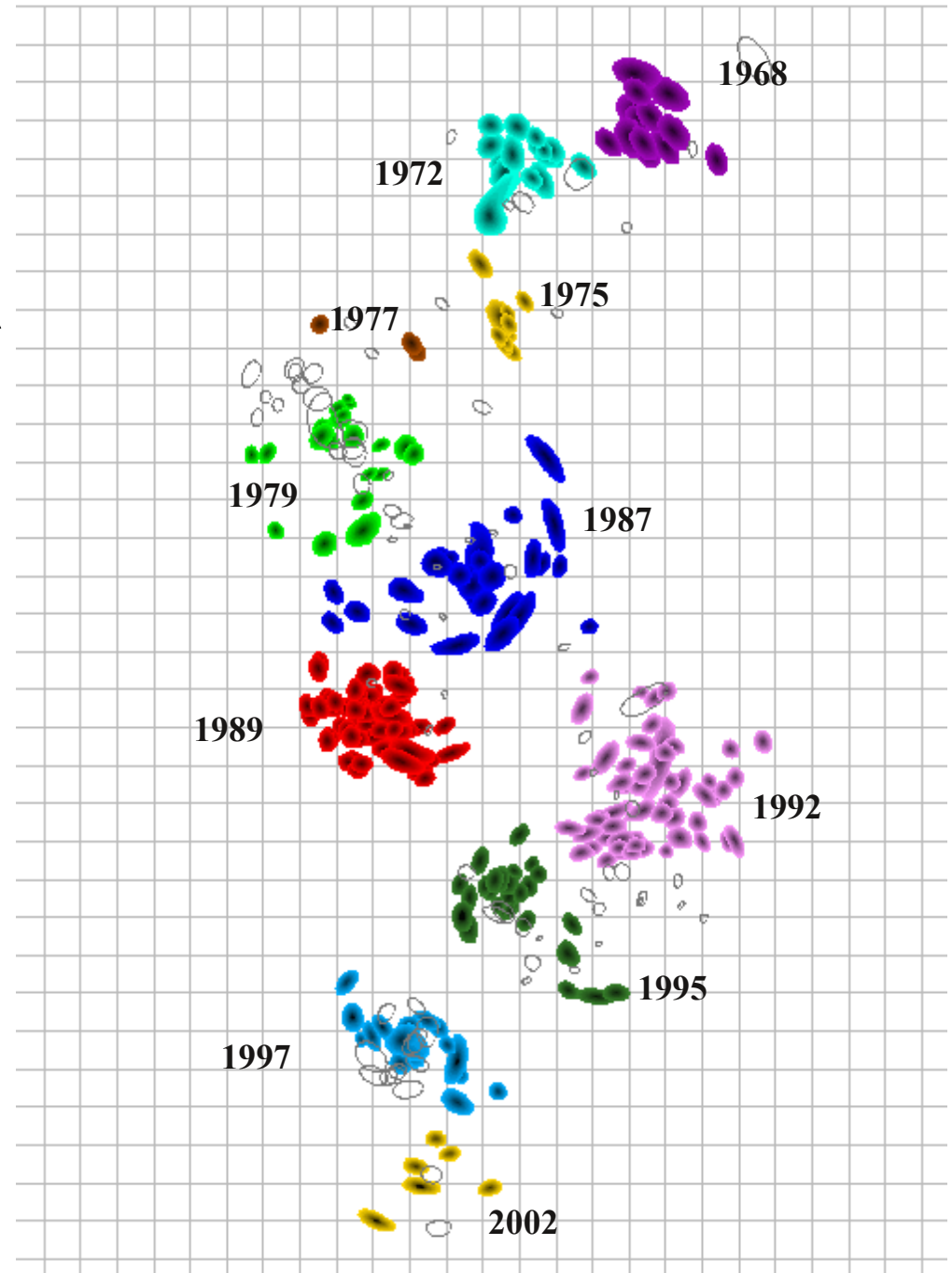
# Haemagglutinin-inhibition (HI) assay data

## HEMAGGLUTINATION INHIBITION REACTIONS OF INFLUENZA H3 VIRUSES (08/11/05)

STRAIN DESIGNATION			REFERENCE FERRET ANTISERA													
REFERENCE ANTIGENS	CDC ID#	Date collected	A	B	C	D	E	F	G	H	I	J	K	L	M	
			WY/03 REASS	WEL/01	CA/7	CA/7 REASS	VIC/500	SN/37 REASS	WI/19	NY/55	NY/55 REASS	HK/4593	NH/14	HK/2831 NEW	MS/5 NEW	
1	WY A/WYOMING/03/2003 X-147	2003715730	REASS	1280	80	160	320	320	40	80	80	40	80	320	40	160
2	NZ A/WELLINGTON/01/2004	2004729358	1/26/04	160	1280	640	1280	640	30	160	320	160	640	640	160	640
3	CA A/CALIFORNIA/07/04	2005705486	9/16/04	160	160	640	2560	1280	320	320	320	160	640	640	160	320
4	CA A/CA/07/04 x PR/8 CDC	2005712034	REASS	320	320	1280	2560	2560	640	640	1280	320	1280	1280	160	1280
5	AS A/VICTORIA/500/2004	2005707652	8/13/04	160	160	320	1280	640	80	160	160	80	320	320	80	320
6	AS A/SINGAPORE/37/2004 IVR-140	2005707637	REASS	160	160	640	1280	640	640	1280	640	160	640	640	80	320
7	WI A/WISCONSIN/19/2004	2005705540	11/6/04	160	320	1280	2560	640	320	1280	1280	160	640	640	160	640
8	NY A/NEW YORK/55/2004	2005705561	11/18/04	160	160	1280	1280	1280	320	640	640	320	640	640	160	640
9	NY A/NY/55/04 A/PR/8/34 X-157	2005711905	REASS	40	160	640	1280	640	160	160	320	640	320	640	160	1280
10	HK A/HONG KONG/4593/04	2005710182	10/18/04	160	160	320	2560	640	320	320	320	160	640	640	160	640
11	NH A/NEW HAMPSHIRE/14/04	2005710231	12/13/04	160	160	640	1280	640	320	640	320	160	640	640	160	640
12	HK A/HONG KONG/2831/05	2005743649	5/25/05	160	40	320	320	80	40	80	160	160	160	640	320	160
13	MS A/MISSISSIPPI/5/04	2005707679	12/14/04	40	80	320	640	320	80	80	160	320	160	320	160	640
TEST ANTIGENS	REPEATS															
14	TH 02-504942 ORIGINAL	2005707539		160	320	640	1280	2560	320	320	320	160	640	640	80	640
15	TH 02-504957 ORIGINAL	2005707540		160	160	320	1280	640	160	320	320	160	640	640	80	640
16	TH 02-505107 ORIGINAL	2005707542		160	320	640	1280	640	160	160	320	160	640	640	80	320
17	MS 05020500027 ORIGINAL	2005712779	2/1/2005	160	160	640	1280	640	320	640	640	160	640	640	160	640
18	UR A/KHABAROVSK/9/05	2005714223	1/27/2005	320	160	640	2560	320	160	320	320	320	640	640	320	320
19	DJ 2004914389	2005714244	12/17/2004	320	160	640	1280	640	320	640	320	160	640	640	160	640
20	DJ 2005900911	2005714283	1/26/2005	320	160	640	2560	640	320	640	320	160	640	640	160	640
21	DJ 2005900992	2005714289	1/29/2005	160	160	640	640	640	160	320	320	160	640	640	160	640
22	DJ 2005901307	2005714306	2/2/2005	160	320	640	1280	1280	320	320	320	160	640	640	160	320
23	DJ 2005901309	2005714307	2/4/2005	640	320	1280	2560	1280	640	640	640	160	1280	1280	160	1280
24	DJ 2005901315	2005714308	2/4/2005	640	320	1280	2560	1280	640	1280	1280	320	1280	1280	320	1280
25	DJ 2005901330	2005714309	2/4/2005	2560	640	2560	5120	5120	2560	2560	2560	640	5120	5120	640	5120
26	DJ 2005902082	2005714334	2/12/2005	320	640	1280	2560	2560	1280	1280	1280	320	2560	2560	320	2560
27	DJ 2005902100	2005714335	2/14/2005	640	320	1280	5120	2560	640	1280	2560	320	2560	1280	320	1280
28	DJ 2005902111	2005714336	2/24/2005	320	160	640	2560	2560	640	640	640	320	640	1280	160	640
29	PN 490730	2005740130	6/22/2005	320	160	640	2560	1280	640	640	1280	160	1280	640	160	640
30	PN 490728	2005740131	6/21/2005	20	80	160	640	160	80	40	80	80	80	160	160	160
31	PN 490714	2005740132	6/18/2005	320	320	1280	5120	2560	1280	1280	1280	320	1280	2560	320	1280
32	UR A/EKATERINBURG/1/05	2005740144	1/14/2005	160	40	320	320	80	80	80	160	80	160	160	80	80
33	UR A/YAROSLAVL/13/05	2005740147	2/1/2005	640	320	640	2560	640	1280	640	640	640	1280	2560	320	640
34	UR A/OMSK/23/04	2005740149	1/20/2004	160	160	320	640	640	160	40	80	80	160	160	160	320
35	UR A/VLADIMIR/37/05	2005740150	2/17/2005	160	80	320	640	320	80	320	320	160	320	320	160	320
36	UR A/KHABAROVSK/51/05	2005740151	3/5/2005	5	5	5	5	5	5	5	5	5	5	5	5	5
37	UR A/TOMSK/27/04	2005740153	1/7/2004	160	160	160	320	320	80	40	80	80	80	160	80	320
38	HK A/HONG KONG/4421/05	2005740176	6/8/2005	80	20	160	160	20	40	40	160	80	80	160	160	40
39	HK A/HONG KONG/4281/05	2005740177	6/16/2005	40	20	80	160	20	40	40	40	80	40	80	160	40
40	HK A/HONG KONG/4241/05	2005740179	6/16/2005	80	40	320	160	80	80	80	160	160	160	320	640	80
41	HK A/HONG KONG/4607/05	2005740181	6/14/2005	40	10	40	40	20	10	20	40	40	40	20	40	10
42	EG 2005904864	2005740677	4/20/2005	80	20	160	160	80	40	80	80	80	160	160	80	80
43	OM 2005906943	2005740704		80	20	160	160	10	20	40	80	80	160	80	80	40
44	SN A/SINGAPORE/03/05	2005741146	2/16/2005	320	160	640	640	320	160	160	320	160	320	640	160	320
45	SN A/SINGAPORE/07/05	2005741152	3/13/2005	80	10	80	80	20	20	40	80	80	40	80	40	20
46	SN A/SINGAPORE/34/05	2005741162	5/31/2005	40	20	160	160	20	20	40	80	80	80	160	160	80
47	SN A/SINGAPORE/35/05	2005741164	5/31/2005	40	10	20	40	10	10	10	20	40	20	40	80	20

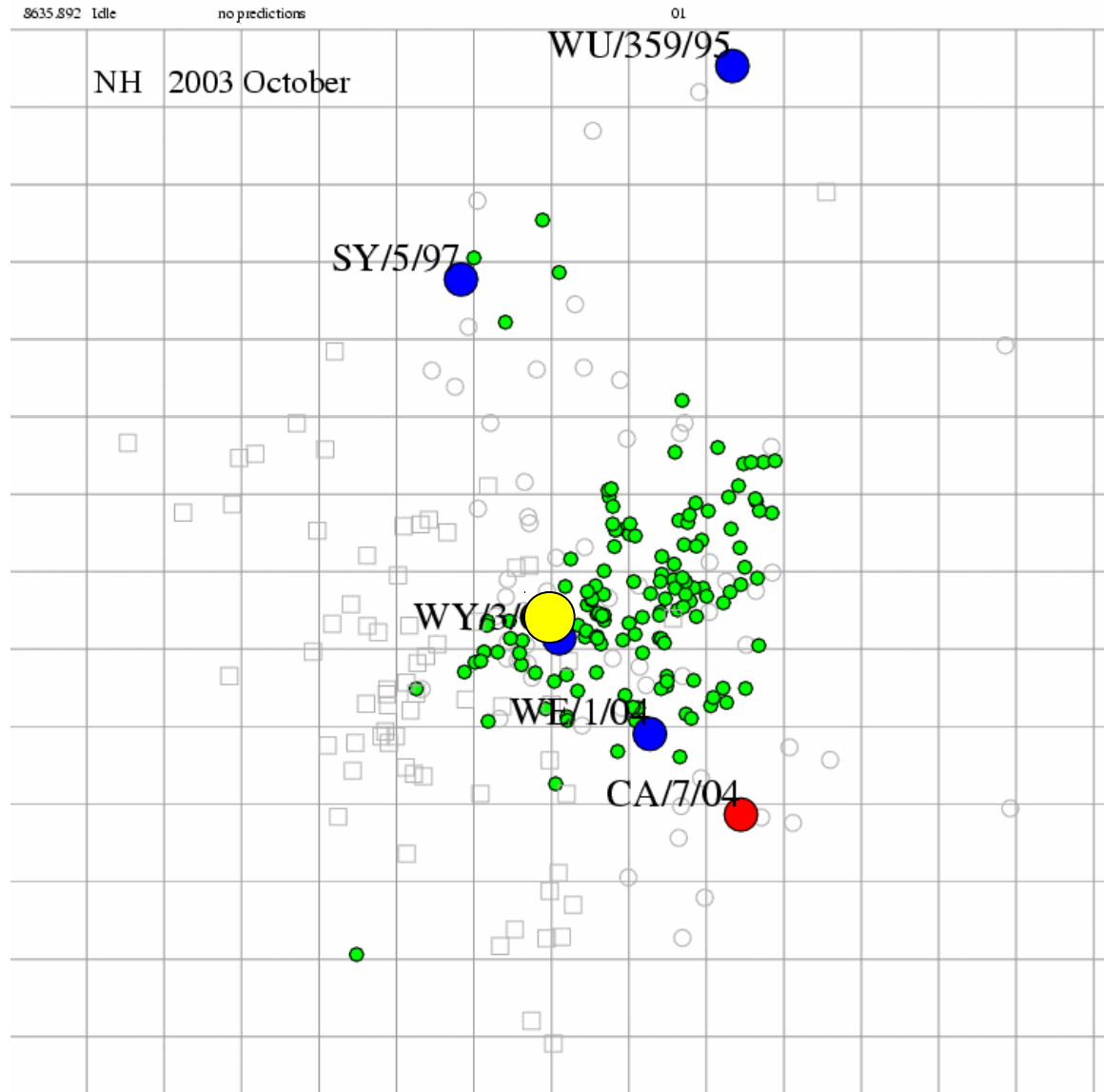
# “Antigenic map” of Influenza H3N2 1968-2003

- ▶ A genotype to phenotype map for HA
- ▶ Two dimensional
- ▶ Mostly linear
- ▶ Forms Clusters
- ▶ Chronologically ordered
- ▶ Equal time between clusters
- ▶ Equal distance between clusters



Smith *et al* 2004 Science

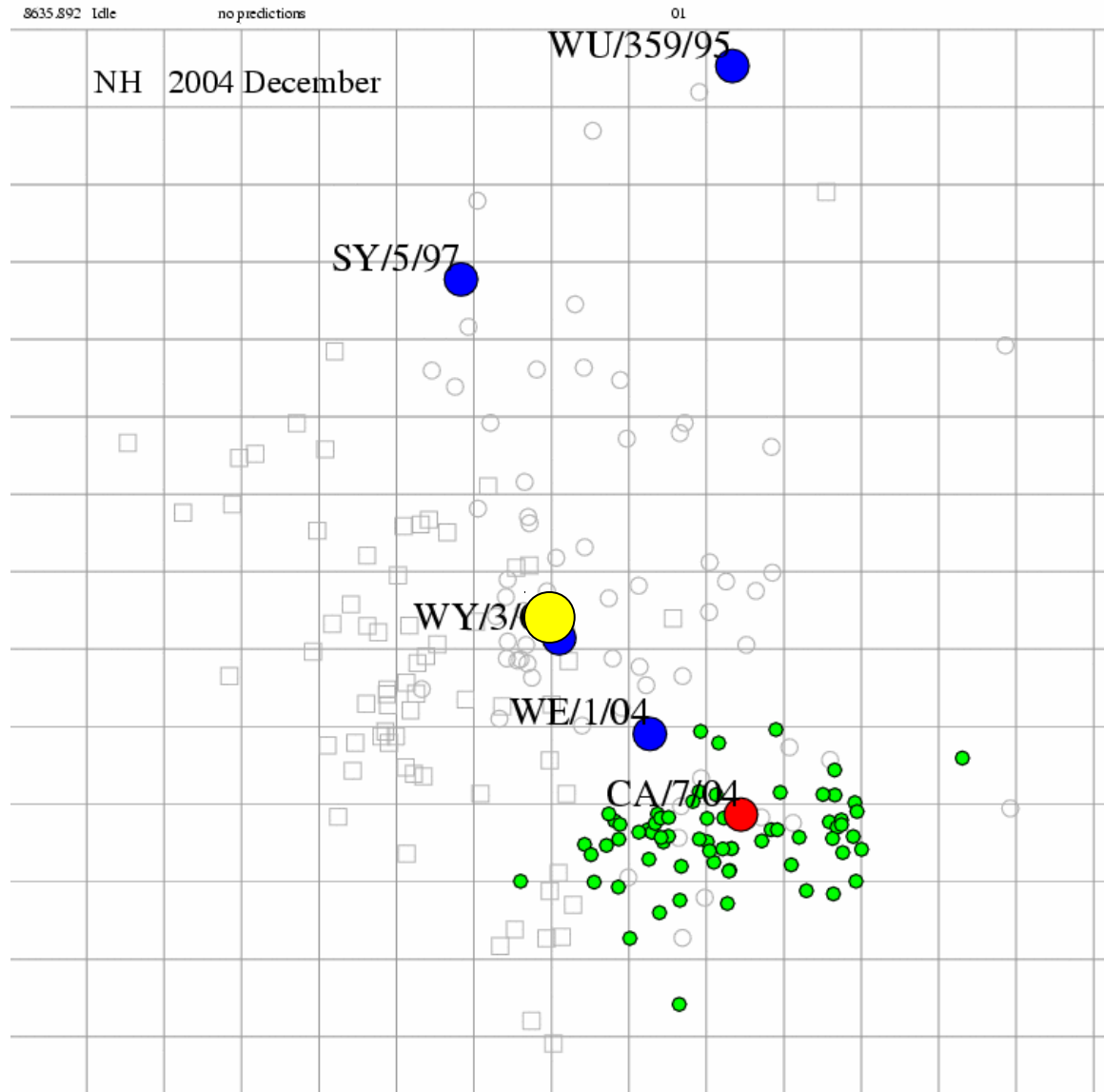
# Northern Hemisphere October 2003



- Currently circulating antigens
- Previous cluster
- Current vaccine strain
- Potential vaccine strain

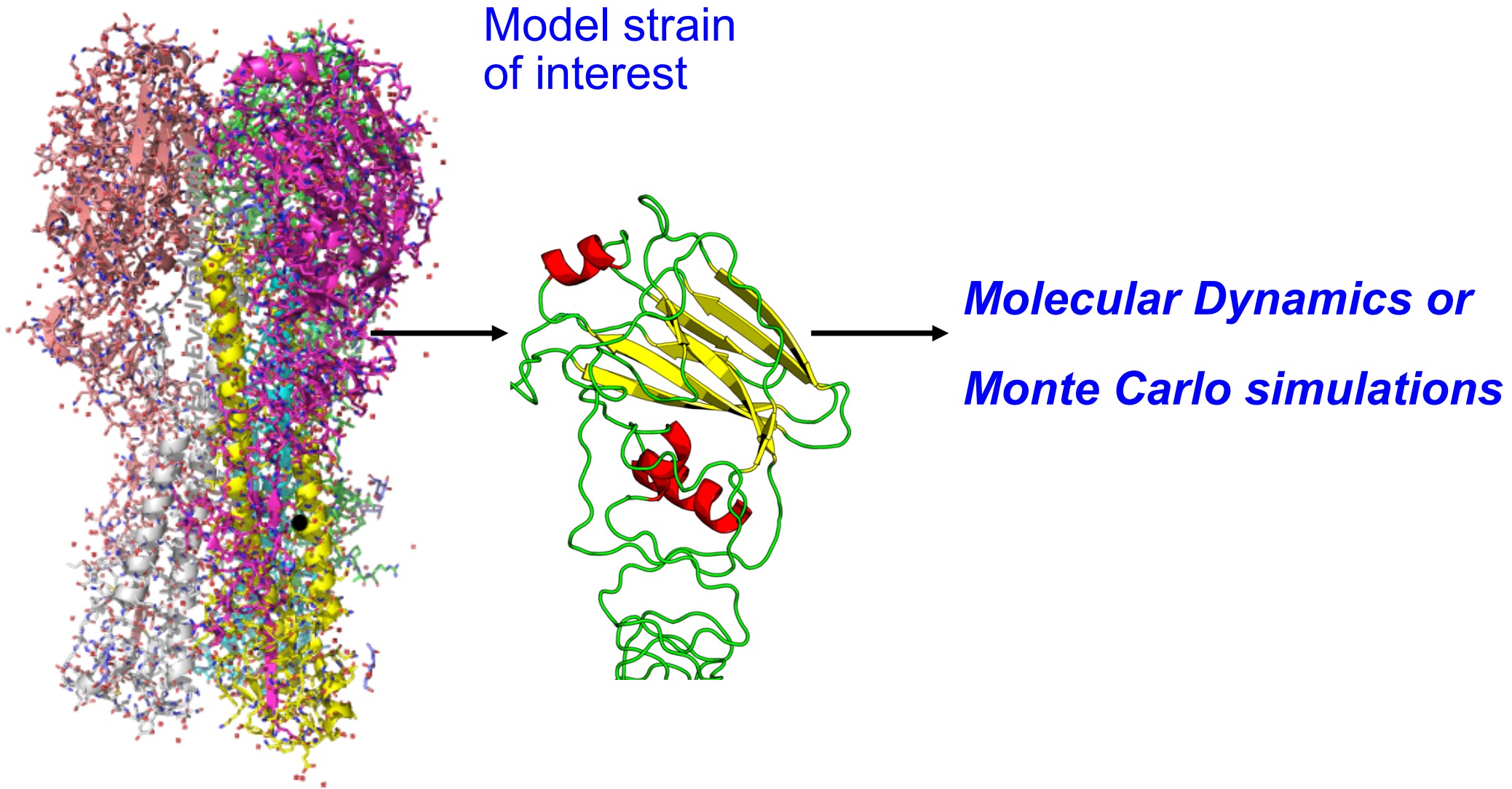


# Northern Hemisphere December 2004



- Currently circulating antigens
- Previous cluster
- Current vaccine strain
- Potential vaccine strain

# *In silico* predictions of the structure of the virus



Xray structure of a strain of HA

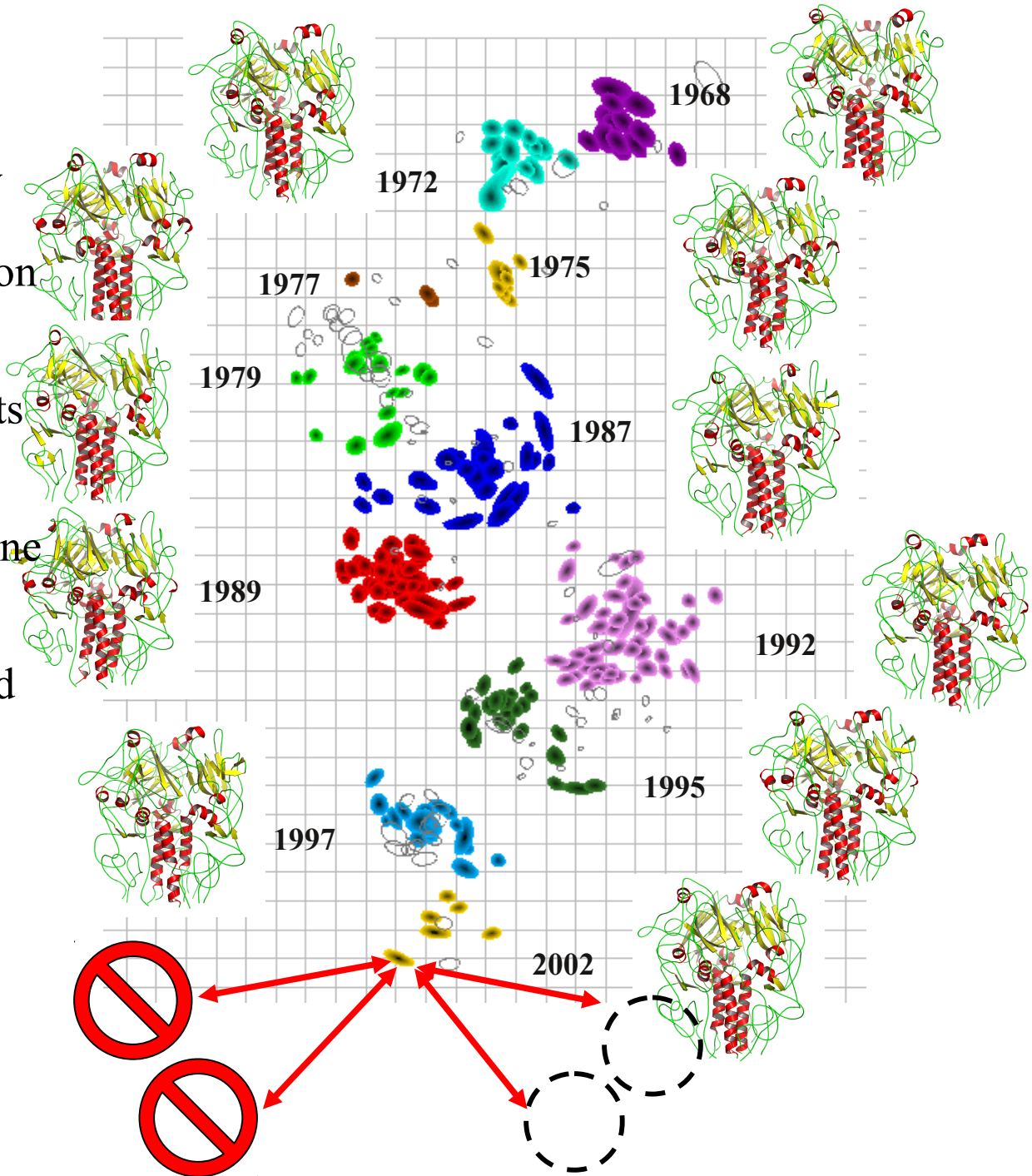
# *In Silico* Genotype to phenotype prediction of the evolution of HA

Given the sequence (genotype) of any HA, we could predict its tertiary structure and its phenotype and position on the 'map'.

Predict which mutations would alter its phenotype.

This would aid surveillance and vaccine selection

Pre-emptive vaccination could be used to block the virus from moving in a particular direction



# What are we interested in modelling?

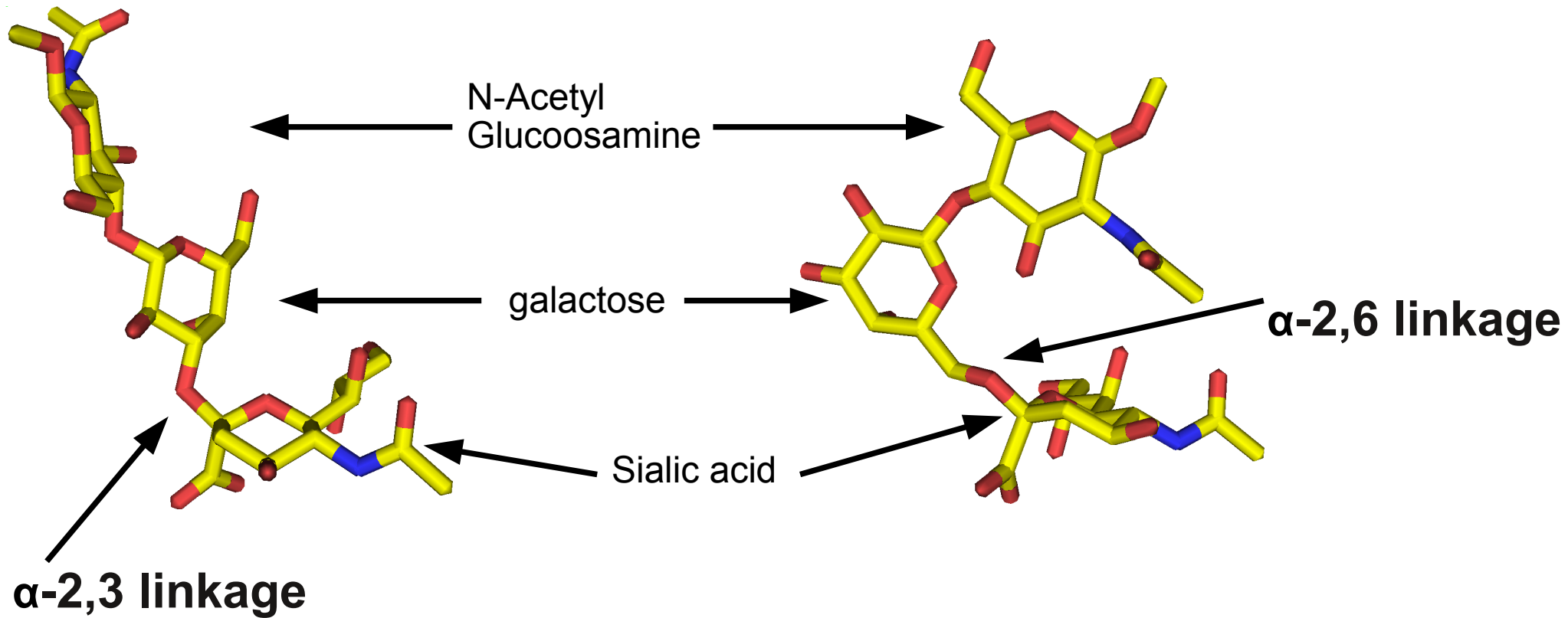
## Antigenic Variation

Genetic changes which lead to vaccine breakdown

## Species Specificity

Binding preference (human/avian) and zoonotic potential

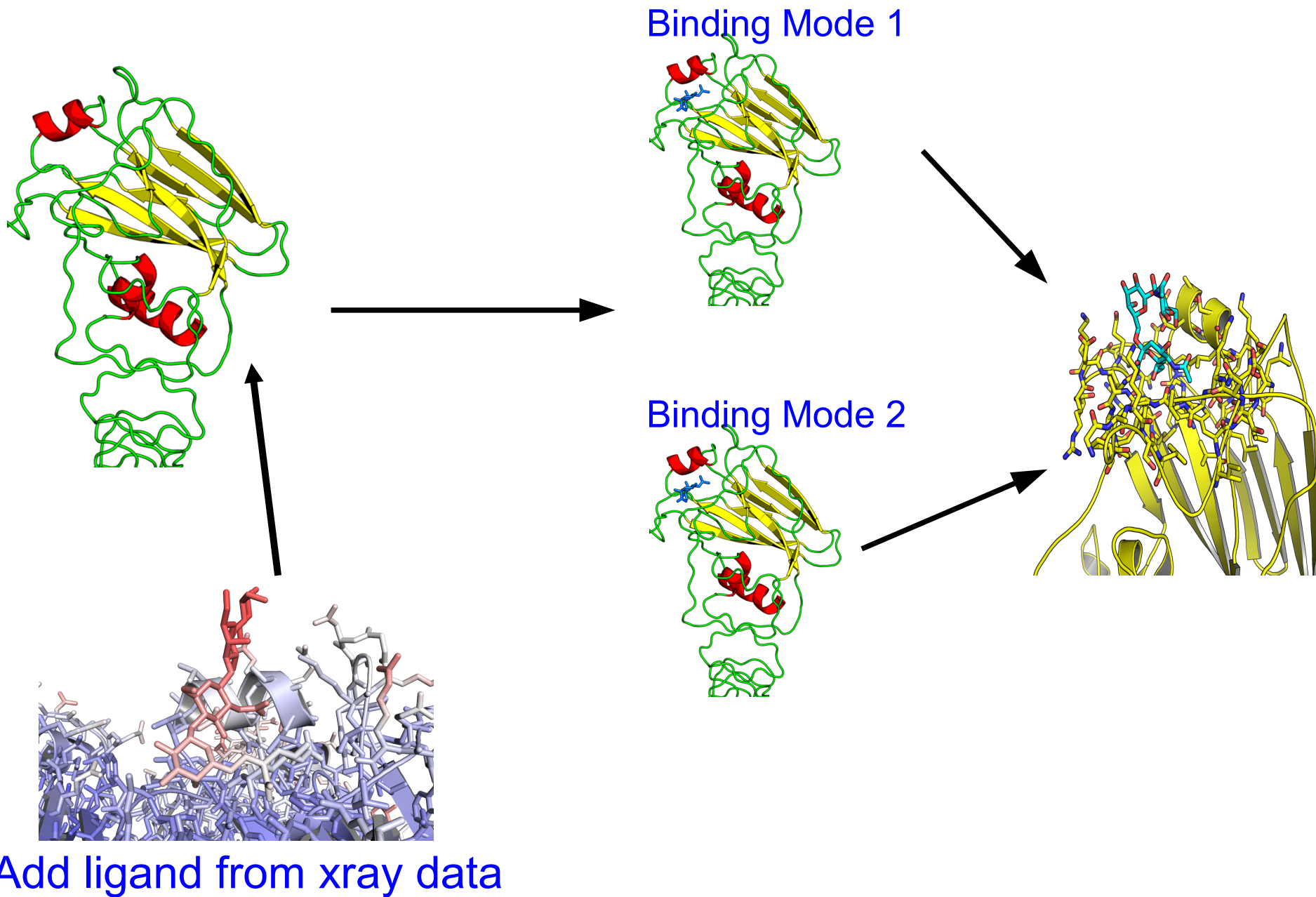
# Determinants of ligand specificity



Avian, Equine & Pig  
adapted influenza  
viruses

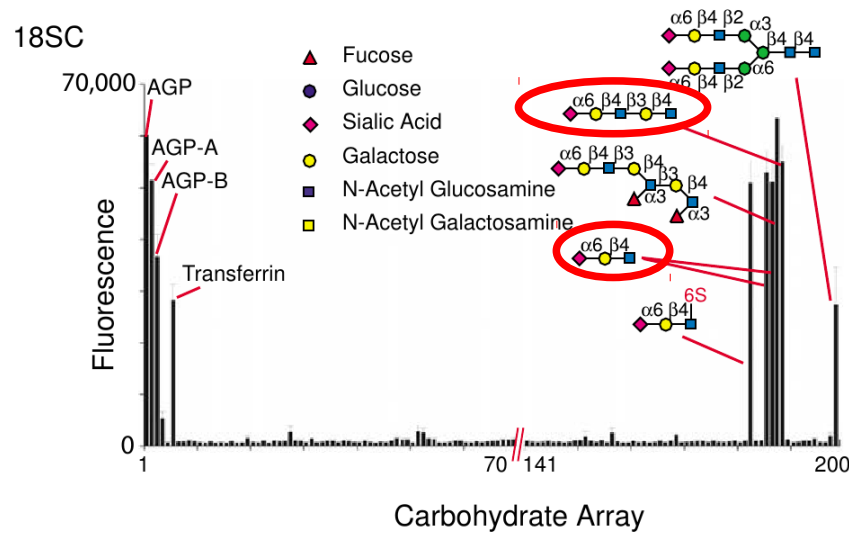
Human & Pig  
adapted influenza  
viruses

# *In silico* predictions of ligand binding

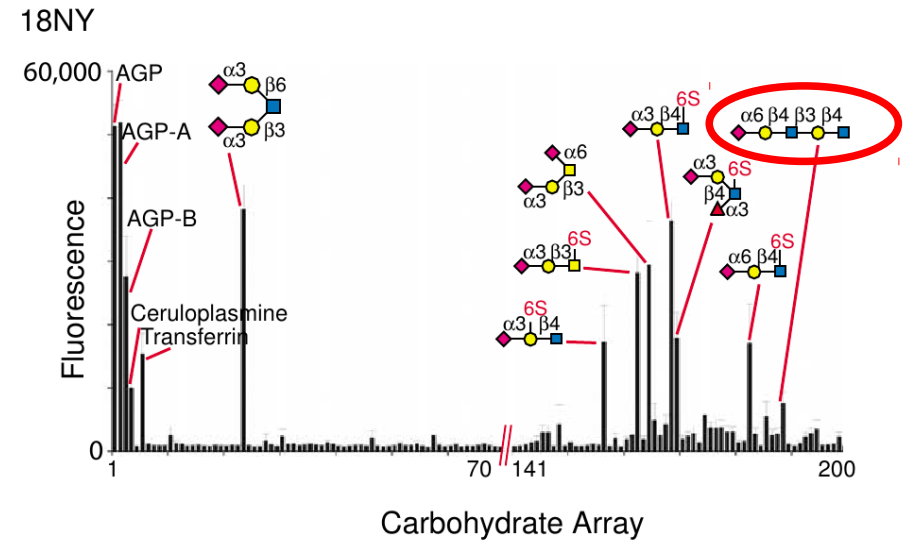


# Sugar binding properties of H1N1 A/South Carolina/1/1918

## A/South Carolina/1/1918



## A/South Carolina/1/1918 + D222G

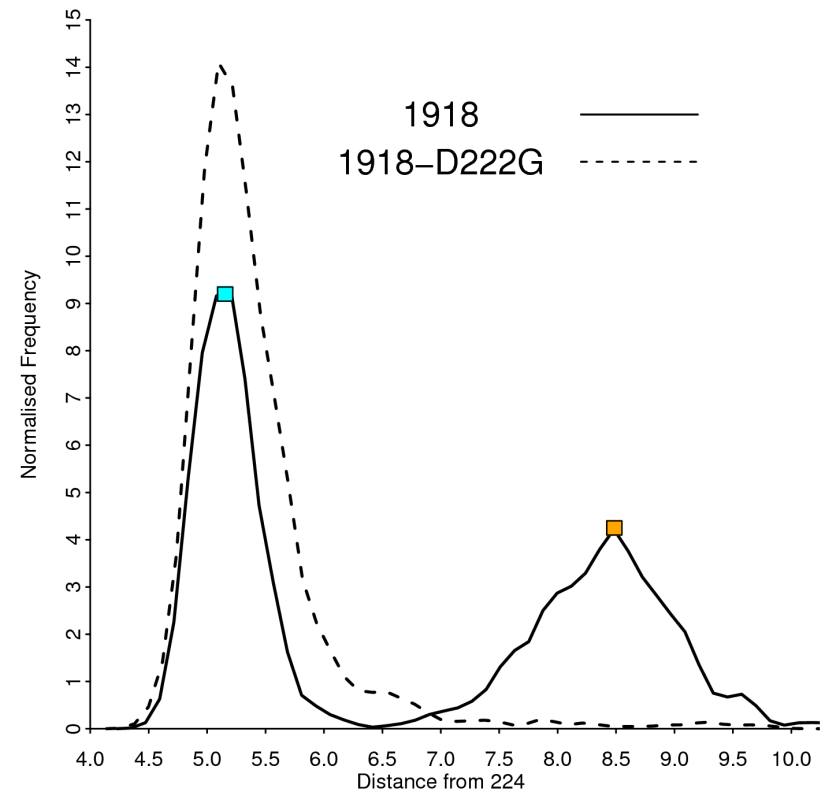
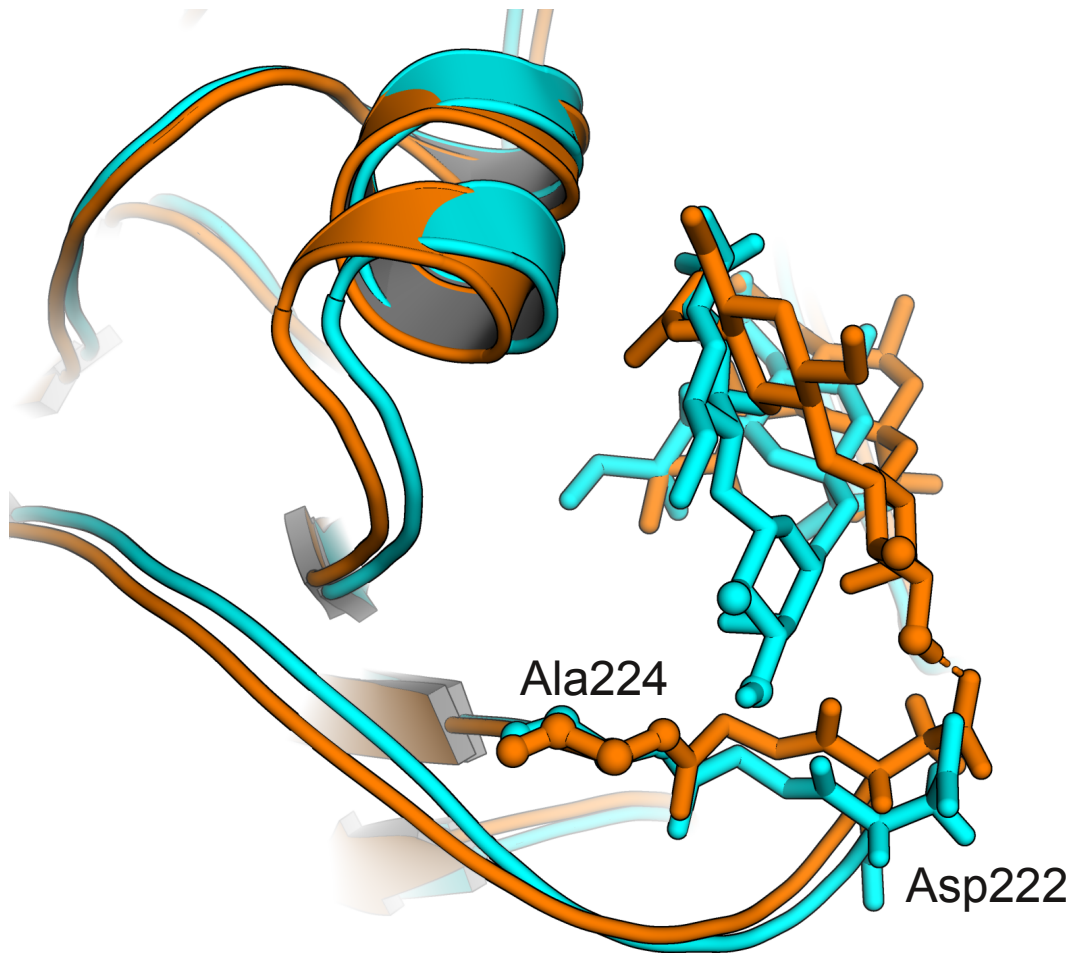


### The single mutation D222G

- ◆ Gains binding to sulphated  $\alpha 2,3$  glycans (increased mortality)
- ◆ Loses binding to  $\alpha 2,6$  glycans (poor transmission)

How does the D222G mutation affect binding in 2009 pdm H1N1 strain?

# Molecular Dynamics of A/South Carolina/1/1918 with $\alpha$ 2,6 glycan



■ ——— ■  
mode 1

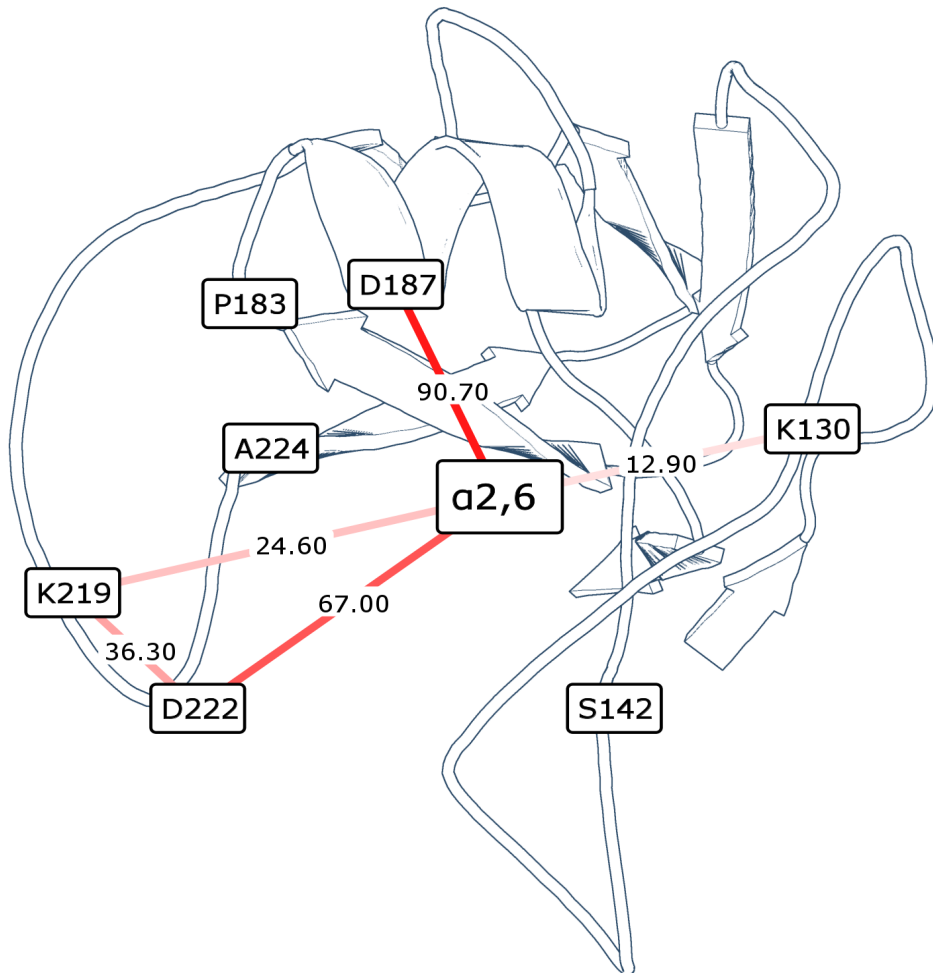
■ ——— ■  
mode 2

~1300 structures extracted from each simulation



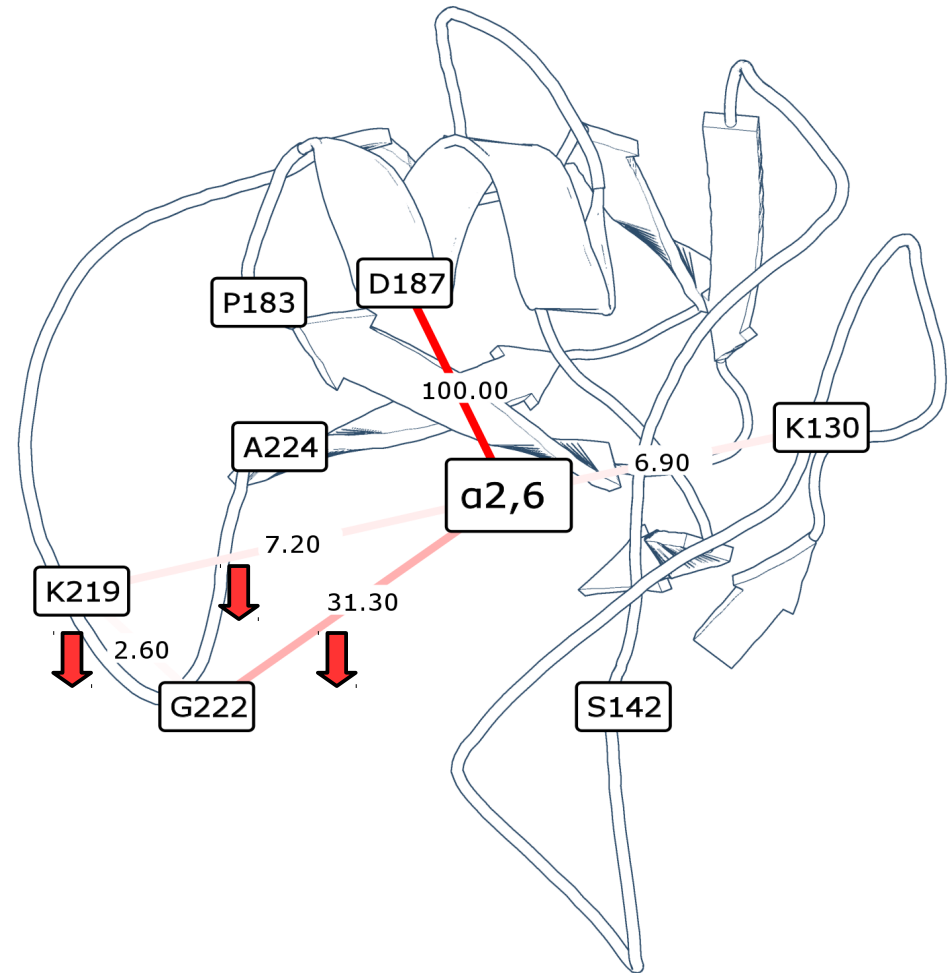
# Hydrogen binding frequencies involving ligand

A/South Carolina/1/1918



Main interactions with sidechain of D222 and D187

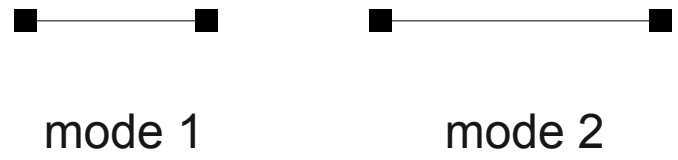
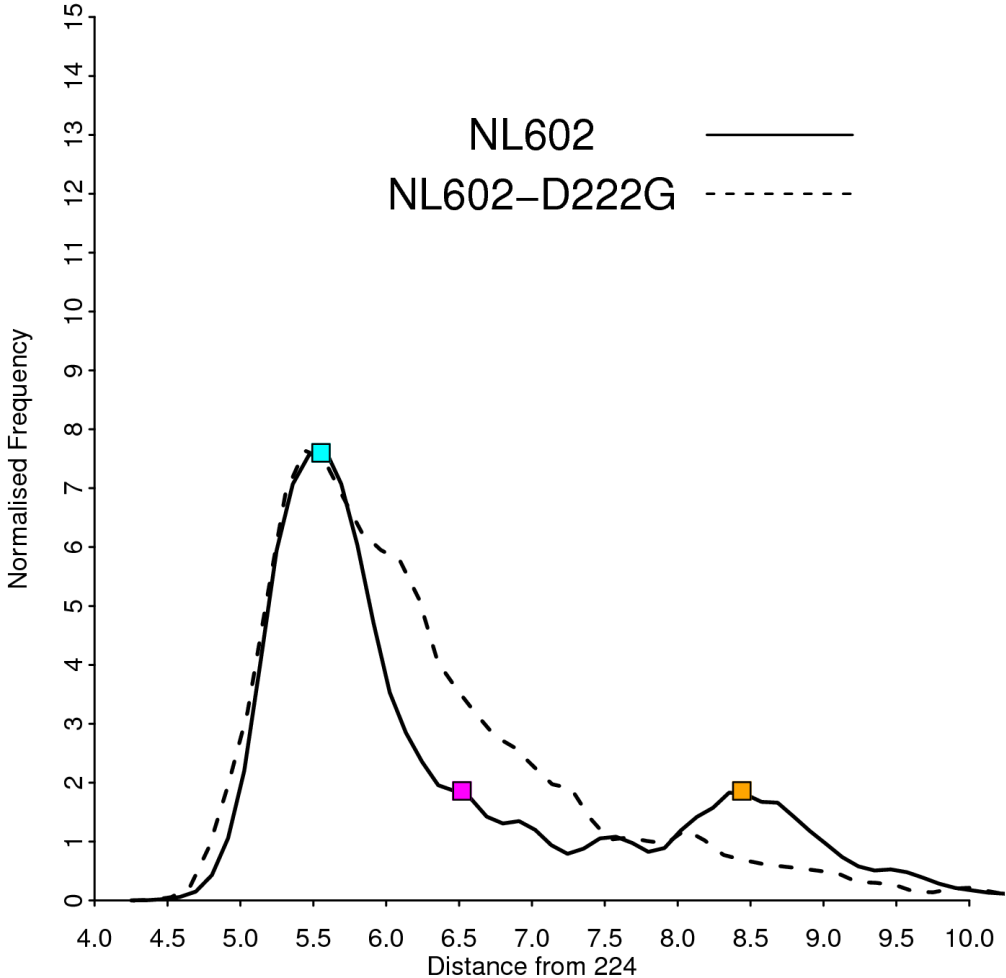
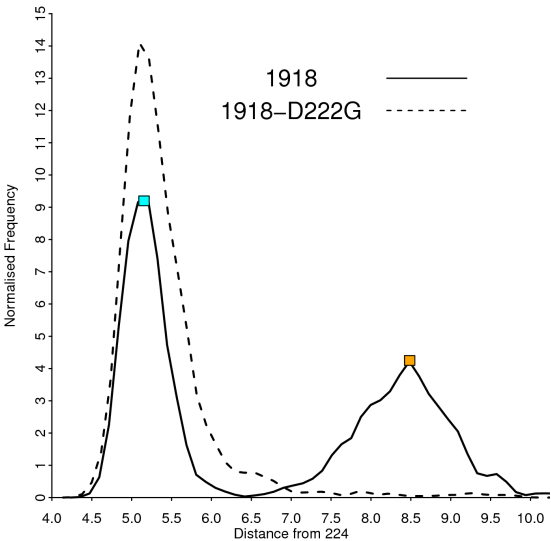
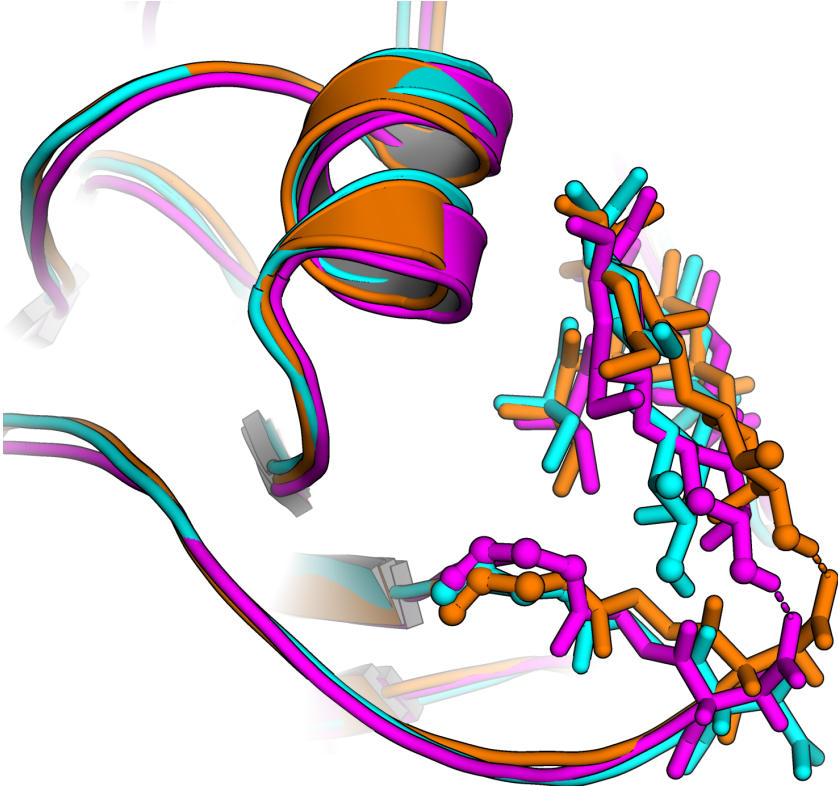
A/South Carolina/1/1918 +D222G



D222G Reduces possible interactions

D222G strain makes fewer interactions with  $\alpha 2,6$  glycan than wild type.

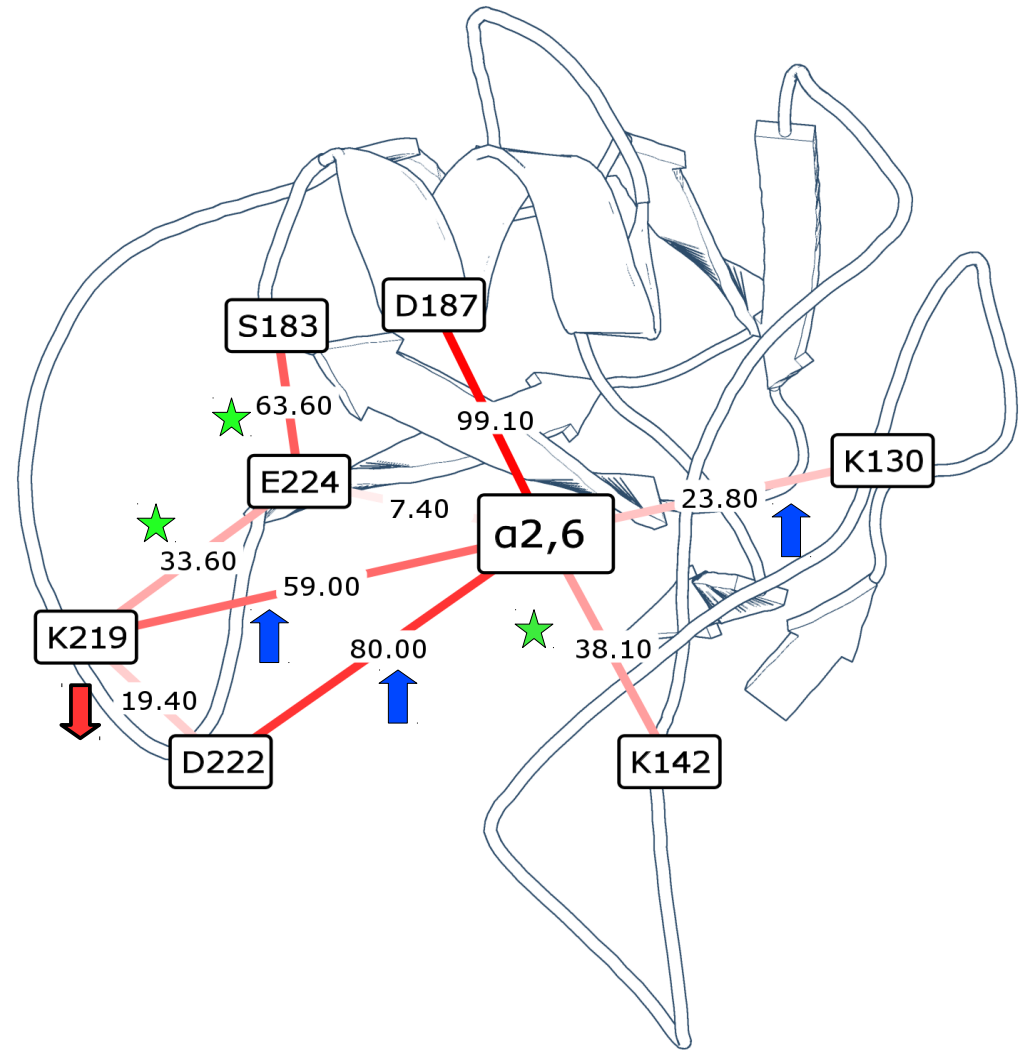
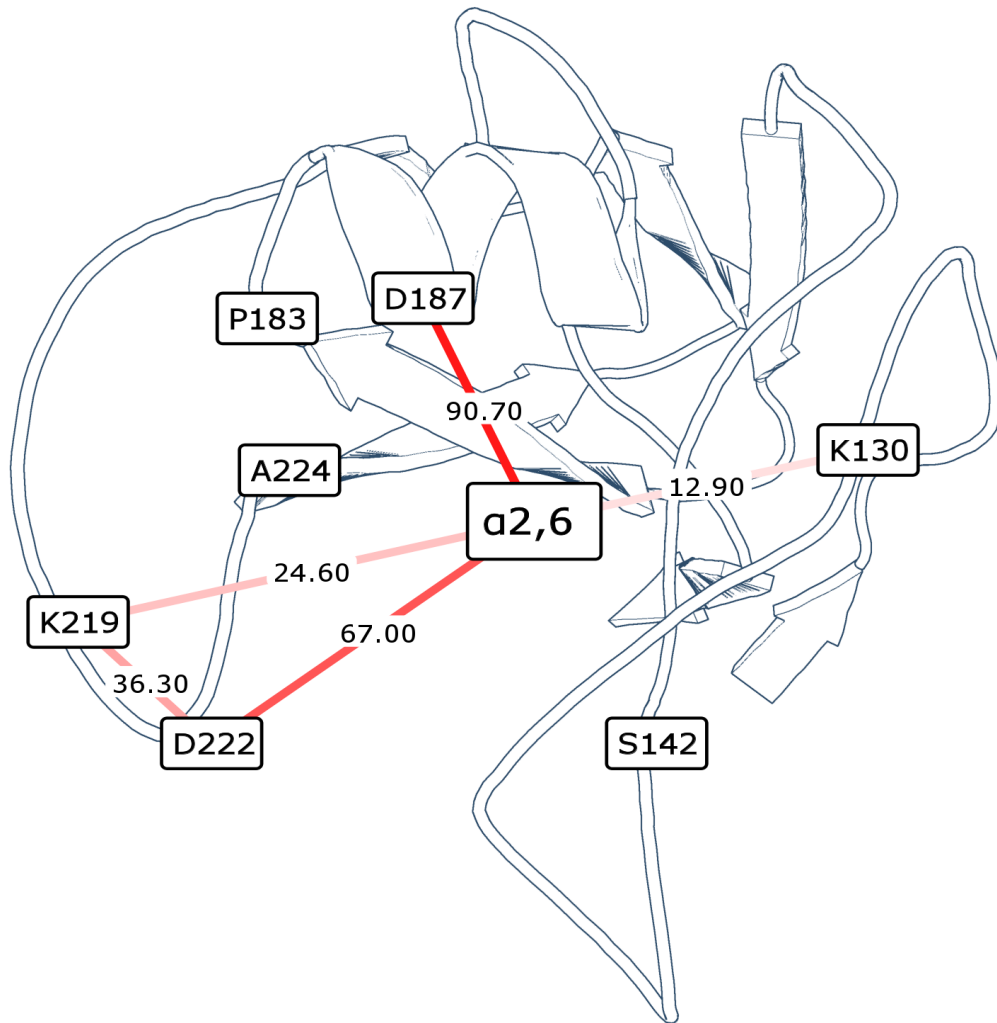
# 2009 pdmH1N1 A/NL/602/2009



More continuous distribution between two modes

**A/South Carolina/1/1918**

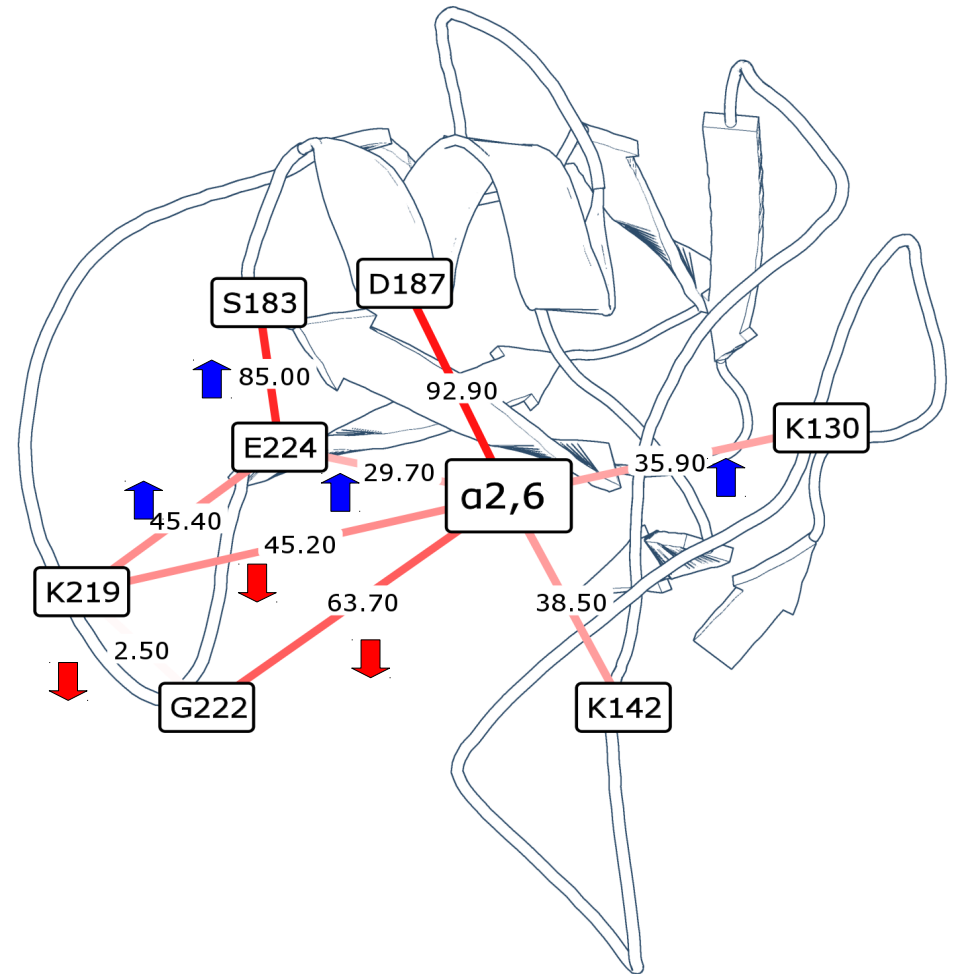
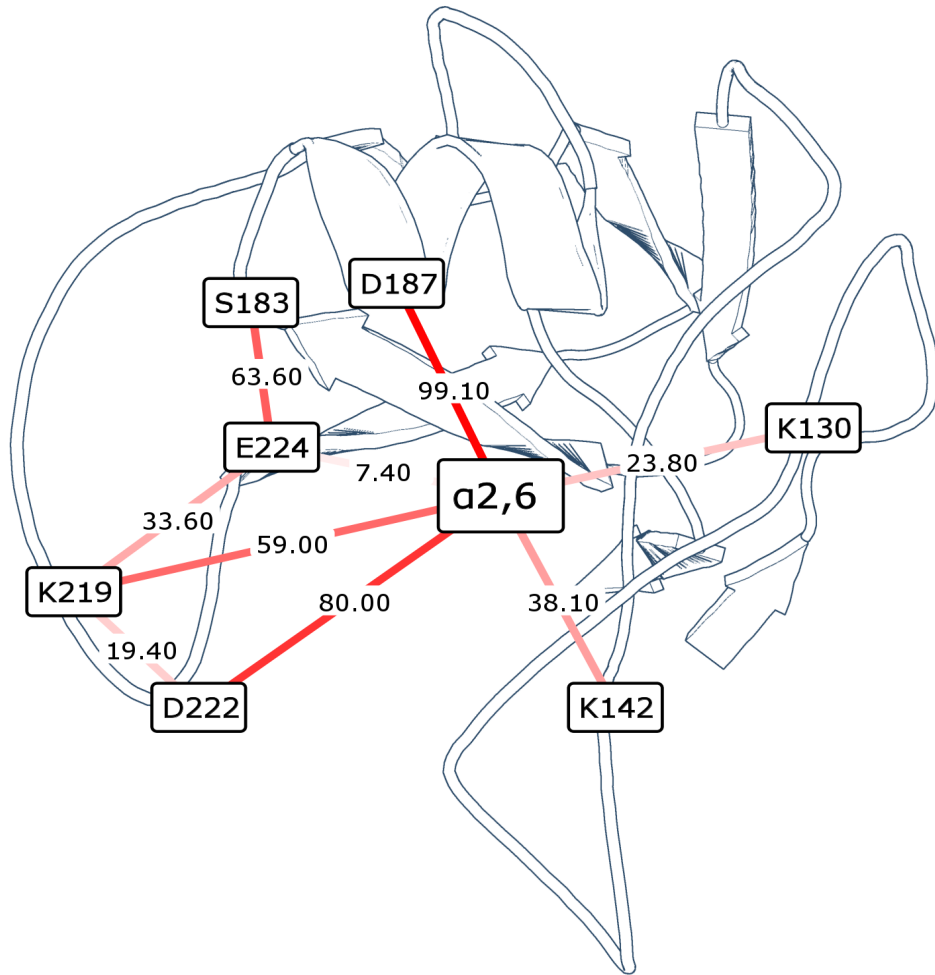
**PDM A/Netherlands/602/09**



Genetic differences provides extra networks of interactions which stabilise intermediate binding mode

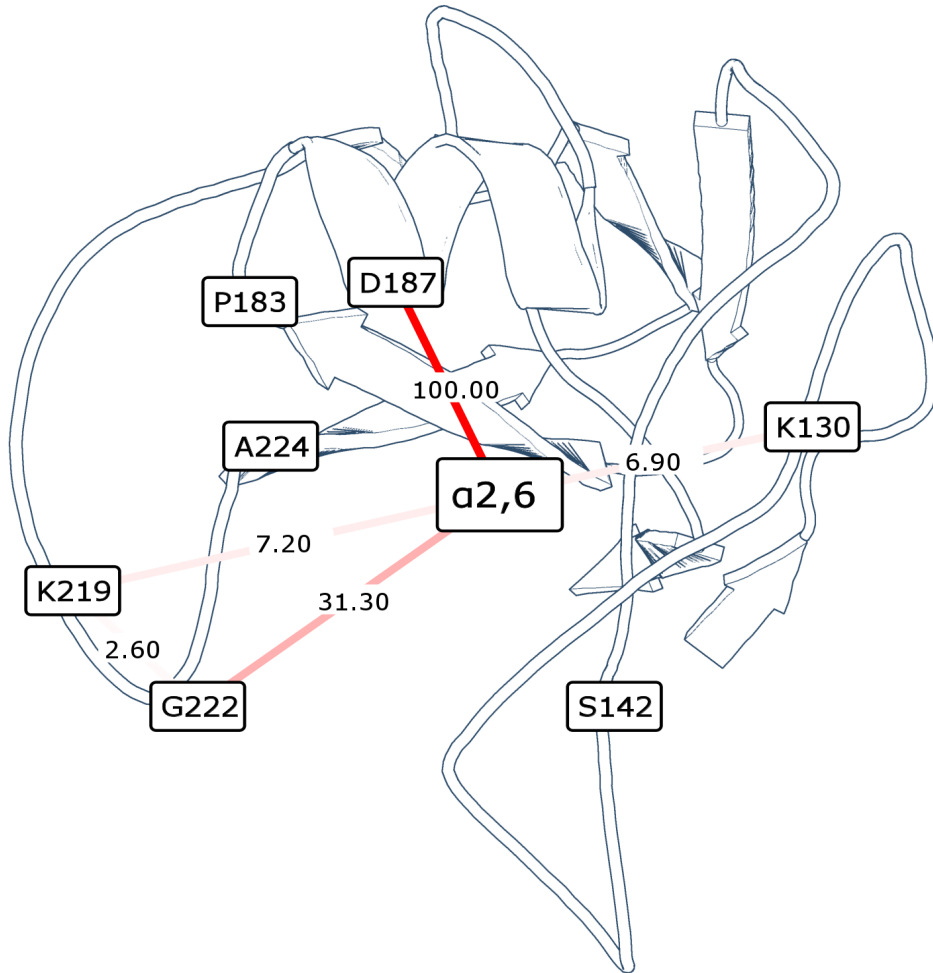
# A/Netherlands/602/09

# A/Netherlands/602/09+D222G

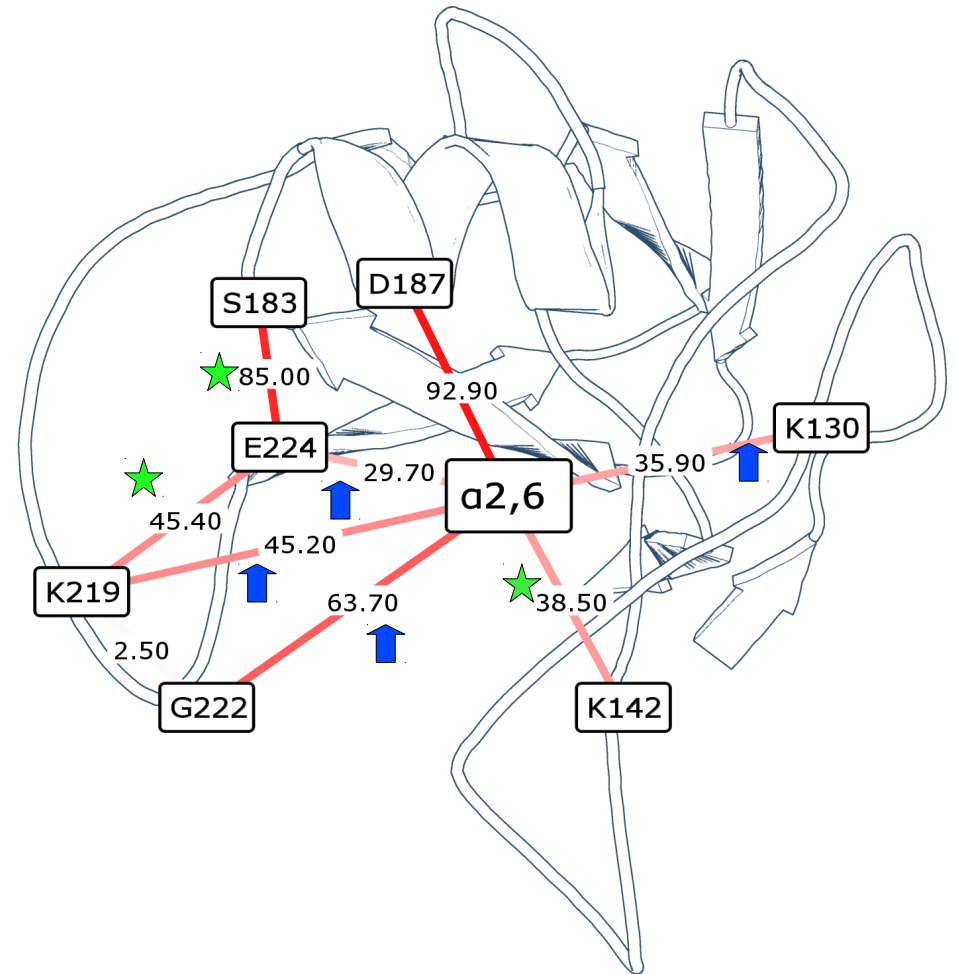


Decrease of 222 interactions compensated for by increase in interactions with E224, S183 and K130.

# A/South Carolina/1/1918 +D222G



# A/Netherlands/602/09 +D222G



We would predict that, unlike A/SC/1918+D222G, A/Netherlands/602/09+D222G binds just as well as A/Netherlands/602/09

# Predictions from the MD simulation

- ▶ Two distinct binding modes of the galactose sugar for SC1918
- ▶ A more continuous distribution for A/Netherlands/602/09
- ▶ A role for E224, S183, K130 in binding
- ▶ Reduction of interactions for SC1918+D222G
- ▶ Conservation of binding for A/Netherlands/602/09+D222G

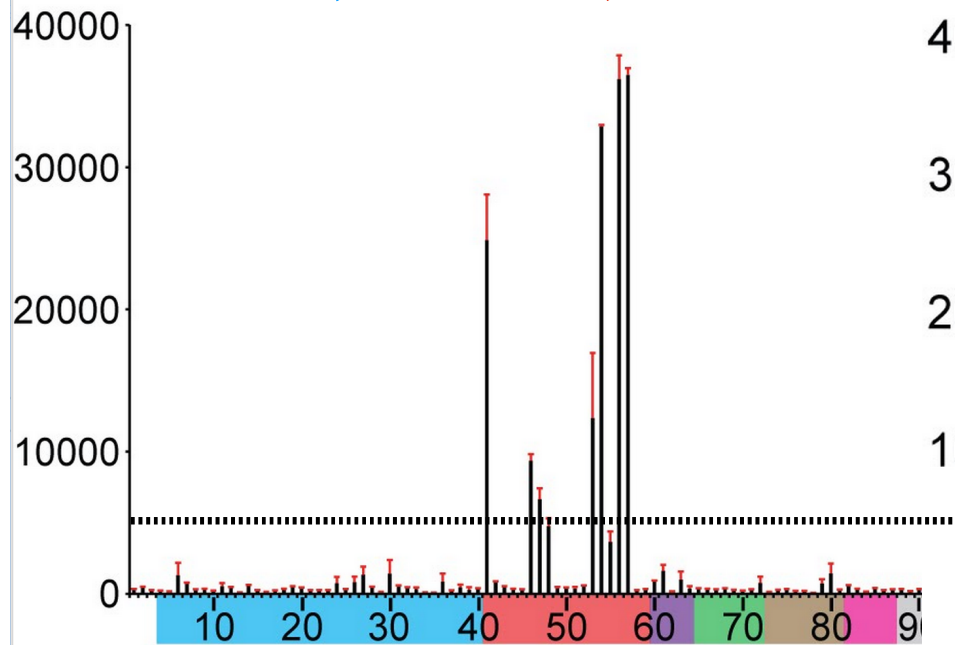
Hua Yang, Paul Carney, and James Stevens.

## Structure and Receptor binding properties of a pandemic H1N1 virus hemagglutinin.

PLoS Curr Influenza. 2010

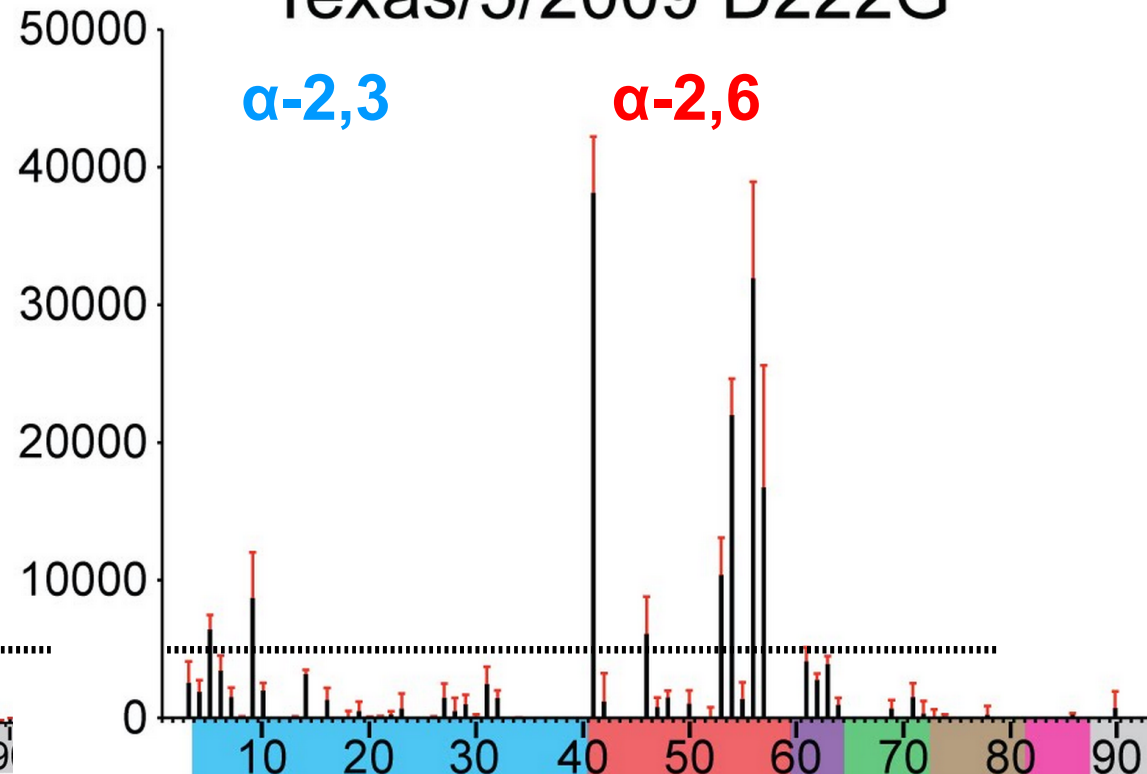
Texas/5/2009

$\alpha$ -2,3       $\alpha$ -2,6



Texas/5/2009 D222G

$\alpha$ -2,3       $\alpha$ -2,6



2009 PDM + D222G maintains  $\alpha$ 2,6 binding

# Future Work

- ♦ Improvements in quantitative binding calculations
- ♦ Investigate binding to  $\alpha$ 2,3 glycans
- ♦ Investigate evidence for multiple binding modes
- ♦ More robust docking and sampling of the glycans into the RBS
- ♦ Longer simulations for benchmark set for improved statistics



## University of Cambridge

Derek Smith  
Terry Jones  
Colin Russell  
Nicola Lewis (& AHT)  
Dan Horton (& VLA)  
Ana Mosterin  
Eugene Skepner  
David Wales (Chemistry)  
Kyle Sutherland-Cash (Chemistry)  
Chris Whittleston (Chemistry)

## **CamGrid sys admins**

## Erasmus Medical Centre

Ron Fouchier  
Salin Chutinimitkul  
Jan de Jong  
Bjorn Koel  
Miranda de Graaf  
Vincent Munster  
Guus Rimmelzwaan  
Walter Beyer  
Theo Bestebroer  
Ruud van Beek  
Ab Osterhaus

## Funding

NIH Director's Pioneer Award,  
Fogarty International Center,  
HFSP, IFPMA, CIDC,  
EU Framework 5 Novafllu,  
EU Framework 6 Virgil

## WHO global influenza surveillance

CDC: Nancy Cox, Sasha Klimov,  
Michael Shaw  
MELB: Ian Gust, Ian Barr, Aeron Hurt,  
Alan Hampson  
NIMR: Alan Hay, Y-P Lin, Vicky Gregory  
NIID: Tashiro Masato, Takato Odagiri  
WHO: Wenqing Zhang, Klaus Stohr  
NICs: Critical and enormously valuable

## Santa Fe Institute

Alan Lapedes